Chapter 12
Optimized Design of Large-Scale Social Welfare Supporting Systems on Complex Networks

Jaroslav Janáček, Ľudmila Jánošíková, and Ľuboš Buzna

Abstract Our contemporary societies are supported by several systems of high importance providing large-scale services substantial for citizens everyday life. Typically, these systems are built or rely on various types of complex networks such as road networks, railway networks, electricity networks, communication networks etc. Examples of such systems are a set of emergency medical stations, fire or police stations covering the area of a state, social or administration infrastructure. The problem of how to design these systems efficiently, fairly, and reliably is still timely and it brings along many new research challenges. This book chapter presents a brief survey of optimization models and approaches applicable to the problem. We pay special attention to the methods based on the branch and bound principle and show how their computational properties can be improved. Furthermore, we discuss how some of these models can be rearranged in order to allow using the existing solving techniques as approximative methods. The presented numerical experiments are conducted on realistic data describing the topology of the Slovak road network. On the one hand, we hope that this chapter can come handy to researchers working in the area of complex networks, as it presents efficient methods to design public service systems on the networks. On the other hand, we can picture the benefits potentially resulting from the knowledge of the network properties and possibly being utilized in the algorithms design.

Keywords Public service systems • Large-scale emergency systems

J. Janáček • Ľ. Jánošíková • Ľ. Buzna (✉)
Faculty of Management Science and Informatics, Department of Transportation Networks, University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovak Republic
e-mail: Jaroslav.Janacek@fri.uniza.sk; janosik@frdsa.uniza.sk; buzna@frdsa.uniza.sk

M.T. Thai and P.M. Pardalos (eds.), *Handbook of Optimization in Complex Networks: Theory and Applications*, Springer Optimization and Its Applications 57,
DOI 10.1007/978-1-4614-0754-6_12,

12.1 Introduction to Public Service Systems

The topology of many real-world systems can be described as a network. In some cases, the system organization and the network topology result from some long-term evolution processes directed by billions of interactions [15, 28, 39]. In other cases this is given by human-controlled decision-making activities, supported by sophisticated optimization methods [1, 18, 42]. In this book chapter, we revise some of the techniques that can be applied to the problem of designing the public service systems operating on all sorts of transportation networks. We focus mainly on road networks, where the typical applications are locating emergency medical centers, police or fire stations [38, 41] or other types of social and administration infrastructures [19, 27]. Nevertheless, these approaches can also be applied to many other man-made systems such as railway networks, telecommunication networks [37], power grid networks or gas networks [34, 40].

The road network is the example of the spatial complex network featuring many non-trivial topological properties [31]. So far, it is relatively little known how these properties influence the efficiency of optimization algorithms. For location problems it is, for example, known that the corresponding solving methods perform much more efficiently on real-world networks than on their random counterparts [30]. The explanation is in the spatial structure of the underlying network, which generates suitable diversity in cost coefficients. This influences the structure of lower and upper bounds and finally results in shorter computational time.

The road network is a substrate for the public service systems. These systems provide various kinds of services to inhabitants of a certain geographical area. The services can include goods delivery, presence of some necessary facilities or they can be some kind of civil services, such as medical care, fire-fighting or house maintenance. Contrary to the private service systems, none of the demands listed above can be ignored. Hereafter, a serviced person or a group of serviced people, characterized by their demand and located within the same municipality, will be referred to as a customer. The customer's demand will be characterized by its weight. If a customer is a group of people, the weight can be proportional to the number of the people in the given group.

Even if the serviced population is concentrated in municipalities, the number of municipalities is usually too large to have a source of the services at each customer's location. That is why placing the sources needs to be really thought through. Hereby, we assume that there is only a finite set of possible service center locations within the serviced area. Nevertheless, the set can be very large, e.g. it can be composed of all municipalities. Therefore, if an appropriate number of service center locations is to be found, a rather large combinatorial problem arises.

When a public service system is designed, two essential types of objectives, namely the economic efficiency and the fairness, should be taken into account [9]. The economic efficiency can be measured through how much the system and its operation cost. The costs of a service system usually consist of fixed expenses paid for opening and operating service centers, and of the cost of transportation related to the serving the customer [13].

This scheme applies only if the system provider serves customers at their locations. However, some systems can provide services only at the service centers. In this case, a customer has to travel to the service center, and obviously, the travel cost is not included in the system costs. The servicing costs are often considered as being proportional to the network distance between the customer and the service center.

Fairness or welfare in the public systems design is related to the notion of justice perceived by the customers resulting from the distribution of limited resources. This topic has been extensively studied, notably in social sciences, welfare economics, and engineering. The overview of basic concepts of fairness can be found in the reference [4]. The customer's welfare is difficult to quantify. Therefore, one possibility is to convert it into a customer's discomfort, expressing the accessibility of a service. The discomfort of an individual customer can be, for example, estimated by considering the distance or travel time required to get from the customer's residence to the nearest service center. When we use this approach, several possibilities emerge. The discomfort can be expressed directly as the distance or only as a part of the distance which exceeds a given accepted real valued threshold $D_{\max}$. The simplest option is to utilize the information whether the distance is longer than $D_{\max}$. If the distance is longer or equal to $D_{\max}$, then the customer's discomfort is considered to be affected.

Next, two extreme cases, the average customer discomfort or the worst case of customer's discomfort, can be considered as possible criteria expressing the quality of the proposed design. In the first case, the estimations of individual discomforts are summed up, and the resulting design minimizes the total discomfort. The second criteria (sometimes denoted as Rawlsian criterion) measures the quality of the design by the discomfort perceived by the worst positioned customer [36]. This approach can be extended by repeating the minimization for the second, third, fourth, etc., worst positioned customer. This extension has been studied throughly in the context of flow networks, where it is known as the max-min fair allocation of flows [33], however, only little attention has been paid to it in the context of location problems [10]. Another possibility, known as the proportional fairness, is based on the minimization of the utility function, which is the sum of logarithms of individual discomforts [35].

If the system costs and the customer's welfare are defined, it is necessary to decide which type of the public service system design should be preferred. Is it the cost-oriented design or the customer's welfare-oriented design? The cost-oriented design searches for the system optimum, minimizing the system costs, while assuring the desired level of the welfare. The welfare-oriented design, on the other hand, minimizes the customer's discomfort subject to a constraint keeping the system costs under a given limit. Both approaches can be combined with an arbitrary evaluation criteria for either cost or discomfort.

In Sect. 12.2, we show some basic examples of how to formulate costs and discomfort mathematically. With each individual case we demonstrate how to build mathematical models which can be then solved by optimization tools. In Sect. 12.3, as examples of solving methods, we briefly review the available universal solvers and exact algorithms to solve the uncapacitated facility location problem. To conclude, in Sect. 12.4, we describe a case study where we examine the network of emergency medical service (EMS) centers operating in the Slovak Republic.

12.2 Modeling Approaches to Designing Public Service Systems

Let us assume that the serviced area consists of the municipalities located at the nodes of the graph $G(N,E)$, where N is a set of nodes and E is a set of edges. The municipalities forming the finite set $J \subseteq N$ are considered as customers. The demand or the number of inhabitants living in the node $j \in J$ is denoted as b_j. The service system design problem can be reduced to the task to decide where to locate the centers within the large finite set I of possible locations, so that the value of the chosen criterion is minimal.

The shortest path length between the nodes $i \in N$ and $j \in N$ is denoted as d_{ij}, and the associated travel time as t_{ij}. The fixed charge f_i is paid to locate the service center at the node i, whereas the costs to satisfy the j-th customer from the service center i can be expressed as $(ed_{ij} + g_i)b_j$, where g_i are the costs spent to satisfy one unit of the demanded volume, and e are the travel costs per one unit of volume and one unit of distance.

