Control and Cybernetics

vol. **34** (2005) No. 4

On the application of statistical kernel estimators for the demand-based design of a wireless data transmission system

by

Piotr Kulczycki[1], Jacek Waglowski[2]

[1] Polish Academy of Sciences, Systems Research Institute
Newelska 6, 01-447 Warszawa, Poland

[2] Polish Academy of Sciences, Systems Research Institute
Doctoral Course of Study
Newelska 6, 01-447 Warszawa, Poland

Abstract: The subject of this paper is the task of designing the LMDS (Local Multipoint Distribution System) wireless broadband data transmission system. The methodology of statistical kernel estimators and fuzzy logic using operations research and mathematical programming is applied to find optimal locations for its base-stations. A procedure which allows to obtain such locations on the basis of potential customer distribution and their expected demand, also in the cases of uncertain and non-stationary data, is investigated.

Keywords: telecommunication, wireless broadband data transmission systems, LMDS, statistical kernel estimators, fuzzy logic, operations research, mathematical programming.

1. Introduction

The Local Multipoint Distribution System (LMDS) is applied by telecommunications operators for wireless broadband data transmission purposes. The dynamic development of information system applications observed in recent years caused a considerable increase in such business activity, and a spectacular example of that is a rapid increase in interest in gaining access to the Internet. The LMDS system allows to connect the operator's network node to the buildings in which customers are located, without the necessity of constructing an expensive cable infrastructure. Thus, data is transmitted between the base-stations distributed across a metropolitan area, and those stations service the regular connections with subscriber-stations located within the effective coverage of transceivers belonging to base-stations. Subscriber-stations installed on building roofs or facades then transmit data to customers through local, e.g. cable, networks.

An essential factor which often decides about economic justification of the LMDS system implementation is to determine base-station locations so that the highest profit can be achieved, within the available investment funds. Presently, there is no methodology to allow for a general solution of the problem formulated in that way: in practice, heuristic methods are applied, largely based on intuition, or other methods from related fields are adopted; see books by Laiho, Wacker and Novosad (2001) and Rappaport (1996) and papers by Bsagni (2001), Franceschetti, Cook and Bruck (2004), Gupta and Kumar (2000), Lee and Kang (2000), Sherali, Pendyala and Rappaport (1996), Tran-Gia, Leibnitz and Tutschku (2000), Tutschku and Tran-Gia (1998), Vohra and Hall (1993), also to find rich bibliography. The task of LMDS base-station location planning is not easy due to the requirement of taking into account a number of technical conditions, as well as economic ones, also in the situation of data uncertainty and non-stationarity.

Basic technical constraints include the theoretical, maximal coverage radius of base-station transceivers, as well as their maximal bitrate, i.e. the largest total data quantity which can be transmitted in a time unit. What is also required for ensuring data transmission is the line of sight between the base-station and the subscriber-station antennas. For that reason, due to complex land shaping, or such obstacles as tall buildings in the base-station coverage area, there may exist shadow areas, in which it is not possible to transmit between the base-station and the buildings located in such areas. Thus, there is a limited number of sites well visible due to their elevation, which can be selected as potential locations for base-stations.

In addition to the above-mentioned technical constraints, another problem facing planners constitutes the estimation of future demand. In practice, such a process can be developed only on the basis of imprecise and incomplete data concerning potential service users located in a given area. Despite such uncertainty, estimation is indispensable in defining a spatial distribution of the predicted demand. This problem gets even more difficult when planning is long-term, especially with non-stationarity of data.

Consequently, the planning task requires a choice of those possible base-station locations which ensure the maximum profit from services, while the number of stations is limited by the availability of funds. This paper will present an algorithm for designing optimal LMDS base-station locations. The method of statistical kernel estimators has been applied for the purpose of describing the spatial distribution of demand for data transmission services. Due to natural uncertainty of demand values, fuzzy logic elements have also been used. In addition, the issues of existence of shadow regions in the coverage areas of base-stations and the problem of their limited bitrates have been taken into account. It is also possible to apply that method when planning with a horizon of several years.

The preliminary version of this paper was presented as Kulczycki and Waglowski (2003).

2. Estimation of the distribution of spatial demand: statistical kernel estimators

The distribution of the spatial demand for data transmission services in the area under consideration has strict point structure connected with particular potential customers. Such a model has, however, limited application significance since it is practically unidentifiable in a metropolitan area. In this paper, statistical kernel estimators will be applied for this purpose. Based on the most representative objects, this approach allows for ensuring the continuity of the model and proper average of spatial demand, making such an approach convenient for optimization tasks.

Let the n-dimensional random variable X, with a distribution having the density function f, be given. Its kernel estimator $\hat{f} : \mathbb{R}^n \to [0, \infty)$ is calculated on the basis of the m-element random sample $x_1, x_2, \ldots, x_m$, obtained experimentally from the variable X, and is defined in its basic form by the following formula:

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right) \quad , \tag{1}$$

where the function $K : \mathbb{R}^n \to [0, \infty)$, which is Borelian and radially symmetrical relative to zero, has a weak global maximum at this point, fulfilling the condition $\int_{\mathbb{R}^n} K(x)\ \mathrm{d}x = 1$, and is called the kernel, whereas the positive coefficient h is known as the smoothing parameter; for interpretation, see also Fig. 1. The form of the kernel K and the value of the smoothing parameter h is selected most often on the basis of the criterion of the minimum mean square error; for proper algorithms, see books by Kulczycki (2005), Silverman (1986), Wand and Jones (1994). It turns out that the form of the function K has no essential importance from the statistical point of view, and for that reason, it is possible when selecting this function to take into account primarily the features of the estimator required in the case of a particular task.

