Uncertainty in Modelling and Its Impact on Optimisation Domain; Network Services Pricing

Mariusz Kamola, Piotr Arabas, Krzysztof Malinowski

Research and Academic Computer Network (NASK), ul. Wąwozowa 18, 02-796 Warszawa, Poland {Mariusz.Kamola,Piotr.Arabas, Krzysztof.Malinowski} @nask.pl

Institute of Control and Computation Engineering, Warsaw University of Technology, ul. Nowowiejska 15/19, 00-661 Warszawa, Poland {M.Kamola,P.Arabas,K.Malinowski} @ia.pw.edu.pl

Keywords: QoS pricing, performance optimization, simulation optimization, market modelling, uncertainty.

Abstract

This paper focuses on two problems of modelling used for objective index computation in optimisation. First is the optimisation domain formed by disjoined sets. Such domain can be the result of implicit constraints imposed on some model variables. Second is the problem of modelling randomness of human decision, especially if the information as to the nature of this randomness is incomplete. The model of a commercial simulator of the market of networking products is considered as an example. The conclusions are that more accurate modelling of the randomness can mitigate difficulties presented by the optimisation, and that some universal hybrid optimisation framework for the problems containing inaccurate models should be developed.

INTRODUCTION

The task of designing a market model, i.e. the formulas modelling sales of particular goods, is usually a difficult one, due to a number of reasons. Probably the most demanding of them is the detection of the actual factors that influence customer behaviour, assessing them quantitatively, and finding the way they interact so that purchase decision is made. This is followed closely by another one, the inherent uncertainty of a human decision. When many customers are considered, their decisions average to some mean, but still that mean alone is often not enough to model the dependent economic parameters, as income or cost, correctly. Once the modelling formulas are chosen, there is still the problem of setting their parameters that should be based on the past (or comparable in some sense) sales data. Usually, scarcity of those data and limitations of parameter tuning procedures demand those formulas to remain simple. In such conditions, only a number of well-known general-purpose modelling functions are in use (cf. e.g. [7, 9]).

This task is even mode demanding when the system to be modelled is a market of network services. In such case one often faces the problem that some data for model identification come in great abundance, while others are dramatically few. The former can be e.g. network usage log files, unveiling such subtleties of data flow as the exact time, source, destination and type of each packet transmitted. The latter can be e.g. the number of long-term contracts made by an Internet Solution Provider (ISP) with large customers — usually such contracts are tailored (w.r.t. both their structure and pricing) to customer's particular needs — and are, in some sense, unique. Therefore, model accuracy must match the amount of available real-life data.

Market model construction, tuning and utilisation for optimal pricing of services — in conditions as those described above — was one of the major objectives of Quality of Service and Pricing Differentiation for IP Services (QOSIPS) project. Authors' participation in the project is the main source of observations on the nature of optimisation problems market models can create, and those observations will be presented in this paper. First, market model developed in QOSIPS will be introduced. Then it will be used to produce some examples of optimal pricing problems with unconnected domains. (Unconnected domains consist of disjoined sets, and present difficulties to most optimisation routines.) It will be shown that the existence of such domains remains in close relation to the way an uncertainty is modelled. The paper closes with conclusions as to the procedure that could be applied to treat such difficult optimisation problems in a uniform way.

MARKET MODEL

The main goal of market modelling in QOSIPS was to reflect properly the complexity of ISP products, the market segmentation, sales, network utilisation and the resulting Quality of Service (QoS) experienced by the customers. Appropriately constructed and initialised market model can be then embedded in an optimisation routine that finds prices satisfying ISP's goals, which are mainly short-term and long-term profits. One can also consider using such model as a tool supporting pricing of new products being introduced to the market.

On one hand, the design of the developed model is flexible enough to be applicable in similar branches of economy (e.g. in cellular telephony); this is made possible by hierarchically organised product structure and by the ability to influence every modelled value by any number of other internal model values, indicated freely by the user. On the other hand, it is focused on QoS-related issues, as multiple QoS metrics, which are usually specified by Service Level Agreements (SLA's). Their detailed presentation has been made in [1,2]. Let us occupy here with only those model features which are relevant to the subject of the paper.

Model Dynamics

The market is perceived in QOSIPS as a dynamic system with the number of customers currently subscribed to an ISP as the state variable $\mathbf{x}$. (Actually, since ISP offer can consist of several products, $\mathbf{x}$ is a vector of state variables $x_1, x_2, \ldots$, for subsequent products.) Market modelling means describing behaviour of that dynamic system over some period, given the initial conditions: number of customers $\mathbf{x}(0)$ and the quality of service $\mathbf{q}(0)$ experienced by them prior to simulation start time. The main model parameter is the vector of prices $\mathbf{p}$ for the corresponding products. Simulation is performed in discrete time. In each step, the following are calculated:

- number of new subscribers:

$$x_i^{\text{in}}(t+1) = \alpha_i^{\text{in}} q_i(t) \prod_{k=1}^{\dim \mathbf{x}} p_i^{\beta_{ik}^{\text{in}}} \quad , \tag{1}$$

- number of churning subscribers:

$$x_i^{\text{out}}(t+1) = \alpha_i^{\text{out}} \left(1 - q_i(t)\right) \prod_{k=1}^{\dim \mathbf{x}} p_i^{\beta_{ik}^{\text{out}}} \quad , \tag{2}$$

which make the current number of subscribers as

$$\mathbf{x}(t+1) = \mathbf{x}(t) + \mathbf{x}^{\text{in}}(t+1) - \mathbf{x}^{\text{out}}(t+1) \quad . \tag{3}$$

The coefficients $\boldsymbol{\alpha}^{\text{in}}$, $\boldsymbol{\beta}^{\text{in}}$, $\boldsymbol{\alpha}^{\text{out}}$, $\boldsymbol{\beta}^{\text{out}}$, appearing in (1) and (2), are the scaling ratios and elasticities, and the formulas are modified Cobb-Douglas modelling functions. The modifications make Cobb-Douglas formulas take into account not only the influence of ISP own prices on the client behaviour, but also on the QoS they experienced in the previous step (usually, the previous month).