Let $s(j)$ be the index of the center which is the closest to the node j, considering either the time or the distance, and which belongs to the set $I_1 \subseteq I$ of nodes where a service center is located. Then, the total system costs can be expressed as:

$$\sum_{i \in I_1} f_i + \sum_{j \in J} (ed_{s(j),j} + g_{s(j)})b_j. \tag{12.1}$$

Here, we ask the question: How could we estimate the discomfort of the customers? Let j be the customer's location and $s(j)$ be the nearest service center as defined above. If the individual customer's discomfort $u_{s(j),j}$ can be expressed as being linearly proportional to the distance between the customer j and the nearest service center, then $u_{s(j),j}$ is equal to $d_{s(j),j}b_j$, considering b_j as the weight. In the case, when only the part of the distance exceeding the given threshold $D_{\max}$ is considered as being proportional to the discomfort perceived by the customer, then $u_{s(j),j}$ is equal to $(d_{s(j),j} - D_{\max})b_j$ subject to $d_{s(j),j} \geq D_{\max}$, and otherwise the discomfort $u_{s(j),j}$ is equal to zero. Whenever only the information whether the distance is longer than $D_{\max}$ is used, the customer's discomfort $u_{s(j),j}$ equals b_j subject to $d_{s(j),j} \geq D_{\max}$; otherwise, the discomfort $u_{s(j),j}$ is equal to zero.

Equivalent definitions of customer's discomfort can be used for travel time $t_{s(j),j}$ and the threshold value $T_{\max}$.

If the average discomfort is considered to be an appropriate measure, then the objective function expressing the discomfort experienced by the population can be expressed by the formula (12.2). Alternatively, if the worst customer's discomfort is some more characteristic quantity, then the expression (12.3) is used instead, to describe the general discomfort affecting the serviced population.

$$\sum_{j\in J} u_{s(j),j} \tag{12.2}$$

$$\max\{u_{s(j),j} : j \in J\} \tag{12.3}$$

To formulate the mathematical model, both the set $I_1 \subseteq I$ and assignments $s(j)$ are expressed as decision variables. Each possible location $i \in I$ is subject to the decision whether to provide a service center or not. This decision can be modeled by the variable y_i, which takes the value of one if a service center is located at the node $i \in I$ and the value of zero otherwise. The assignment $s(j)$ representing the center i to be assigned to the customer j is expressed by the zero-one decision variable z_{ij}. The decision variable z_{ij} takes the value of one, if $i = s(j)$.

Using the variables y_i and z_{ij} and the substitution $c_{ij} = (ed_{ij} + g_i)b_j$, expressions (12.1), (12.2), and (12.3) can be rewritten as expressions (12.4), (12.5), and (12.6), respectively:

$$\sum_{i\in I} f_i y_i + \sum_{i\in I}\sum_{j\in J} c_{ij} z_{ij} \tag{12.4}$$

$$\sum_{i\in I}\sum_{j\in J} u_{ij} z_{ij} \tag{12.5}$$

$$\max\left\{\sum_{i\in I} u_{ij} z_{ij} : j \in J\right\} \tag{12.6}$$

The non-linearity of the expression (12.6) can be removed by adding a new variable which is minimized and by adding a set of constraints ensuring that this variable takes the values greater or equal to the discomfort perceived by each customer.

In this subsection, we revised the basic forms of the objective functions used when designing public service systems. For more systematic overviews we refer the reader to the references [6, 16, 22].

12.2.1 Cost-Oriented Service System Design

The cost-oriented design leads to the system, which minimizes the system cost, assuring a certain level of the welfare. The structure of the resulting system is described by the above-introduced variables y_i, which determine the nodes, at which the service centers are located. The quality criterion corresponds, for example, to the expression (12.4). In addition, some consistency constraints must be now imposed on the decision variables. With each customer j, only a single allocation variable

z_{ij} is allowed to take value of one, and, furthermore, if the variable z_{ij} takes the value of one, then the associated variable y_i must take the value of one as well. If no other restriction on the customers' welfare is put except being served from a service center, then the model of the cost-oriented system design can be stated as follows:

$$\text{Minimize} \sum_{i\in I} f_i y_i + \sum_{i\in I}\sum_{j\in J} c_{ij} z_{ij} \tag{12.7}$$

$$\text{subject to} \sum_{i\in I} z_{ij} = 1 \quad \text{for } j \in J \tag{12.8}$$

$$z_{ij} \le y_i \quad \text{for } i \in I,\ j \in J \tag{12.9}$$

$$y_i,\ z_{ij} \in \{0,\ 1\} \quad \text{for } i \in I,\ j \in J \tag{12.10}$$

The constraints (12.8) ensure that each municipality (customer) is assigned to exactly one location. Whenever the customer j is assigned to the location i, the constraints (12.9) (so called linking constraints) force the variable y_i to take the value of one. This problem, introduced in [2], is known as the uncapacitated facility location problem (UFLP), or the simple plant location problem. This problem is NP-hard [21] and its properties and solving techniques are broadly discussed, for example, in [3, 32].

If a certain level of welfare is supposed to be preserved, the model (12.7) – (12.10) is supplemented with the conditions that keep the discomfort described by the expression (12.6) below a given threshold $U_{\max}$. Thus, for this purpose, either the set of constraints (12.11) or (12.12) is added to the model (12.7)–(12.10).

$$\sum_{i\in I} u_{ij} z_{ij} \le U_{\max} \quad \text{for } j \in J \tag{12.11}$$

$$u_{ij} z_{ij} \le U_{\max} \quad \text{for } i \in I,\ j \in J \tag{12.12}$$

Note that the extended model can be easily turned into the original uncapacitated facility location problem by redefining the cost coefficients c_{ij}. It can be done for example by setting c_{ij} to sufficiently high value, whenever the inequality $u_{ij} \le U_{\max}$ does not hold. In case the total costs of the designed system do not include the costs c_{ij}, then much simpler model can be formulated instead. As the relation between the customer j and the center location i becomes irrelevant, with the exception of the constraints (12.12), we can even avoid using the allocation variables z_{ij}. We define the coefficients a_{ij} so that a_{ij} takes the value of one if and only if the inequality $u_{ij} \le U_{\max}$ holds, and otherwise it takes the value of zero. We use the zero-one location variables y_i as before, and formulate the following model:

$$\text{Minimize} \sum_{i\in I} f_i y_i \tag{12.13}$$

$$\text{subject to} \sum_{i\in I} a_{ij} y_i \ge 1 \quad \text{for } j \in J \tag{12.14}$$

$$y_i \in \{0,\ 1\} \quad \text{for } i \in I \tag{12.15}$$

The constraints (12.14) are satisfied, if at least one service center is located in the neighborhood of the customer j, so that the discomfort constraint $u_{ij} \leq U_{\max}$ is met. The problem (12.13)–(12.15), known as the set covering problem, was for the first time formulated in [41]. The problem is NP-hard [29] and its solving algorithms are over-viewed, for instance, in [11].