In practical problems, additional procedures are used for improving the properties of kernel estimators. In the methodology investigated here, the so-called modification of the smoothing parameter is preferred. The specific procedure can be performed as follows:

(A) the kernel estimator $\hat{f}$ is specified in accordance with formula (1);

(B) the modifying parameters $s_i > 0 \quad (i = 1, 2, \ldots, m)$ of the form:

$$s_i = \left(\frac{\hat{f}(x_i)}{s^{\sim}}\right)^{-1/2} \tag{2}$$

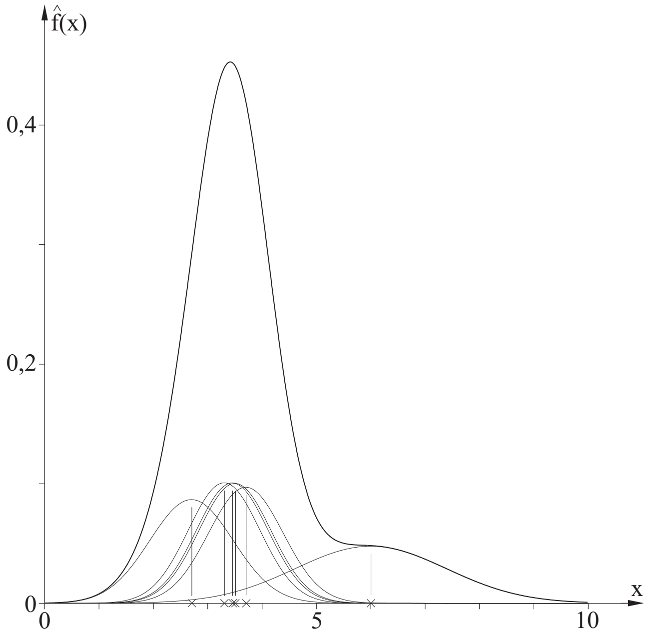

Figure 1. A kernel estimator with the modification of the smoothing parameter (4) for the one-dimensional case (i.e. $n = 1$)

are calculated, where $s^{\sim}$ is the geometric mean of the numbers $\hat{f}(x_1)$, $\hat{f}(x_2), \ldots, \hat{f}(x_m)$, given by the logarithmic equation

$$\log(s^{\sim}) = m^{-1} \sum_{i=1}^{m} \log(\hat{f}(x_i)) \; ; \tag{3}$$

(C) the kernel estimator with the modification of the smoothing parameters finally assumes the form of

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} \frac{1}{s_i^n} K\left(\frac{x - x_i}{hs_i}\right) \; . \tag{4}$$

As a result of the introduction of the above concept, the areas where estimator (1) takes on small values are additionally flattened, contrary to parts of its large values where the characterization of specific features of distribution improves; see Fig. 1. A broader discussion of the issues presented above can be found in books by Kulczycki (2005), Silverman (1986), Wand and Jones (1994). Exemplary applications of kernel estimators are described in papers of Kulczycki (2000, 2001, 2002a,b), Kulczycki and Wisniewski (2002), Schiøler and Kulczycki (1997).

In the problem investigated here, the kernel estimator will be used for the characterization of distribution of spatial demand for data transmission services in the area under consideration. The variable X is therefore two-dimensional, i.e. $n = 2$, while its particular coordinates represent longitude and latitude. The kernel estimator with the modification of the smoothing parameter (4) will be applied after additional mapping of the coefficients $w_i > 0$ for $i = 1, 2, \dots, m$ to every kernel; therefore,

$$\hat{f}(x) = \frac{1}{h^2 \sum\limits_{i=1}^{m} w_i} \sum_{i=1}^{m} \frac{w_i}{s_i^2} K\left(\frac{x - x_i}{h s_i}\right) . \tag{5}$$

Because of the convenience of analytical calculations, the Cauchy kernel

$$K(x) = \frac{1}{\pi(||x||^2 + 1)^2} \tag{6}$$

will be used.

Finally, the statistical kernel estimators methodology will be applied in this paper, to identify the distribution of the spatial demand for data transmission services in the whole area under consideration. With its use one can obtain a continuous characteristic of such distribution derived from data with a point feature, while taking into account the fundamental traits of quality of a considered area: density of client concentration and the corresponding expected level of demand. Particular subscriber buildings in the data base are characterized by their geographical position $x_i = [x_{i1}, x_{i2}]^{\mathrm{T}}$ and the coefficient w_i representing potential demand for data transmission services corresponding to this location $(i = 1, 2, \dots, m)$. The identified distribution is properly made continuous owing to the properties of statistical kernel estimators. Moreover, due to the averaging aspects of such estimators, it is possible to use a simplified data base, including only the locations of main subscriber buildings, and taking into account in the corresponding coefficients w_i also smaller objects in their neighborhood. Thanks to this, the influence of imprecise and/or incomplete identification of potential subscriber location on the final result will be significantly lowered, as will the cost associated with its execution. The smoothing and averaging properties of statistical kernel estimators considerably simplify the most difficult and expensive phase of the procedure of planning optimal locations of LMDS base-stations investigated in this paper.

3. Base-station system performance index

In practice, it is not difficult to identify a limited number of sites for installing base-stations, including e.g. tall buildings and telecommunications towers. Having defined in the previous section the function $\hat{f}$ which characterizes the spatial distribution of demand for data transmission services, one can map for particular locations the values resulting from that function's integration, within the coverage areas of the respective transceivers. Next, in the case of a base-station system, the integral for the whole area covered by the ranges of particular transceivers defines the capacity of meeting the total demand, being also a criterion of the appraisal of the system's quality. This section will present the basic algorithm for calculating the integral's value.

Let the set of k potential locations of base-stations at sites $x_j = [x_{j1}, x_{j2}]^{\mathrm{T}}$, with $j = 1, 2, \ldots, k$, be given. The following notations are introduced:

$$E_j = \int\limits_{C_j} \hat{f}(x)\ \mathrm{d}x \tag{7}$$

$$E_{j_1, j_2, \ldots, j_n} = \int\limits_{C_{j_1} \cap C_{j_2} \cap \ldots \cap C_{i_n}} \hat{f}(x)\ \mathrm{d}x\ , \tag{8}$$

where C_j denotes the j-th circle with the center at x_j and the positive radius r_j (representing maximal range of the transceiver mapped to the j-th location), and $j_1, j_2, \ldots, j_n \in \{1, 2, \ldots, k\}$ are different, while $2 \leq n \leq k$. The total demand characterizing the quality of the base-station system, is given by the formula

$$\begin{aligned} E &= \int\limits_{C_1 \cup C_2 \cup \ldots \cup C_k} \hat{f}(x)\ \mathrm{d}x \\ &= \sum_{j=1}^{k} E_j - \sum_{\{j_1, j_2\}} E_{j_1, j_2} + \sum_{\{j_1, j_2, j_3\}} E_{j_1, j_2, j_3} + \ldots + (-1)^k E_{1,2,\ldots,k}\ . \end{aligned} \tag{9}$$

The text below presents an algorithm for calculating the values of formulas (7) and (8), which exhausts the procedure allowing to define the capacity of meeting the demand for data transmission services within the fixed base-station system, in accordance with formula (9), which characterizes system quality.