12.2.2 Customer's Welfare-Oriented Service System Design

The welfare-oriented design aims at minimizing the customer's discomfort, provided the system costs do not exceed the given level $C_{\max}$. Also, in this case, the structure of the designed system is described by the location variables y_i and allocation variables z_{ij}. The quality criterion expressing the average customer's welfare could correspond to the expression (12.5). The consistency constraints imposed on the decision variables y_i and z_{ij} are the same as before. A simple model illustrating the system design oriented at the average customer's welfare can be stated as follows:

$$\text{Minimize} \sum_{i\in I}\sum_{j\in J} u_{ij} z_{ij} \tag{12.16}$$

$$\text{subject to} \sum_{i\in I} z_{ij} = 1 \quad \text{for } j \in J \tag{12.17}$$

$$z_{ij} \leq y_i \quad \text{for } i \in I,\ j \in J \tag{12.18}$$

$$\sum_{i\in I} f_i y_i + \sum_{i\in I}\sum_{j\in J} c_{ij} z_{ij} \leq C_{\max} \tag{12.19}$$

$$y_i, z_{ij} \in \{0,\ 1\} \quad \text{for } i \in I,\ j \in J \tag{12.20}$$

This problem can be solved directly, using a universal optimization tool. However, if the sets I and J are too large, then the problem can be rearranged as the uncapacitated facility location problem, using the Lagrangian relaxation. If the positive Lagrangian multiplier λ is introduced, and we apply the relaxation to the constraint (12.19), we obtain a new objective function:

$$\begin{aligned}
&\sum_{i\in I}\sum_{j\in J} u_{ij} z_{ij} + \lambda\Big(\sum_{i\in I} f_i y_i + \sum_{i\in I}\sum_{j\in J} c_{ij} z_{ij} - C_{\max}\Big) \\
&\quad = \sum_{i\in I} \lambda f_i y_i + \sum_{i\in I}\sum_{j\in J} (u_{ij} + \lambda c_{ij}) z_{ij} - \lambda C_{\max}
\end{aligned} \tag{12.21}$$

For the given value of λ we can minimize the expression (12.21) subject to the constraints (12.17), (12.18), and (12.20) as an instance of the uncapacitated facility location problem. Then the optimal or the near-optimal solution of the original problem (12.16)–(12.20) can be obtained by iterative process, where the bisection method is used to find such value of λ that the resulting solution $(\mathbf{y}, \mathbf{z})$ fulfills the constraint (12.19) as the equality with an arbitrary precision.

The same approach can be applied when the total costs do not include the costs c_{ij}. Thus, the constraint (12.19) is replaced by simpler constraint (12.22). If the parameter p limits the number of the located service centers instead of constraining the costs, then the constraint (12.19) can be replaced by the constraint (12.23).

$$\sum_{i \in I} f_i y_i \leq C_{\max} \tag{12.22}$$

$$\sum_{i \in I} y_i \leq p \tag{12.23}$$

If the discomfort is expressed as $u_{ij} = b_j$ for $d_{ij} > D_{\max}$ and $u_{ij} = 0$ otherwise, then the instance of the set covering model can be formulated. The coefficients a_{ij} are defined as above, i.e. a_{ij} takes the value of one, if and only if the inequality $d_{ij} \leq D_{\max}$ holds, otherwise it takes the value of zero. Also, the location variables y_i are defined as before and, in addition, the auxiliary variables x_j are introduced for each customer $j \in J$. It is expected that the variable x_j takes the value of one, if there is no located service center within the radius $D_{\max}$ from the customer j. Then the following set covering model describes the welfare-based system design problem:

$$\text{Minimize} \sum_{j \in J} b_j x_j \tag{12.24}$$

$$\text{subject to} \sum_{i \in I} a_{ij} y_i \geq 1 - x_j \quad \text{for } j \in J \tag{12.25}$$

$$\sum_{i \in I} f_i y_i \leq C_{\max} \tag{12.26}$$

$$y_i \in \{0, 1\} \quad \text{for } i \in I \tag{12.27}$$

$$x_j \in \{0, 1\} \quad \text{for } j \in J \tag{12.28}$$

The term (12.24) expresses the overall volume of uncovered demands which is minimized. The constraints (12.25) ensure that the variables x_j take the value of one, if and only if there is no service center located within the radius $D_{\max}$ from the customers location j. The constraint (12.26) puts the limit $C_{\max}$ on the system costs. More complicated situation arises if the welfare-oriented system is designed and the discomfort of the worst positioned customer is used as the quality criterion. The model of this type assumes the following form:

$$\text{Minimize } h \tag{12.29}$$

$$\text{subject to} \sum_{i \in I} u_{ij} z_{ij} \leq h \quad \text{for } j \in J \tag{12.30}$$

$$\sum_{i \in I} z_{ij} = 1 \quad \text{for } j \in J \tag{12.31}$$

$$z_{ij} \leq y_i \quad \text{for } i \in I, j \in J \tag{12.32}$$

$$\sum_{i \in I} f_i y_i + \sum_{i \in I} \sum_{j \in J} c_{ij} z_{ij} \leq C_{\max} \tag{12.33}$$

$$y_i, z_{ij} \in \{0,\ 1\} \quad \text{for } i \in I, j \in J \tag{12.34}$$

The transformation of the model (12.29)–(12.34) either into the uncapacitated facility location problem or into the set covering problem is possible, although only under very strong and simplifying assumptions on both customers' discomforts and cost coefficients.

The family of the above reported models can comprise more complex service system design problems than we have explained. Thus, it is often necessary to add some additional constraints which can capture restrictions such as limited capacity of service centers, capacity and time constraints on the communication links (roads) etc. Another issue is how to find a reasonable trade-off between various objectives. For example, what should be preferably minimized the total discomfort of all customers or the worst individual discomfort? As shown above, each of these criteria leads to a specific objective function. The criterion of the total discomfort is described by the term (12.16), and the criterion of the worst situated customer discomfort is described by the term (12.29). To a certain extent, these two approaches can be combined in one optimization process as the following example shows. Let us introduce the real valued threshold $U_{\max}$, and define the penalty P as

$$P = \sum_{j \in J} \max\{u_{ij} : i \in I\}. \tag{12.35}$$

Then we define the new disutility $\underline{u_{ij}}$ so that $\underline{u_{ij}} = u_{ij}$ if $\underline{u_{ij}} \leq U_{\max}$ and $u_{ij} = P$ otherwise. The optimal solution of the problem described by (12.16)–(12.20) with the new coefficients $\underline{u_{ij}}$ does not allow to assign a customer to a service center, if it should cause bigger discomfort than $U_{\max}$, under the assumption that such a solution exists. Using this construction, an optimization process can be formulated. The process starts with a big value of the threshold $U_{\max}$, and repeatedly solves the problem (12.16)–(12.20) for smaller and smaller $U_{\max}$ until the value of (12.16) exceeds penalty P. The last but one solution is a good compromise between these two approaches. Another option is to combine different objectives into one model whereby a multi-objective problem is created.

12.3 Solving the Uncapacitated Facility Location Problem to Optimality

As we have demonstrated in the previous section, the UFLP can be often either directly or indirectly used to tackle the public service system design problem. For this reason we describe in this section the methods enabling to solve this problem efficiently. There are usually two options of solving the optimization problem derived from the real-world situation. We might either choose a specialized algorithm restricted to the problem, which almost always requires the implementation of the algorithm using a convenient programming language, or, if the problem can be seen as the instance of a standard class of optimization problems, we can use a universal optimization tool.

12.3.1 Universal Optimization Tools

A universal optimization tool is a software package designed to solve standard classes of optimization problems, such as linear programming problems, linear integer programming problems or quadratic programming problems. Typically, such a tool consists of a modeler which allows to separate the mathematical model from the data, describing the particular instance of the optimization problem. Thus, the formulation of the mathematical model is independent on parameter values, which makes it easier to maintain both, the model and the data. Nowadays, the most popular tools are CPLEX (http://www.cplex.com), XPRESS-MP (http://www.fico.com) or MOSEK (http://www.mosek.com).

Modern optimization packages offer programming languages to simplify the work with mathematical models, debugging environments and interfaces for graphical output. Moreover, they can be embedded into development environments, such as AIMMS (http://www.aimms.com). To supplement the results provided by specialized algorithms, presented in the next section, we compare them with the results obtained by XPRESS-MP.

12.3.2 Specialized Algorithms

As we have already mentioned in Sect. 12.2, the uncapacitated facility location problem is broadly applicable in the design of public service systems. The corresponding solving technique can be used not only to design the cost-optimal two-levels distribution system [24] but it can be extended in order to solve more complex location problems. As it was discussed in the reference [20], it is possible to turn the maximum distance problem, the maximum covering problem and the p-median problem into the form of the UFLP. Also, the capacitated version of the UFLP or the p-center problem can be approximatively solved using the solving algorithms for UFLP [26].

Many scholars have dealt with UFLP [14]. Nevertheless, as far as the exact algorithms are concerned, the now seminal procedure DualLoc proposed by Donald Erlenkotter [17] is still one of the most efficient methods, and it enables to find an optimal solution in tractable computational times [12]. Inspired by this approach in [30], Manfred Koerkel proposed several successful modifications, which speed up the solving process and the resulting algorithm was named PDLoc.

In [25], we have shown how both algorithms, DualLoc and PDLoc, can be accelerated by modifying the famous procedure of dual ascent (DA) introduced in [5]. We implemented both algorithms using the integrated programming development environment Delphi. In our implementation, we restricted the values of coefficients f_i and c_{ij} to integers. For historical reasons, from now on we will refer to our implementation of the DualLoc algorithm as to the BBDual algorithm [23].

12.3.2.1 Algorithm BBDual

If the variables y_i are known, the optimal values of the variables z_{ij} can be found easily. It is sufficient to assign the customer j to the facility i, for which the value of coefficient c_{ij} is minimal. Thus, the most difficult problem is to determine the setting of the variables y_i. The BBDual algorithm is based on the branch and bound method, in which two subproblems emerge by fixing the variables y_i either to zeros or ones. The algorithm uses of the depth-first strategy. To decide if a given subproblem should be processed or excluded from the searching process, a lower bound of high quality is needed. Such lower bound can be obtained by successively performing the dual ascent (DA) and the dual adjustment algorithms (DAD). The former, the DA procedure, starts from an arbitrary feasible solution of the dual problem, and it subsequently increases the value of its objective function. The latter procedure enables a further improvement of the dual solution by searching for a configuration in which a small decrease of the objective function will allow larger increase. These two procedures provide dual feasible solution and the corresponding value of the objective function serves as the lower bound. Furthermore, a corresponding primal feasible solution is generated. This is done by the PRIMA procedure, which follows the complementary conditions holding between an LP relaxation of the problem (12.8)–(12.10) and its dual. The best-found primal solution is stored and its objective function value constitutes the upper bound for the optimal solution.

12.3.2.2 Algorithm PDLoc

Algorithm PDLoc comprehends a number of effective modifications and improvements of procedures originally proposed by Erlekotter, and in addition, it is enhanced by several new procedures. Similarly to the BBDual algorithm, the PDLoc employs the branch and bound method to determine the optimal solution, but in contrast to the BBDual, the strategy of the lowest lower bound is used to process the searching tree. Varying the order of customers in the PRIMA procedure enables to open new locations, and to explore a broader spectrum of primal solutions. This leads to faster decrease of the upper bound and to faster termination of the searching process.

The number of steps in the incremental build-up process that is used to construct a dual solution by the DA procedure depends on the gap between the f_i and c_{ij} values. Therefore, more rapid incrementation leads to considerable improvements in the cases when the fixed charges f_i are considerably higher than the allocation costs c_{ij} are. This is ensured by using the dual multi-ascent procedure (MDA) instead of the DA procedure.

Another improvement results from applying the simple exchange heuristic right after the first primal solution is found. Moreover, if a large difference between the upper and the lower bound occurs, it is reduced by the modified dual adjustment procedure. This procedure consists of two phases. The first of them is called the primal-dual multi adjustment (PDMAdj) and the second is the primal-dual adjustment (PDAdj). In a loop, both procedures alternatively decrease and increase the reached lower bound in order to find a combination of operations which would allow to increase the lower bound. The difference between those two is in the scheme that is used to decide which and how many variables are modified in one step.

The used branch and bound searching scheme allows to fix the selected locations to be permanently open ($y_i = 1$) or closed ($y_i = 0$) and thereby to reduce the size of the solved problem. To fix a variable, special conditions have to be met. The evaluation of the conditions is time consuming, especially, if the searching process is nested deeply in the searching tree. Therefore, fixing variables is preferably done on the top of the searching tree (pre-processing). If the processed branch is far down from the root, the variables are fixed only if there is a chance to fix a couple of them simultaneously.

12.3.2.3 Benchmarks

Benchmarks that we used to test our implementation of algorithms were derived from the real-world network. The set G700 consists of 700 problems derived from the road network of the Slovak Republic. This set includes ten subgroups the sizes of which range from $100 \times 2{,}906$, $200 \times 2{,}906$, to as large as $1{,}000 \times 2{,}906$. The first number is the number of candidates for a facility location ($|I|$), and the second number refers to the size of the customer set ($|J|$). Each subgroup contains 70 benchmarks. For each size of the benchmark ten different random subgraphs of the road network graph of corresponding size were generated. Each subgraph was used as a base for creating seven benchmarks by modifying the coefficients c_{ij} and f_i to cover evenly the whole spectrum of the centers located by the optimal solution.

For instance, for the problem of size $100 \times 2{,}906$ the optimal numbers of located facilities were 1, 17, 33, 50, 66, 83 and 100, respectively. We will provide the source code of the algorithms upon request. Our benchmarks were uploaded onto the supplementary Web page http://frdsa.uniza.sk/~buzna/supplement.

12.3.2.4 Preliminary Experiments

We solved all problems to get the frequency in which the particular procedures are executed. These numerical experiments were performed on a PC equipped with Intel 2.4 GHz processor and 256 MB RAM. The average computational time distributed among the inner procedures is shown in Fig. 12.1.

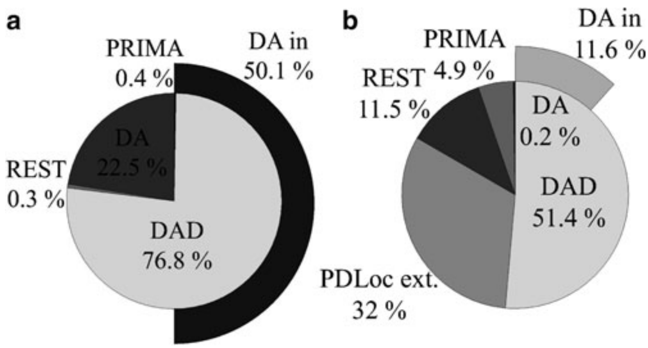

Fig. 12.1 The average distribution of computational time among the procedures of the BBDual algorithm in (**a**) and the PDLoc algorithm in (**b**) (these results were obtained for the set G700)

The abbreviations "DA," "DAD," and "PRIMA" denote the relative average time taken by the corresponding procedure. The "REST" includes the time consumed by the branch and bound method, lower bound computation, as well as necessary memory operations. "PDLOC ext." represents the time spent on running the procedures which are exclusively included only in the PDLOC algorithm. The label "DA in" stands for the time taken by the procedure DA, which was called from other procedures (e.g. the DA procedure is called inside the DAD procedure). This also explains why we plotted this value outside the pie graph. The results clearly show that the BBDual algorithm took in average 72.6% by performing the DA procedure, while the PDLOC algorithm devoted only 11.8% of the computational time to this procedure. The time consumed by the DAD procedure is approximately equal with both algorithms, although its distribution among the time consuming activities differs considerably.

In the case of the BBDual algorithm, the DAD procedure took 50.1% of the time. On the contrary, the PDLoc algorithm needs only a small portion of the time to perform the DA procedure nested in the DAD procedure. This implies that the PDLoc algorithm focuses more on intensive searching for improving operations. More detailed comparison of the computational performance can be found in Table 12.1.