Due to the selection of the kernel in the form (6), it is possible to calculate an analytical formula for the integral from the function of the single kernel K_i with the parameters h, s_i and w_i, on the circle C_j, with the radius r_j and the distance $d_{i,j}$ between the centers of the circle and the kernel (for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, k$), expressed through

$$E_j = \frac{1}{2 \sum\limits_{i=1}^{m} w_i} \sum_{i=1}^{m} w_i \left(\frac{r_j^2 - d_{i,j}^2 - h^2 s_i^2}{\sqrt{r_j^4 + 2(h^2 s_i^2 - d_{i,j}^2) r_j^2 + (h^2 s_i^2 + d_{i,j}^2)^2}} + 1 \right). \tag{10}$$

For the purpose of the problem under consideration, the above formula may be simplified to the form

$$E_j = \sum_{i=1}^{m} w_i \left(\frac{r_j^2 - d_{i,j}^2 - h^2 s_i^2}{\sqrt{r_j^4 + 2(h^2 s_i^2 - d_{i,j}^2) r_j^2 + (h^2 s_i^2 + d_{i,j}^2)^2}} + 1 \right), \tag{11}$$

because the values calculated on that basis will be subject to comparison in the search for an optimal element, therefore, multiplication of the performance index by the positive constant $2\sum_{i=1}^{m} w_i$ does not affect the results obtained in such a task. Following the procedure proposed below one can successively conclude that such operation results only in a "rescaling" of value of performance index (7).

Next, the analytical calculation of the integral value, in the case when an integration set is the intersection of any number of circles, is practically unexecutable. A practical approximate procedure will therefore be investigated below.

First, the case of the kernel K_i and two circles C_{j_1} and C_{j_2} will be considered. Owing to a possibility of renumbering, it can be assumed that $r_{j_1} \leq r_{j_2}$. Let D_{j_1,j_2} denote the distance between the circles' centers. One of the following relationships may occur between those circles:

(A) $D_{j_1,j_2} \geq r_{j_1} + r_{j_2}$, implying disjunction of the circles or edge contact; then, $E_{j_1,j_2} = 0$;

(B) $D_{j_1,j_2} \leq r_{j_2} - r_{j_2}$, which means that the smaller circle is their intersection; then, $E_{j_1,j_2} = E_{j_2}$, whose value may be calculated from formula (11);

(C) neither of the previous cases occurs; the circle intersection has the shape of a lens; the method of calculating the approximate value of E_{j_1,j_2} is given below.

This method is realized by replacing the lens with a circle, for which formula (11) can be applied. By guaranteeing equal fields of the circle and the lens, and with proper location of the circle's center, the difference between the values of function (5) on the areas of the lens and of the circle is not large, while the error of integration (having the averaging nature) on them is fairly insignificant. It is worth noticing that the largest values of the error occur when the lens is considerably flattened, therefore, when its field, together with the value E_{j_1,j_2}, is relatively small, in the region of few percent.

Let $\hat{D}_{j_1,j_2}$ denote the distance between the points of intersection of the circles C_{j_1} and C_{j_2}; according to the assumptions of case (C): $\hat{D}_{j_1,j_2} > 0$. The calculation of the value $\hat{D}_{j_1,j_2}$ is not difficult, based on non-complex procedures of analytical geometry (Waglowski, 2005). The field of the lens L_{j_1,j_2} can be defined, in the case of a flat lens, i.e. when $r_{j_2} \geq \sqrt{r_{j_1}^2 + D_{j_1,j_2}^2}$, by formula

$$
\begin{aligned}
L_{j_1,j_2} = & \frac{\hat{D}_{j_1,j_2}}{2}\left(\sqrt{r_{j_1}^2 - \left(\frac{\hat{D}_{j_1,j_2}}{2}\right)^2} + \sqrt{r_{j_2}^2 - \left(\frac{\hat{D}_{j_1,j_2}}{2}\right)^2}\right) \\
& + r_{j_1}^2 \arcsin\left(\frac{\hat{D}_{j_1,j_2}}{2r_{j_1}}\right) + r_{j_2}^2 \arcsin\left(\frac{\hat{D}_{j_1,j_2}}{2r_{j_2}}\right) - \hat{D}_{j_1,j_2} D_{j_1,j_2}\,, \quad (12)
\end{aligned}
$$

however, in the case of a convex lens, i.e. when $r_{j_2} < \sqrt{r_{j_1}^2 + D_{j_1,j_2}^2}$, as

$$L_{j_1,j_2} = \frac{\hat{D}_{j_1,j_2}}{2}\left(\sqrt{r_{j_1}^2 - \left(\frac{\hat{D}_{j_1,j_2}}{2}\right)^2} - \sqrt{r_{j_2}^2 - \left(\frac{\hat{D}_{j_1,j_2}}{2}\right)^2}\right) + r_{j_1}^2 \arcsin\left(\frac{\hat{D}_{j_1,j_2}}{2r_{j_1}}\right) - r_{j_2}^2 \arcsin\left(\frac{\hat{D}_{j_1,j_2}}{2r_{j_2}}\right) + \hat{D}_{j_1,j_2} D_{j_1,j_2}\,. \quad (13)$$

Upon calculation of the lens field L_{j_1,j_2} from formulas (12) or (13), one may easily calculate the radius of the substitute circle r_{j_1,j_2}:

$$r_{j_1,j_2} = \sqrt{\frac{L_{j_1,j_2}}{\pi}}\quad . \quad (14)$$

Its center can be defined in the following manner. The straight line crossing the centers of the circles C_{j_1} and C_{j_2} is also crossing each of them at two different points, one on each lens edge. Let the center between those points be the center of the substitute circle. Calculation of its coordinates is not difficult using the analytical geometry methods (Waglowski, 2005). Once the center and the radius of the substitute circle are known, it is possible to calculate the value of E_{j_1,j_2}, based on formula (11).