12.3.2.5 Amendments of Algorithms BBDual and PDLoc

Amendments of algorithms BBDual and PDLoc were inspired by preliminary experiments, which showed that the DA procedure consumes large portion of the computational time. In order to describe the procedure DA, we need to consider the following form of the dual problem [30], derived from the LP relaxation of the primal problem (12.7)–(12.10).

$$\text{Maximize } z_D = \sum_{j \in J} v_j \tag{12.36}$$

Table 12.1 Average time in seconds and corresponding standard deviation obtained for benchmarks G700

	BBDual		BBDual*		PDLoc		PDLoc*	
Size of problems	t[s]	Std D	t[s]	Std D	t[s]	Std D	t[s]	Std D
100 × 2,906	5.343	12.64	0.28	0.29	1.44	0.93	0.77	0.34
200 × 2,906	27.16	68.33	0.41	0.39	1.80	0.99	0.87	0.22
300 × 2,906	52.95	143.11	0.74	0.64	2.40	1.48	1.20	0.90
400 × 2,906	127.06	337.58	1.01	0.47	2.74	1.82	1.46	0.88
500 × 2,906	134.17	340.52	1.75	1.05	5.29	6.52	2.83	0.90
600 × 2,906	277.59	700.73	2.54	1.78	5.21	5.54	3.64	2.88
700 × 2,906	277.70	704.57	3.90	2.77	6.12	5.45	4.63	4.38
800 × 2,906	497.26	1,248.42	5.07	4.23	8.56	8.87	6.45	6.35
900 × 2,906	640.44	1,652.65	7.24	6.31	11.45	11.25	8.89	8.83
1,000 × 2,906	644.88	1,595.72	7.07	5.89	10.60	11.19	7.47	8.08

Procedure DA(J^+)
While $|J^+| > 0$ do:
    Get next $j \in J^+$.
    Set $d_1 = min_i(s_i)$ with $i \in \{i: c_{ij} \leq v_j\}$.
    Set $d_2 = min_i(c_{ij} - v_j)$ with $i \in \{i: c_{ij} > v_j\}$.
    If $d_1 > d_2$, set $d = d_2$. Else, delete j from J^+ and set $d = d_1$.
    If $d > 0$, then:
        For each $i \in \{i: c_{ij} \leq v_j\}$, set $s_i = s_i - d$.
        Set $v_j = v_j + d$ and $z_D = z_D + d$.
Terminate.

Fig. 12.2 The original procedure DA as it was described in [17] or in [30]

$$\text{subject to} \sum_{j \in J} \max\{0, v_j - c_{ij}\} + s_i = f_i, \qquad \text{for } i \in I. \tag{12.37}$$

$$s_i \geq 0 \qquad \text{for } i \in I. \tag{12.38}$$

The original DA procedure (see Fig. 12.2) solves the problem (12.36)–(12.38) by processing the set of relevant customers J^+, customer by customer, in the order which follows the unordered input sequence. At each step the variable v_j corresponding to the customer $j \in J^+$ is incremented by the value d, whereas the value d is determined as the maximal value, which satisfies the constraints (12.6). Hence, it becomes obvious, that this variable cannot be increased in the followings steps, and the customer j can be excluded from the set J^+. This procedure is repeated until J^+ is emptied.

As we will demonstrate, the performance of this procedure depends on the ordering in which the set of relevant customers J^+ is processed. This drawback is illustrated by the example with two possible locations, i_1 and i_2, and three customers j_1, j_2 and j_3, as it is depicted in Fig. 12.3. The locations i_1 and i_2 are associated with two slack variables, s_{i1} and s_{i2}, respectively. The edge connecting the customer j

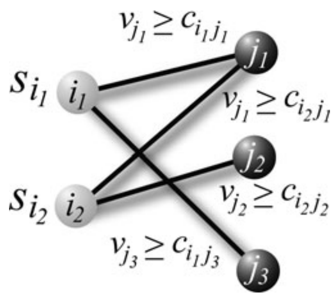

Fig. 12.3 The scenario where the ordering of customers improves the efficiency of the DA procedure. *White-colored* nodes represent the candidates for the facility location and *black* nodes are customers. The variables s_{i1} and s_{i2} are slack variables (see the constraints (12.37)) and v_j are dual variables corresponding to customers. The location i is connected with the customer j by the edge only if the inequality $v_j \geq c_{ij}$ holds

Procedure DA*(J⁺)

Order the set J^* of relevant customers into the sequence $j_1, j_2, \ldots, j_k, \ldots, j_n$ ascendingly with respect to the cardinalities K_j.
For $k = 1$ to $|J^+|$ do:
  Set $j = j_k$.
  $d = \min\{\min\{s_i : i \in I, c_{ij} \leq v_j\}, \{c_{ij} - v_j : i \in I, c_{ij} > v_j\}\}$.
  If $d > 0$, then:
    Update s_i for $i \in I, c_{ij} \leq v_j$ by $s_i := s_i - d$. Set $v_j := v_j + d$ and $z_D = z_D + d$.
    If the cardinality of K_j has increased, then reorder the sequence of customers $j_k, \ldots, j_n$.
Terminate.

Fig. 12.4 Modified DA* procedure

with the location i symbolizes that the inequality $v_j \geq c_{ij}$ holds. By the symbol K_j we denote the cardinality of the set $\{i \in I : c_{ij} \leq v_j\}$ for the customer j. From Fig. 12.3 we get $K_{j_1} = 2$, $K_{j_2} = 1$ and $K_{j_3} = 1$.

The procedure DA (see Fig. 12.2) processes the set of customers J^+ in the order which is given by the sequence in which they enter the procedure. Thus, the first to be processed is the customer j_1 followed by j_2 and j_3. If the processing of the customer j_1 enables to increase the variable v_{j_1} by a value β, then the lower bound z_D is increased exactly by β. The maximal theoretical increase of the lower bound z_D is thus given by the sum of variables s_{i1} and s_{i2}. In the example from Fig. 12.3, the increment β has to be subtracted from both slack variables s_{i1} and s_{i2}, in order to meet the constraints (12.37). In summary, the theoretical capacity $s_{i1} + s_{i2}$ is reduced by 2β, in order to increase the lower bound z_D by β.

Considering the theoretical capacity and its possible decrease by modifying the variables v_j, we formulate a new DA* procedure [25] (see Fig. 12.4). This procedure exploits better the potential of increasing the lower bound. This approach is based on prioritizing those customers, which preserve better chances for the increase of the lower bound z_D in the following steps. We order the relevant customers J^+ ascendingly according to the cardinalities K_j. The benefit of this modification is demonstrated by the example in Fig. 12.3.

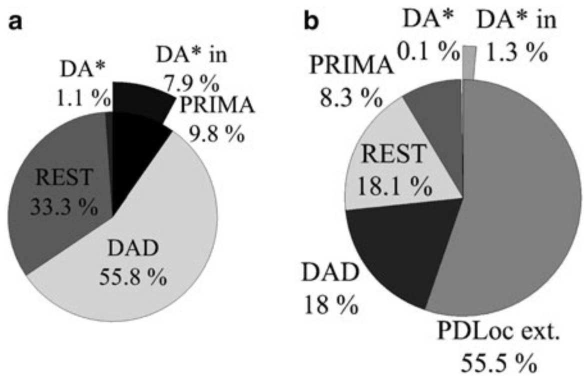

Fig. 12.5 The average distribution of computational time among the procedures of the BBDual* algorithm in (**a**) and the PDLoc* algorithm in (**b**) (these results were obtained for the set G700)

Having applied the above mentioned ordering of customers, they will be processed in the order j_2, j_3 and j_1. If processing the customer j_2 enables to increase the variable v_{j_2} by β, then, considering the constraints (12.37), only slack variable s_{i2} has to be reduced by β. This way the theoretical capacity is reduced by the value β, and the lower bound z_D increases by β. Compared to the unordered case, the ordering may reduce the sum of slack variables s_i less than the original DA procedure does, while preserving the same growth rate for the variables v_j.

12.3.2.6 Verification of the Proposed Amendments

Verification of the proposed amendments was performed with the same set of benchmarks as described previously. We applied the new DA* procedure to both algorithms. We inserted the MDA procedure into the BBDual algorithm, as it had proved itself to be very efficient with the PDLoc algorithm when dealing with the cases in which the costs f_i are considerably higher than the costs c_{ij}. We also amended the evaluation of subproblems in the BBDual algorithm. Both incoming subproblems are solved simultaneously, and the most perspective subproblem is processed as the first. We would like to point out that we have tested all these modifications separately [25]. However, none of them brought any remarkable improvements compared to when applied together. To distinguish the original and the new versions of the algorithms BBDual and PDLoc, the modified version is denoted with the superscript "*."

The effects of the proposed modifications were extensively examined by numerical experiments [25]. Similarly, as in the preliminary experiments, we evaluated the average computational time and its distribution among the inner procedures. Figure 12.5 gives the evidence of the significant change in the time distribution of the inner procedures. The total usage of the DA procedure was reduced from

72.6% to 9% for the BBDual* algorithm and from 11.8% to 1.4% for the PDLoc* algorithm. These results also suggest that the new DA* procedure has a significant influence on the performance of both algorithms. Table 12.1 compares the results achieved by the original and modified versions of the algorithms. The results indicate that the proposed modification can save time considerably. Please note that using the MDA procedure in the BBDual algorithm contributed to this significant reduction, especially for the benchmarks where $f_i \gg c_{ij}$. However, the improvement brought by the new DA* is perceptible in the whole range of the parameter values. More details on this computational study can be found in the reference [25].

12.4 Case Study

In this section, we show how the choice of the particular criteria can influence the resulting design of the public service system. As an illustration example we use the system of EMS. More precisely, we present the optimized location of ambulance stations in the area of the Slovak Republic. Another purpose of the following text is to introduce improving modifications of the mathematical model resulting in better quality of the final design. When we were carrying out this study, the deployment of EMS stations was defined by the regulations of the Ministry of Health of the Slovak Republic Nos. 30/2006 and 365/2006. In accordance with these regulations, there were 264 EMS stations located in 223 cities and villages, including urban districts of the capital Bratislava and of the second largest Slovakian city Košice (see Fig. 12.6). They altogether served 2,916 municipalities populated with 5,379,455 inhabitants.

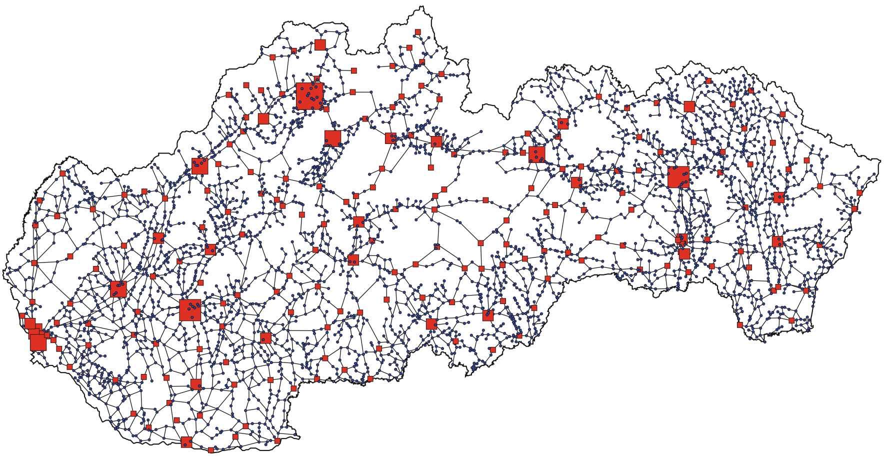

Fig. 12.6 Road network of the Slovak Republic. Locations of the existing EMS stations are marked with *squares*. The size of the *squares* is proportional to the number of the stations located in the given municipality

We evaluate four different criteria:

$$C1: \sum_{j \in J} b_j t_{s(j),j}, \tag{12.39}$$

expressing the total traveling time from all customers to the closest EMS location weighted by the number of inhabitants b_j,

$$C2: \sum_{\substack{j \in J \\ t_{s(j),j} > T_{\max}}} b_j (t_{s(j),j} - T_{\max}), \tag{12.40}$$

summing up the travel time exceeding the threshold value $T_{\max}$ weighted by the number of inhabitants b_j,

$$C3: \sum_{\substack{j \in J \\ t_{s(j),j} > T_{\max}}} b_j, \tag{12.41}$$

representing the number of inhabitants which are further from the closest EMS than the threshold $T_{\max}$ states, and

$$C4: \sum_{\substack{j \in J \\ t_{m(j),j} > T_{\max}}} b_j, \tag{12.42}$$

expressing the number of inhabitants which are covered by fewer than two EMS stations, where we define $m(j)$ as the index of the second closest EMS station to the customer j.

We combine the criterion $C1$ with the location–allocation type of model (12.7)–(12.10) and the criteria $C3$ and $C4$ with the set covering type of model (12.13)–(12.15). In all three cases we propose the deployment of 264 stations, which is optimal with respect to the particular criteria.

The set I of the candidate stations consists of the existing EMS stations defined by the official regulations, and of the municipalities with at least 300 inhabitants. Altogether we found 2,283 cities and villages meeting the conditions. Following the recommendation of the Ministry of Health to deliver medical care within 15 min from an emergency call, we regard $T_{\max} = 15$ min as the threshold value.

As it was indicated in Sect. 12.2, different criteria lead to different kinds of mathematical models, and therefore we also apply different solution techniques. The Location–allocation model was solved using the modified algorithm BBDual*, described in Sect. 12.3. The size and structure of the covering models are simpler, and they allow to use the universal optimization tool Xpress-MP (http://www.fico.com). The experiments were performed on a personal computer equipped with the Intel Core 26,700 processor with the following parameters: 2.66 GHz and 3 GB RAM. The computational time for the BBDual* algorithm was about 116 min. Xpress-MP managed to solve the covering models in 1.5 s and 0.3 s, respectively.

Table 12.2 Values of criteria calculated for three solutions obtained by the optimization (Location–allocation model, Set covering model, Double covering model) and for the existing design of EMS stations (Man-made)

	Criterion C1	Criterion C2	Criterion C3	Criterion C4	Number of changed EMS stations
Location–allocation model (C1)	13,677,537	35,109	15,613	285,324	104
Set covering model (C3)	23,613,635	180	91	188,937	207
Double covering model (C4)	25,597,594	182	91	11,684	199
Man-made	16,378,985	92,032	31,672	431,415	0

The last column shows the number of changed EMS locations proposed by the optimization compared to the existing design

Table 12.2 summarizes the values of the criteria (12.39)–(12.42) for the optimal solutions obtained by three selected models and for the current deployment. The results for the Location–allocation model with the criterion (12.39) are in the first row, the results for the Set covering model with criterion (12.41) are in the second row, and the results for the Double covering model (12.42) are listed in the third row. The Man-made row in the table corresponds to the current distribution of EMS stations over the area of the Slovak Republic.

Comparing the results of mathematical modeling to the current situation, we can observe that the first model, the Location–allocation model, is able to improve all of the defined criteria. Although this model minimizes the total travel time from ambulance locations to all potential patients, the best improvement is achieved by criterion C3, reflecting the discomfort of uncovered people. It means that the solution obtained by this model is more social, than the current system offers, in terms of customer's equity in the access to the provided service. As expected, the highest level of equity is assured for the set covering models. They were able to locate the ambulances in such a way that almost all municipalities were covered. The only uncovered village was Runina with 91 inhabitants. This small village in the most eastern part of Slovakia has no candidate location within the radius of 15 min.

To achieve a more fair design, the covering models change substantially the current deployment of EMS stations (see the last column in Table 12.2). To implement these solutions would thus require to restructure the existing infrastructure considerably what might be too costly. If this issue is of importance, it can be practical to reduce the number of changed EMS stations by including the following constraint

$$\sum_{i \in I_0} (1 - y_i) \leq r, \tag{12.43}$$

where I_0 denotes the set of current center locations and r is the upper bound for the acceptable number of changes.