The above procedure may be easily generalized in the recurrent manner in the cases of intersection of any number of circles, exceeding two. Upon ordering circles according to increasing radius size, it is necessary to calculate the substitute circle parameters for the lens obtained from the first pair, followed by subsequent iterations for the substitute circle and subsequently considered ones, repeating such iterations until the list of circles is exhausted. The result is a substitute circle for the area being an intersection of all the circles under consideration. It is possible to apply formula (11) to the resulting circle.

The above process completes the basic calculation algorithm necessary to apply formula (9), allowing to characterize the quality of the given base-station system. In the next two sections, the algorithm will be modified to take into account shadow areas and limited bitrates of base-stations.

4. Performance index modification to account for shadow areas

In the previous considerations, the integration set of the density function characterizing spatial demand distribution, was assumed to be an area being the union of circles resulting from base-station coverages (see formula (9)). As it was mentioned in the introduction, within that area the shadow area occurs, in which transmission is impossible due to uneven land or obstacles, e.g. tall buildings. To account for a shadow area, it is necessary to subtract the integral value on those fields from the value given by formula (9). It should be pointed out that the shadow area of one base-station may be covered, in whole or in part, by another base-station, and, by the same, such a shadow area will not be treated from the point of view of the whole system as a shadow, at least partially.

In the practice of designing teletransmission networks, one may come across various approaches to accounting for shadow areas. Maps resulting from the spatial analysis software, obtained through aerial image processing, are often applied for this purpose. Sometimes, however, to reduce cost and accelerate analysis, only approximate sketches are produced, followed by the approximation of shadow areas by simple geometric figures, while those figures are treated as circles, or, generally, circle unions. With this approach, the algorithm developed in the previous section allows for easy calculation of the integral from the density function of the spatial distribution of the demand for data transmission services in shadow areas, in analogy to formula (9), followed by subtraction of that value from the index calculated in Section 3.

5. Performance index modification to account for limited base-station bitrate

The performance index of the particular base-station system, defined by formula (9), represents the capability of meeting the total demand for teletransmission services provided within the system's transceiver coverage. However, in especially attractive city areas, the coverage demand may not be met due to limited transceiver bitrates. In this section, the procedure allowing to account for limited bitrates of particular base-stations will be worked out.

To reduce the dimensionality of the optimization problem considered below, first, it is necessary to exclude from investigation those base-stations belonging to the system whose bitrates are higher than or equal to the demand under coverage, i.e. the value of $\int_{C_j} \hat{f}(x)\,\mathrm{d}x$. Moreover, if the set, being the union of the areas within the coverage of the stations remaining after the above activity is not connected (i.e. it is composed of disjoint subsets), it is necessary to decompose the task by conducting the considerations described below, separately for each of such subsets.

Let $b_j > 0$ with $j = 1, 2, \ldots, k^*$ where $k^* \leq k$, mean maximal bitrates of particular transceivers belonging to a subsystem of k^* base-stations. The set, being the union of the areas within the base-stations' coverage, is divided by the circles constituting coverage edges of particular transceivers into a finite number of subsets with nonempty interior (the maximal possible number is $2^{k^*} - 1$). Those sets, denoted further as Z_i, will be numbered with the index $i = 1, 2, \ldots, I$. Using the algorithm presented in Section 3, the approximate of the integral $\int_{Z_i} \hat{f}(x)\,\mathrm{d}x$ for each $i = 1, 2, \ldots, I$, can be calculated.

Let the matrix A with the dimension $k^* \times I$ and nonnegative elements, be given. Particular rows of the matrix are connected with subsequent base-stations of the system under consideration, while columns – with particular subsets Z_i. If the i-th subset is outside of the range of the j-th station, one should assume that $a_{j,i} = 0$. The following performance index will be considered, and the decision variables will be all the elements of the matrix A whose value was not assumed above as zero (the respective set will be denoted below as $\{a^*_{j,i}\}$):

$$\max_{\{a^*_{j,i}\}} \sum_{\substack{j=1,2,\ldots,k^* \\ i=1,2,\ldots,I}} a_{j,i} \,, \tag{15}$$

with the constraints

$$a_{j,i} \geq 0 \quad \text{for} \quad j = 1,2,\ldots,k^* \quad \text{and} \quad i = 1,2,\ldots,I \tag{16}$$

$$\sum_{i=1}^{I} a_{j,i} \leq b_j \quad \text{for} \quad j = 1,2,\ldots,k^* \tag{17}$$

$$\sum_{j=1}^{k^*} a_{j,i} \leq \int_{C_i} \hat{f}(x)\, \mathrm{d}x \quad \text{for} \quad i = 1,2,\ldots,I. \tag{18}$$

This is a typical linear optimization task. Calculations are facilitated by the fact that the task is "sparse": the elements of the matrix A have mostly the value of zero. Moreover, by arranging the elements of the set $\{a^*_{j,i}\}$ in a vector, the task may be described in a canonical form and solved by the generally available simplex method, see, e.g. Garfinkel and Nemhauser (1972), Wagner (1975). Each of the elements $a_{j,i}$, obtained in accordance with the above procedure, indicates which portion of demand from the area Z_i should be served by the j-th station in order to meet the largest possible demand for telecommunication services for the given base-station system, taking into account limited bitrates of the respective transceivers.

6. Selection of the optimal base-station system

Once the base-station system performance index has been worked out in accordance with Sections 3 to 5, one may start solving the basic task of the present paper, i.e. selection of the optimal base-station system. For that purpose, the methods originating from operations research (Garfinkel and Nemhauser, 1972; Wagner, 1975) will be applied.