Let us look more closely at the results obtained by the Double covering model. In the solution presented in Table 12.2, there are all candidates for EMS stations equivalent regardless their population or the distance from a hospital. As a consequence of such simplification the model deploys, for example, only two stations in the capital Bratislava and none in the county capital Žilina. At first sight, this was not a reasonable design, which was also later confirmed by a computer simulation. To get a more realistic solution, the model needs to be modified. We add constraint $y_i \geq 1$ for the selected cities, to ensure that at least one station is located there. In further experiments, we apply this constraint to those cities that have a hospital.

The existing situation and the solution of the Double covering model were evaluated using computer simulations. The goal of the simulations is to assess how efficiently the system operates in real conditions, and to evaluate the parameters and properties that are not explicitly captured by mathematical models. From a customer's (patient's) point of view, the simulation should give answers to the following questions (evaluated separately for every municipality, as well as for the entire region):

1. What is the real coverage rate, i.e. what is the percentage of the calls processed within the required time limit?
2. What are the average and maximal waiting times for an ambulance arrival?

The analysis combining mathematical optimization with simple Monte Carlo simulations allows to verify the arrangements improving the system performance, such as the number of the ambulances allocated to a station. Simulating the EMS system can be viewed as a queuing system with the Poisson arrival of events and exponentially distributed service time [8]. Supposing that every station is equipped with one ambulance only, the system has as many service lines as the number of the stations is. Random events are the emergency calls that come from the populated areas. Every municipality has a specific arrival rate λ_j (calls per time unit). The statistics describing the number of calls for particular municipalities were not available to us. Therefore we used aggregated statistics for the Slovak Republic mapping the year 2001 (http://www.kezachranka.sk), reporting that the overall number of patients was 232,904. We calculated the rate λ per one inhabitant, and used it to estimate $\lambda_j = \lambda b_j$, where b_j is the number of inhabitants registered in the municipality j. Processing a call requires to take the following steps:

1. Call handling (a dispatcher has to obtain the address, and assess how serious the call is, decides which ambulance to assign to it, and contacts the ambulance crew, then the crew has to reach the vehicle);
2. Driving the ambulance from the station to the patients location;
3. Treating the patient by the ambulance crew;
4. Transporting the patient to the nearest hospital;
5. Passing the patient to the hospital staff;
6. Driving the ambulance back to the station.

Table 12.3 Comparison of the real-world situation from the year 2006 with the deployment of the stations proposed by the mathematical models

	Man-made	Double covering	Double covering with EMS located at hospitals
Number of municipalities with at least one EMS station	223	217	245
Multiply covered inhabitants [%]	91.90	99.77	99.49
Calls not serviced within 15 min [%]	1.22	1.25	0.25
Average waiting time [min]	3.4	5.4	3.9
Maximal waiting time [min]	43.5	30	28.4

When modeling the service time, we neglect the call handling and we assume deterministic travel times (defined by the distance and the average speed). According to [8], the treatment time can be considered as to be exponentially distributed with the mean value 10 min. Hence, as we did not have the relevant statistical data on the EMS operation, we supposed that every patient was transported to the nearest hospital, and the time the ambulance spent waiting at the hospital was constant (10 min).

There are many possible dispatching policies for allocating the calls to EMS stations. We always assign a call to the nearest station. If the nearest ambulance is busy, the used policy differs with each situation. The call either waits until the ambulance returns back to the station or it can be reallocated to the next nearest station with an idle ambulance. A simulation model allows to experiment with dispatching policies and consequently to choose the best one. With the experiments, reported in Table 12.3, we suppose that we know in advance the time when the ambulance is returning back to the station. The call is then allocated to the station by which the patient gets served the earliest.

Table 12.3 compares the characteristics of the EMS system given by the official regulations from the year 2006 (column Man-made), with the results obtained by the Double covering model and with the design proposed by the modified Double covering model with stations located at hospitals. In all three cases there are located 264 stations. In Table 12.3, we can see that the solutions proposed by mathematical modeling increase the occurrence of multiply covered inhabitants from 91.9% to almost 100%. The last three rows list the performance characteristics of the system obtained by the computer simulations.

Comparing the solution of the Double covering model with the current deployment shows that the rate of calls not serviced within 15 min remains almost identical (approximately 1.2%). The average waiting time for an ambulance to arrive increases (by 59%), while the maximal waiting time decreases (by 31%). Comparing the solution of the modified Double covering model (with stations located at hospitals) with the current deployment we can observe significant improvement in both quantities: The rate of calls not serviced within 15 min declines to one fifth, and the maximal waiting time is shortened by one third.

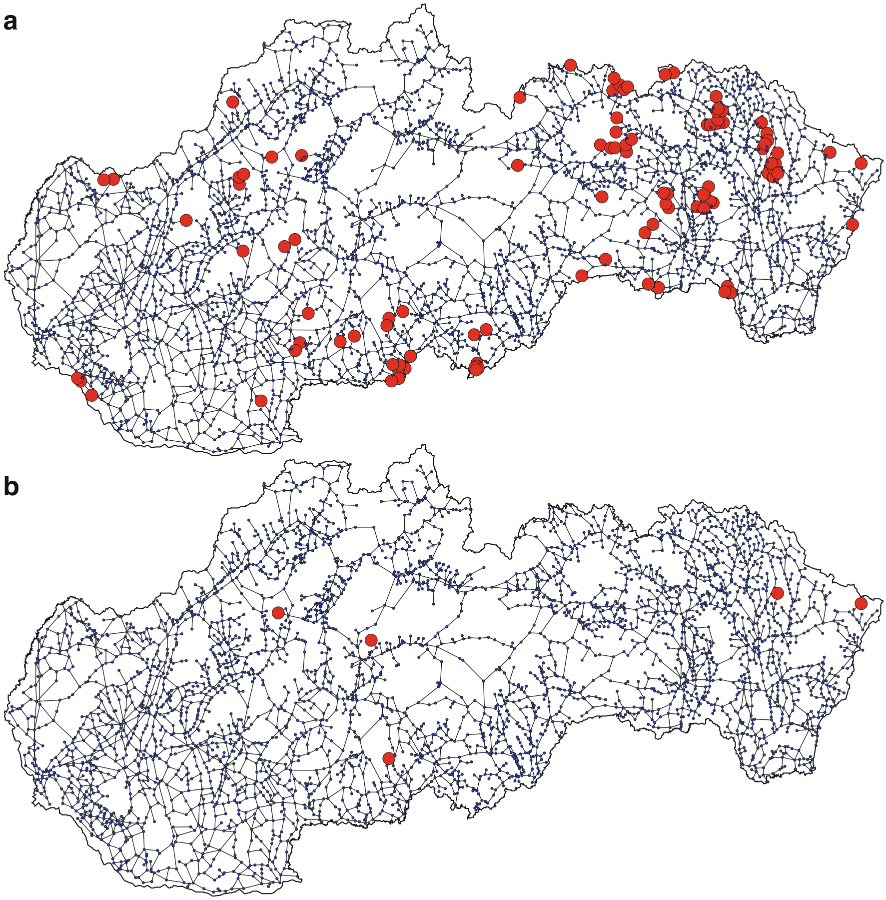

Fig. 12.7 Municipalities with the average waiting time for an ambulance arrival longer than 15 min. In (**a**) we show the situation for the existing EMS stations in the year 2006, and (**b**) depicts the results we reached when we performed Monte Carlo simulations for the solution obtained by the Double covering model imposing the stations to be located close to hospitals

Although the average waiting time remains the same, Fig. 12.7a, b confirm that the number of municipalities where the arrival time exceeds 15 min decreases dramatically.

In conclusion, the statistics presented in Table 12.3 indicate that the design proposed by mathematical modeling is better than the current solution with respect to the service availability to patients. The rate of calls not serviced within 15 min declines to one fifth, the maximal waiting time for an ambulance arrival shortens by one third. The number of municipalities with the average waiting time longer than 15 min falls from 102 to 5.