The utilization of radio frequencies made available to the telecommunications operator requires the application of devices with essentially different functional parameters. In the model presented here, a possibility of selecting, in each potential location, one possible version of transceivers, from among p options, while $p \in \mathbb{N}\backslash\{0\}$, is assumed. Particular versions are represented with the following positive parameters: r_i – coverage radius, b_i – maximal bitrate, and c_i – cost of equipment and its installation, where $i = 1, 2, \ldots, p$. The case when no equipment is installed at a location is reflected by $i = 0$ and $c_0 = 0$.

Let the k-dimensional decision vector

$$\begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{bmatrix} \tag{19}$$

be given. Particular coordinates represent potential base-station locations, and assume the values $g_j \in \{0, 1, \ldots, p\}$ for $j = 1, 2, \ldots, k$. To be more precise: if the j-th coordinate is 0, i.e. $g_j = 0$, it means that the transceiver installation at the j-th location is not planned; however, if that coordinate takes on the value i from the range 1,2,...,p, it means that the i-th version of such device is installed at the j-th location. The optimization task consists here of searching for the maximum of the expression

$$\max_{g_1, g_2, \ldots, g_k} E([g_1, g_2, \ldots, g_k]^{\mathrm{T}}) \tag{20}$$

with the constraint

$$\sum_{j=1}^{k} c_{g_j} \leq \tilde{C}, \tag{21}$$

where the positive number $\tilde{C}$ means the maximal amount of available funds, while $E([g_1, g_2, \ldots, g_k]^{\mathrm{T}})$ denotes the value of function (9) for a system of transceivers distributed in accordance with the value of the decision vector $[g_1, g_2, \ldots, g_k]^{\mathrm{T}}$.

The above issue can be reduced to the classical form of a decision tree. Thus, let the k-level decision tree be given (see Fig. 2): particular levels represent subsequent potential base-station locations. Decision tree nodes are assigned subsequently one of the possible values $g_j \in \{0, 1, \ldots, p\}$ for $j = 1, 2, \ldots, k$; if the j-th level is assigned the value g_j, the node represents the case in which the g_j-th version of a transceiver is installed at the j-th location. That also implies assigning to that node the cost c_{g_j} of the given version of a transceiver, which is necessary to verify constraint (21). The solution of the problem under consideration consists in the determination of a path from the first level node to the k-th level node, described by the vector $[g_1, g_2, \ldots, g_k]^{\mathrm{T}}$, for which the function E reaches the maximum, and constraint (21) is fulfilled. To solve so formulated a task, the classical branch-and-bound method has been applied; for an intuitional illustration, see Fig. 2.

An important element affecting the rate of calculation is the effective fathoming ("closing") of those nodes from which a better path than previously

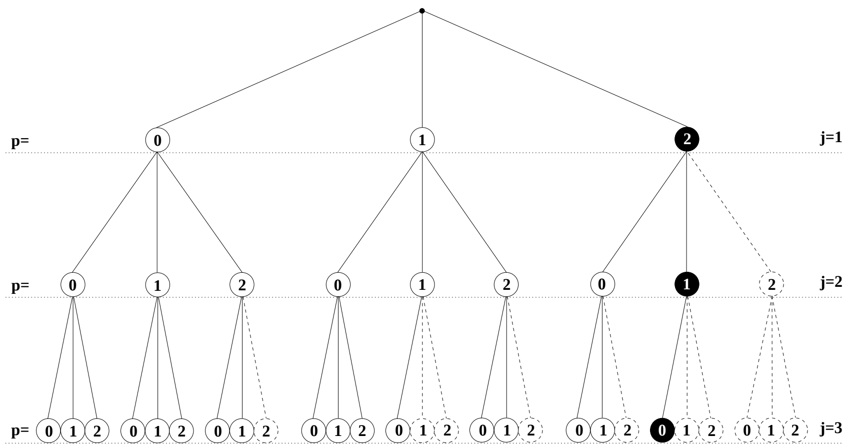

Figure 2. An illustration of the decision tree for $k = 3$, $p = 2$; the available funds are adequate for the purchase of not more than two transceivers, including only one of a more expensive kind; the dotted line marks the paths omitted owing to the fathoming of proper nodes; the black spots mark nodes included in the optimum path.

found cannot be generated. Let the numbering of particular transceiver versions be such that $c_0 \leq c_1 \leq \ldots \leq c_p$, and the numbering of tree levels such that $\int_{C(x_1,r_{\max})} \hat{f}(x)\,\mathrm{d}x \geq \int_{C(x_2,r_{\max})} \hat{f}(x)\,\mathrm{d}x \geq \ldots \quad \ldots \geq \int_{C(x_k,r_{\max})} \hat{f}(x)\,\mathrm{d}x$, i.e. according to the demand level met by a given location for the transceiver version with the largest range $r_{\max}$; ($C(x,r)$ denotes in the above notations a circle with the center x and the radius r). If the node under consideration is located in layer $j \in \{1, 2, \ldots, k-1\}$, it will not be difficult to calculate the number $J \in \mathbb{N}$ stating how many cheapest transceivers may be installed within the available funds, i.e.

$$J = \mathrm{int}\left(\frac{\tilde{C} - \sum_{i=1}^{j} c_{g_i}}{c_{g_1}} \right), \tag{22}$$

where $\mathrm{int}(a)$ denotes the integer part of the number $a \in \mathbb{R}$. Let a fragment of the path above the level j describe the partial decision vector $[g_1, g_2, \ldots, g_j]^{\mathrm{T}}$. If the value E, characterizing according to formula (9) the base-station quality for the decision vector

$$[g_1, g_2, \ldots, g_j, \underbrace{g_{r_{\max}}, \ldots, g_{r_{\max}}}_{J \text{ factors}}, 0, \ldots, 0]^{\mathrm{T}}, \tag{23}$$

where $g_{r_{\max}}$, representing the version of transceivers with the largest coverage, is smaller than or equal to the maximum of previously calculated value E, then such a node should be fathomed because decision vectors of the form $[g_1, g_2, \ldots, g_j, \text{any}]^{\mathrm{T}}$ may not produce a better path than the one found.

7. Long-term planning horizon

The task previously considered was stationary in nature. Its characteristic values, i.e. size and the distribution of spatial demand, available investment funds, prices and parameters of transceivers, and a set of potential base-station locations, were not subject to the changes occurring over time. However, one can expect increased transmission to current customers and inclusion of new customers with time. Also, a gradual increase of funds can be expected owing to current income and the operator's growing creditworthiness. In addition, the parameters of transceivers are also changed. After signing new agreements and expanding urban infrastructure, new base-station locations will become available. The methodology presented in this paper allows for accounting easily for the time factor and for all the above-mentioned aspects.

If the project is considered within $T \in \mathbb{N}\backslash\{0,1\}$ time periods (not necessarily equal ones, although in practice these refer to particular periods, and most often $T = 2$ or $T = 3$), the decision vector (19) should be generalized to the form

$$[g_{1,t=1}, g_{2,t=1}, \ldots, g_{k_1,t=1},\; g_{1,t=2}, g_{2,t=2}, \ldots, g_{k_2,t=2},\; \cdots,\\ g_{1,t=T},\; g_{2,t=T}, \ldots, g_{k_T,t=T}]^{\mathrm{T}}\,, \tag{24}$$

where the parameter $t = 1, 2, \ldots, T$ characterizes particular time periods (the above notation also allows for the number of potential base-station locations $k_1, k_2, \ldots, k_T$ to change in particular periods). Constraint (21) assumes the form of T independent conditions

$$\sum_{j=1}^{k_t} c_{g_j,t} \le \tilde{C}_t \quad \text{for} \quad t = 1, 2, \ldots, T, \tag{25}$$

where the parameters occurring above were correspondingly indexed by t, characterizing transceiver prices $c_{g_j,t}$ and the maximal value of available funds $\tilde{C}_t$ in particular years. Performance index (20) becomes a linear combination of subsequent components corresponding to particular time periods

$$\max_{\substack{g_{1,t},\, g_{2,t},\, \cdots,\, g_{i,t} \\ t=1,2,\ldots,T}} \sum_{t=1}^{T} d_t E_t(\,[g_{1,t}, g_{2,t}, \ldots, g_{k,t}]^{\mathrm{T}})\,, \tag{26}$$

while the re-occurring parameters are indexed by $t = 1, 2, \ldots, T$, and the positive weights d_t represent the meanings of particular periods. These parameters are of special importance when the time periods are unequal and in the case when the periods considered in a long-term perspective are less significant in the planning process due to the possibility of making strategic changes in the future. The values of the particular factors E_t may be calculated on the basis of the diverse functions $\hat{f}_t$ characterizing a predicted spatial distribution of demand for data transmission services in the area under consideration, as well as diverse transceiver parameters $r_{i,t}$, $b_{i,t}$, and $c_{i,t}$.

If a replacement of installed device versions is not assumed, the appearance of the element $g_{j,t} \neq 0$ at any position of the decision vector implies the same value of decision vector elements representing a given location for $t+1, t+2, \ldots, T$, which effectively reduces the number of paths in the decision tree. In the opposite case, however, if such a replacement is accepted, then, in the year of a change, the parameter $c_{g_j,t}$ represents, not – as before – the price of equipment and its installation, but the cost of the change itself.

8. Fuzzy nature of demand

The coefficients w_i for $i = 1, 2, \ldots, m$ introduced in formula (5) represent the demand for teletransmission services assigned to particular subscriber-station locations. Their value is estimated in practice by the analysis of the nature of particular service users, based on their affiliations with consumer groups defined for common features. Identification of such uncertainty is in practice conducted on the basis of expert opinions expressed verbally, often on intuitional premises. Consequently, the description of the predicted demand for teletransmission services by a subscriber-station will require fuzzy logic elements (Kacprzyk, 1986). What should also be taken into account is the specific nature of the task under consideration: a lot of fuzzy numbers (equal to the number of subscriber-stations m) necessary to identify and to use in subsequent analysis, as well as the fact that incidentally, the coefficients w_i may be deterministic owing to previously signed agreements. In that situation, especially suitable are fuzzy numbers of the type L-R, whose membership function is assumed here in the following form:

$$\mu_{(w_i,\alpha_i,\beta_i)}(x) = \begin{cases} L\left(\dfrac{w_i - x}{\alpha_i}\right) & \text{for} \quad x \leq w_i \\ R\left(\dfrac{x - w_i}{\beta_i}\right) & \text{for} \quad x \geq w_i \end{cases}, \tag{27}$$

where w_i, α_i, $\beta_i > 0$, the fixed function $L : (-\infty, 0] \to [0,1]$ is nondecreasing as well as $R : [0, \infty) \to [0,1]$ is nonincreasing, and $L(0) = R(0) = 1$. The parameter w_i may be interpreted as a modal value, while α_i and β_i describe left- and right-hand concentration around that value, respectively.

The fuzzy number $\mathcal{A}$ of the type L-R may, therefore, be identified by three parameters, which will be denoted as

$$\mathcal{A} = (w, \alpha, \beta) \tag{28}$$

and, consequently, the process of identification requires only determination of the values which are close to intuitional interpretation. Algebraic operations on fuzzy numbers of the type L-R are defined as follows:

$$\mathcal{A} + \mathcal{B} = (w_{\mathcal{A}}, \alpha_{\mathcal{A}}, \beta_{\mathcal{A}}) + (w_{\mathcal{B}}, \alpha_{\mathcal{B}}, \beta_{\mathcal{B}}) = (w_{\mathcal{A}} + w_{\mathcal{B}}, \alpha_{\mathcal{A}} + \alpha_{\mathcal{B}}, \beta_{\mathcal{A}} + \beta_{\mathcal{B}}) \tag{29}$$

$$\mathcal{A} - \mathcal{B} = (w_{\mathcal{A}}, \alpha_{\mathcal{A}}, \beta_{\mathcal{A}}) - (w_{\mathcal{B}}, \alpha_{\mathcal{B}}, \beta_{\mathcal{B}}) = (w_{\mathcal{A}} - w_{\mathcal{B}}, \alpha_{\mathcal{A}} + \alpha_{\mathcal{B}}, \beta_{\mathcal{A}} + \beta_{\mathcal{B}}) \tag{30}$$