12.5 Conclusions

In this chapter, we have discussed some issues arising when designing public service systems on complex networks. We focused mainly on effective solving methods, and we searched for a desirable compromise between fairness and economic efficiency, which can be attributed to the resulting design. As possible directions for further research we consider:

- The analysis of the behavior with focus on customers' preferences based on real-world data, and utilization of the results in the modeling of public service systems,
- More detailed theoretical analysis of the price of fairness in the context of location problems,
- and last but not least, possible utilization of centrality measures and other network properties [7] as early indicators of promising solutions in the design of optimization algorithms.

Acknowledgements The authors are grateful for the financial support provided by the Ministry of Education of the Slovak Republic (project VEGA 1/0361/10) and thank J. Slavík, P. Tarábek, M. Koháni, and two anonymous reviewers for thorough reading of the manuscript and the valuable comments.

References

1. Ahuja, R.K., Magnanti, T., Orlin, J.: Network flows: Theory, algorithms, and applications, Prentice Hall, (1993).
2. Balinski, M.: Integer programming: methods, uses computation, Management Science, 12, pp. 254–313, (1965).
3. Beltran-Royo, C., Vial J.P., Alonso-Ayuso, A.: Semi-Lagrangian relaxation applied to the uncapacitated facility location problem, Computational Optimization and Applications, Online First, DOI: 10.1007/s10589-010-9338-2, (2010).
4. Bertsimas, D., Farias, V.F., Trichakis, N.: The price of fairness, Operations Research, (forthcoming).
5. Bilde, O., Krarup, J.: Sharp lower bounds and efficient algorithms for the simple plant location problem. Annals of Discrete Mathematics, 1, pp. 77–97, (1977).
6. Brandeau, M.L., Chiu, S.S., Kumar, S., Grossmann, T.A.: Location with Market Externalities, In: Z., Drezner, ed., Facility Location: A Survey of Applications and Methods, Springer Verlag, New York, (1995).
7. Brandes, U., Erlenbach, T.: Network Analysis: Methodological Foundations, Springer-Verlag Berlin Heidelberg, (2005).
8. Budge, S., Erkut, E., Ingolfsson, A.: Optimal Ambulance Location with Random Delays and Travel Times. Working paper, University of Alberta School of Business, (2005).
9. Butler, M., Wiliams, H.P.: Fairness versus efficiency in charging for the use of common facilities. Journal of Operational Research Society, 53, pp. 1324–1329, (2002).
10. Butler, M., Wiliams, H., P.: The allocation of shared fixed costs, European Journal of Operational Research 170, pp. 391–397, (2006).

11. Caprara A., Toth, P., Fischetti, M.: Algorithms for the set covering problem, Annals of Operations Research 98, pp. 353–371, (2000).
12. Crainic, T.G.: Long-haul freight transportation. In Handbook of transportation science, New York: Springer, (2003).
13. Daganzo, C.F.: Logistics System Analysis, Springer Verlag Berlin, (1996).
14. Daskin, M.S.: Network and Discrete Location. Models, Algorithms, and Applications. John Wiley & Sons, New York, NY, (1995).
15. Dorogovtsev, S., Mendes, J.: Evolution of networks: From biological nets to the Internet and WWW, Oxford University Press, Oxford (2003).
16. Eiselt, H.A., Laporte, G.: Objectives in Location Problems, In: Drezner, Z., ed., Facility Location: A Survey of Applications and Methods, Springer Verlag, New York, (1995).
17. Erlenkotter, D.: A dual-based procedure for uncapacitated facility location. Operations Research, **26**, 6, pp. 992–1009, (1978).
18. Frank, H., Frisch, I.T.: Communication, Transmission and Transportation Networks, Addison Wesley, Series in Electrical Engineering, (1971).
19. Galvao, R.D., Nascimento, E.M.: The location of benefit-distributing facilities in the Brazilian social security system. Operational Research'90 (Proceedings of the IFORS 1990 Conference), pp. 433–443, (1990).
20. Galvao, R.D.: Uncapacitated facility location problems: contributions, Pesqui. Oper. 2004, **24**, 1, pp. 7–38 . ISSN 0101–7438, (2004).
21. Garey, M.R., Johnson, D.S.: Computers and Intractability: A guide to the theory of NP Completeness, San Francisco, Freeman and Co., (1979).
22. Hakimi, S.L.: Optimum location of switching centers and the absolute centers and medians of a graph, Operations Research, 12, pp. 450–459, (1964).
23. Janáček, J., Kováčiková, J.: Exact solution techniques for large location problems. In Proceedings of the International Conference: Mathematical Methods in Economics, pp. 80–84, Ostrava, (1997).
24. Janáček, J., Buzna, Ľ.: A comparison of continuous approximation with mathematical programming approach to location problems, Central European Journal of Operation Research, **12**, 3, pp. 295–305, (2004).
25. Janáček, J., Buzna, Ľ.: An acceleration of Erlenkotter-Koerkels algorithms for the uncapacitated facility location problem. Ann. Oper. Res. **164**, pp. 97–109, (2008).
26. Janáček, J., Gábrišová, L.: A two-phase method for the capacitated facility problem of compact customer subsets. Transport, **24**, 4, pp. 274–282, (2009).
27. J. Janáček, B., Linda, E., Ritschelová: Optimization of municipalities with extended competence selection, Prague economic papers: quarterly journal of economic theory and policy, **19**, 1 pp. 21–34, (2010).
28. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.L.: The large-scale organization of metabolic networks, Nature, **407**, (2000).
29. Karp, R.M.: Reducibility among Combinatorial Problems: In Complexity of Computer Computations, Miller, E.W., Tatcher J.W. eds., New York, Plenum Press, pp. 85–104, (1972)
30. Koerkel, M.: On the exact solution of large-scale simple plant location problems. Eur. J. Oper. Res. **39**, 2, pp. 157–173, (1989).
31. Lämmer, S., Gehlsen, B., Helbing, D.:Scaling laws in the spatial structure of urban road networks, Physica A, **363**, 1, pp. 89–65, (2006).
32. Mladenovic, N., Brimberg, J., Hansen, P.: A note on duality gap in the simple plant location problem. European Journal of Operational Research, **174**, pp. 11–22, (2006).
33. Nace, D., Doan, L.N., Klopfenstein, O., Bashllari, A.: Max-min fairness in multi-commodity flows, Computers & Operations Research **35**, pp. 557–573. (2008).
34. Nogueirai, T., Mendes, R., Valie, Z., Cardoso, J. C.: Optimal Location of Natural Gas Sources in the Iberian System, Proceedings of the 6th WSEAS International Conference on Power Systems, Lisbon, Portugal, September 22–24, (2006).
35. Pioro, M., Medhi, D.: Routing, flow, and capacity design in communication and computer networks, Morgan Kaufmann Publishers, (2004).

36. Rawls, J.: A Theory of justice. Harward University Press, (1971).
37. Resende, M.G.C., Pardalos, P.M.: Handbook of optimization in telecommunications, Springer Science+Business Media, Inc., (2006).
38. Silva, F., Serra, D.: Locating Emergency Services with Different Priorities: The Priority Queuing Covering Location Problem, The Journal of the Operational Research Society, **59**, 9, pp. 1229–1238, (2008).
39. Srinivasan, U.T., Dunne, J.A., Harte, J., Martinez, N.D.: Response of complex food webs to realistic extinction sequences, Ecology **88**, 617, (2007).
40. Sripetch, A., Saengudomlert, P.: Topology Design of Optical Networks Based on Existing Power Grids, Fifth Annual Conference on Communication Networks and Services Research(CNSR'07), (2007).
41. Toregas, C., Swain, R., ReVelle C., Bergman, L.: The location of emergency service facilities, Operations Research, **19**, pp. 1363–1373, (1973).
42. Whittle, P., Networks optimization and evolution, Cambridge University Press, (2007).