$$c \cdot \mathcal{A} = (cw_{\mathcal{A}}, c\alpha_{\mathcal{A}}, c\beta_{\mathcal{A}}), \tag{31}$$

where $\mathcal{A}$ and $\mathcal{B}$ denote fuzzy numbers, while c is a positive real number. If one adopts the notation in which the real number a is described in the form of three parameters $a = (a, 0, 0)$, those operations may be generalized to addition and subtraction of the fuzzy and real numbers. Moreover, formulas (29)-(31) express then correctly also the operations on two real numbers. Finally, the result is that the fuzzy number of the type L-R in the above range is a generalization of the real number. In this paper

$$L(x) = R(x) = \begin{cases} 1 - x^2 & \text{for} \quad x \in [-1, 1] \\ 0 & \text{for} \quad x \in (-\infty, -1) \cup (1, \infty) \end{cases} \tag{32}$$

has been assumed, in a form typical for the natural expression of experts' opinions in similar tasks, where the values in the neighborhood of an extreme one are treated as not much less likely.

Finally, for each of m locations of subscriber buildings, the coefficient w_i representing potential demand for data teletransmission services, introduced in formula (5), was generalized to the three-parameter fuzzy number suitable for identification and calculations in practice, denoted below as $\mathcal{W}_i = (w_i, \alpha_i, \beta_i)$, where $w_i - \alpha_i \geq 0$ for every $i = 1, 2, \ldots, m$. In a special case, $\mathcal{W}_i = (w_i, 0, 0)$ may represent the real ("non-fuzzy") number w_i.

As one can infer from formulas (9), (11), and the modifications presented in Sections 4 and 5, the performance index of the base-station system under consideration has a form of linear combination of three-parameter fuzzy numbers $\mathcal{W}_i$, and, therefore, due to formulas (29)-(31), it also becomes a three-parameter fuzzy number, denoted below as $\mathcal{E}$. To allow for the comparison of qualities of particular base-station systems, the methodology of fuzzy preference theory (Fodor and Roubens, 1994), will be applied. The preference function P of the fuzzy number $\mathcal{E}$, with the bounded support of the membership function, will be adopted in the form resulting from the decision-making practice (Berger, 1980):

$$P(\mathcal{E}) = \delta \, \frac{\int\limits_{\min \operatorname{supp} \mu_{\mathcal{E}}}^{\max \operatorname{supp} \mu_{\mathcal{E}}} x \mu_{\mathcal{E}}(x) \, \mathrm{d}x}{\int\limits_{\min \operatorname{supp} \mu_{\mathcal{E}}}^{\max \operatorname{supp} \mu_{\mathcal{E}}} \mu_{\mathcal{E}}(x) \, \mathrm{d}x} + (1 - \delta) \min \operatorname{supp} \mu_{\mathcal{E}} \, , \tag{33}$$

where $\delta \in [0, 1]$, $\mu_{\mathcal{E}}$ means the membership function of the fuzzy number $\mathcal{E}$, while $\operatorname{supp} \mu_{\mathcal{E}}$ denotes its support. The value of the membership function is therefore a linear combination with weights δ and $1-\delta$ of the average value of the fuzzy number and the minimum value of its support. The average number corresponds to the Bayes decision rule and expresses a "realistic" operation, while the minimum value of the membership function support results from the minimax rule and represents the "pessimistic" point of view. The parameter δ determines therefore the company's strategy in the range from realistic (assuming average predicted demand) for $\delta = 1$, to pessimistic (assuming the lowest level of predicted demand) for $\delta = 0$. When clear preferences are missing, the value $\delta = 0.5$ can be proposed.

In the case of the three-parameter number $\mathcal{E} = (e, \alpha, \beta)$ for the functions L and R given by formula (32), the value of preference function (33) is expressed by

$$P(\mathcal{E}) = \delta \left(\mathrm{e} + \frac{3(\beta\text{-}\alpha)}{8} \right) + (1-\delta)(e-\alpha) = e - \alpha + \delta \left(\frac{5\alpha + 3\beta}{8} \right). \quad (34)$$

Finally, when two base-station systems characterized by fuzzy performance indexes are considered, the one for which the preference function (34) is larger should be recognized as a "better one".

The preference function may also be used to generalize constraints (16)-(18), treating the real numbers occurring there as three-parameter fuzzy numbers with $\alpha = 0$ and $\beta = 0$.

9. Verification of methodology

The operation of the procedure described in this paper has been positively verified using a computer program written in *Delphi*.

The results obtained were correct both in the case of artificial data (selected tendentiously to achieve intuitionally obvious results) as well as for a small database of the capital city of Warsaw. In the second case, the simplification of data by replacing several subscriber buildings with one properly balanced representation did not significantly change the result obtained. Owing to the averaging properties of statistical kernel estimators, the effect achieved is especially worth highlighting, in particular when large metropolitan areas are studied.

The convergence of the investigated algorithm itself is guaranteed thanks to a finite number of steps carried out at every stage of the method proposed.

While testing the procedure a significant dependence of the calculation time on the parameters m (the number of subscriber buildings available in the data base) and k (the number of potential locations of base stations) was clearly notable. Thus, an increase in the value of the parameter k by 1 results in a growth in calculation time of about 70-percent, while a tenfold increase in m causes a growth of around 6-times. In the case of real values for these parameters, the number of potential base station locations k – even for large cities – is in the teens, whereas the number of buildings in which institutional clients can be located, m, ranges from a few hundred to a few thousand. Moreover, the speed of result generation is influenced by the parameters p (the number of versions of transceivers) and $\tilde{C}$ (the value of available funds). In practice, the value of the parameter p results directly from the number of radio frequencies allocated for use, and – due to their strict rationing and relatively high cost connected with rental – does not exceed 2. Tests carried out on a normal class of PC computer (processor Pentium 233MHz, 64MB RAM), for $m = 1000$, $k = 12$, $p = 2$, and the basic version of the task presented in Section 3, allowed the solution to be achieved in around 3 minutes. Additional increase in calculation time is observed when taking into account shadow areas, as well as, to a lesser degree, the limited bitrate and the fuzzy nature of demand. A long-term planning horizon, however, has notable influence on increase in this time.

The calculations in planning problems are not carried out in real time, and, moreover, they do not require multiple repetitions; therefore, application of this method for tasks with parameters existing in practice is completely possible. The presented procedure may demand – particularly in the case of a long-term planning horizon – the use of computers with relatively large calculation capacity, which are, however, generally applied for GIS packages or database processing.

10. Conclusions

The goal of the research presented in this paper was to develop a practical method allowing for finding an optimal base-station system for the LMDS wireless data transmission system. The respective planning process must take into account a number of constraints related to radio technology conditions, as well as the considerable uncertainty of predictions based on incomplete market data and the resulting investment risk. Due to a combination of all those aspects and the assumptions contained in the operator's strategy, the indication of optimal base-station locations constitutes a very complex problem, while the level of complexity rapidly increases in the case of long-term planning. The methods used in practice until now did not allow for solving such problems, and in many cases, planning decisions were based on intuition.

This paper presents an algorithm that, despite the complexity of the problem, allows to find an optimal system with respect to assumed criteria, within the framework of the existing technical constraints. The algorithm comprises the availability of diverse equipment options, with various parameters (price, coverage, and bitrate), the existence of shadow areas with nominal coverage by transceivers, a possibility of long-term planning, and the existence of demand inaccuracy considered in relation to various operator strategies based on investment security. Owing to the application of statistical kernel estimators, the spatial demand distribution characteristics were averaged, which reduced the method requirements with respect to the potential customer database size, and that in turn is reflected in the efficiency and low costs of the planning process.

The procedure investigated here is based on the elements originating from various fields of science and technology: telecommunications, mathematical statistics, fuzzy logic, operations research, and numerical methods. The procedure presented here is universal in nature and can be easily adapted to related tasks, e.g. planning customer-service points in metropolitan areas.

The current paper comprises the contents of the Ph.D. dissertation of Waglowski (2005) – complete software allowing for the direct use of the methodology presented above is available therein.

References

Berger, J.O. (1980) *Statistical Decision Theory.* Springer-Verlag, New York.

Bosagni, S. (2001) Finding a Maximal Weighted Independent Set in Wireless Networks. *Telecommunication Systems* **18**, 155-168.

Fodor, J. and Roubens, M. (1994) *Fuzzy Preference Modelling and Multicriteria Decision Support.* Kluwer, Dordrecht.

Franceschetti, M., Cook, M. and Bruck, J. (2004) A geometric theorem for network design. *IEEE Transactions on Computers* **53**, 483-489.

Garfinkel, R.S. and Nemhauser, G.L. (1972) *Integer programming.* Wiley, New York.

Gupta, P. and Kumar, P.R. (2000) The Capacity of Wireless Networks. *IEEE Transactions on Information Theory* **46**, 388-404.

Kacprzyk, J. (1986) *Zbiory rozmyte w analizie systemowej.* PWN, Warsaw.

Kulczycki, P. (2000) Fuzzy Controller for Mechanical Systems. *IEEE Transactions on Fuzzy Systems* **8**, 645-652.

Kulczycki, P. (2001) An Algorithm for Bayes Parameter Identification. *Journal of Dynamic Systems, Measurement, and Control* **123**, 611-614.

Kulczycki, P. (2002a) Statistical Inference for Fault Detection: A Complete Algorithm Based on Kernel Estimators. *Kybernetika* **38**, 141-168.

Kulczycki, P. (2002b) A test for comparing distributions functions with strongly unbalanced samples. *Statistica* **62**, 39-49.

Kulczycki, P. (2005) *Estymatory jądrowe w analizie systemowej.* WNT, Warsaw.

Kulczycki, P. and Waglowski, J. (2003) Optimal Base-Stations Locations in the LMDS Wireless Data Transmission System. *Proc. First African Control Conference*, Cape Town, 3-5 December, 375-380, CD: 108.

Kulczycki, P. and Wiśniewski, R. (2002) Fuzzy Controller for a System with Uncertain Load. *Fuzzy Sets and Systems* **131**, 185-195.

Laiho, J., Wacker, A. and Novosad, T., eds. (2001) *Radio Network Planning and Optimization for UMTS.* Wiley, New York.

Lee, C.Y. and Kang, H.G. (2000) Cell Planning with Capacity Expansion in Mobile Communications: A Tabu Search Approach. *IEEE Transactions on Vehicular Technology* **49**, 1678-1691.

Rappaport, T.S. (1996) *Wireless Communications.* Prentice-Hall, Englewood Cliffs.

Schiøler, H. and Kulczycki, P. (1997) Neural Network for Estimating Conditional Distributions. *IEEE Transactions on Neural Networks* **8**, 1015-1025.

Sherali, H.D., Pendyala, C.M. and Rappaport, T.S. (1996) Optimal location of transmitters for microcellular radio communication system design. *IEEE Journal on Selected Areas of Communication*, **14**, 662-673.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Tran-Gia, P., Leibnitz, K. and Tutschku, K. (2000) Teletraffic issues in mobile communication network planning. *Telecommunication Systems* **15**, 3-20.

Tutschku, K. and Tran-Gia, P. (1998) Spatial traffic estimation and characterization for mobile communication network design. *IEEE Journal of Selected Areas in Communications* **16**, 804-811.

Vohra, R.V. and Hall, N.G. (1993) A probabilistic analysis of the maximal covering location problem. *Discrete Applied Mathematics* **43**, 175-183.

Waglowski, J. (2005) Planowanie optymalnego układu stacji bazowych bezprzewodowego systemu transmisji danych LMDS.Ph.D.-thesis, Polish Academy of Sciences, Systems Research Institute, Warsaw.

Wagner H.M. (1975) *Principles of Operations Research.* Prentice–Hall, Englewood Cliffs.

Wand, M.P. and Jones, M.C. (1994) *Kernel Smoothing.* Chapman and Hall, London.