Modelling Usage of Resources and QoS

Next, in each step the system output is calculated:

- product (e.g. link, application, service — depending on the context) utilisation

$$u_i = \alpha_i^{\text{usage}} p_i^{\beta_i^{\text{usage}}} \quad , \tag{4}$$

- quality of service coefficient

$$q_i(t+1) = \begin{cases} 1 & \text{if} \quad c_i > 1 \\ 0 & \text{if} \quad c_i < 0 \\ c_i & \text{otherwise} \end{cases} \quad , \tag{5}$$

$$c_i = a_i + \textstyle\sum_{k=1}^{\dim \mathbf{x}} b_{ik} x_i(t) u_i \quad ,$$

- profit

$$J_i(t+1) = J_i^{\text{income}}(t+1) - J_i^{\text{cost}}(t+1) \quad , \tag{6}$$

  J^{income} and J^{cost} being certain nonlinear functions of $x_i(t)$, $x_i^{\text{in}}(t+1)$, $x_i^{\text{out}}(t+1)$, u_i and $q_i(t+1)$.

In (4), to calculate the utilisation, the Cobb-Douglas formula is used again. In (5), the quality coefficient (representing the fraction of data or time for which SLA was not met) is modelled by a linear function of weighed total usage of all products. This is, certainly, rather simplified way of representing unloaded, partially loaded and congested states of network infrastructure, but it is done without resorting to complex modelling tools as *ns2* (see [5]).

IMPLICIT CONSTRAINTS IN OPTIMISATION

Typically, ISP is interested in adjusting $\mathbf{p}$ to maximise its performance index over a given period. ISP initialises model parameters (utilising historical data or expert knowledge), and then can compute market response for any $\mathbf{p}$ set by hand, or to start an optimisation routine computing $\mathbf{p}$ which maximises the performance function (usually, the profit). Depending on model parameters, the shape of the performance function surface can vary from a very simple to containing multiple optima, due to repeated switching in (5). (See [1] for more details.)

The constraints imposed on the vector of decision (price) variables can be twofold:

- explicit

$$p_i \in [p_i^{\min}, p_i^{\max}] \quad , \tag{7}$$

  forming a hypercube of the search domain;

- implicit, restricting the number of customers in each month not to fall below a certain limit

$$\sum_{k=1}^{\dim \mathbf{x}} x_k(t, \mathbf{p}) \geq x^{\min}, \quad t = 0, 1, \ldots \quad , \tag{8}$$

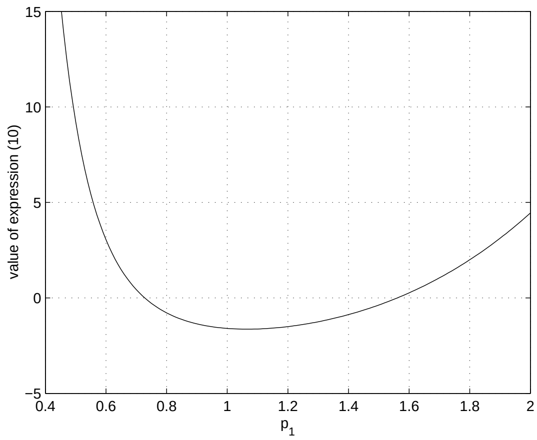

Fig. 1: Graph of expression (10) for $\beta = 2$, $\beta_1^{\text{usage}} = -0.3$ and $a_1 = -1.5$ for $p_1 \in [0.4, 2]$

mapping to the space of decision variables in a complex way, often computable only numerically.

Optimisation in Steady State

In this case the objective is to maximize profit when all transient effects caused by the change of price are already negligible. Therefore, the optimisation problem is to find

$$\arg\max_{\mathbf{p}\in D} \sum_{k=1}^{\dim \mathbf{x}} J_k(t) \quad , \tag{9}$$

i.e. to find prices that maximise profit from all own products for a given month t, sufficiently late after the price change. D is the domain determined by (7) and (8).

As an example, consider still more simple model with just one product offered. Also, assume α_1^{in}, α_1^{out} and α_1^{usage} to be 1, and $\beta \stackrel{\text{df}}{=} \beta_1^{\text{out}} = -\beta_1^{\text{in}}$. In the steady state $\mathbf{x}^{\text{in}}(t) = \mathbf{x}^{\text{out}}(t)$, and this, along with (4), gives

$$x_1 = \frac{\frac{p_1^{-\beta}}{p_1^{-\beta}+p_1^{\beta}} - a_1}{b_{11} p_1^{\beta_1^{\text{usage}}}} \quad .$$

The part of $\partial x_1 / \partial p_1$ that determines its sign is

$$-2\beta + (2a_1 - 1)\beta_1^{\text{usage}} + (a_1 - 1)\beta_1^{\text{usage}} p_1^{-2\beta} + a_1 \beta_1^{\text{usage}} p^{2\beta} \quad . \tag{10}$$

One can easily indicate the parameters that make (10) take both positive and negative values. An example graph of (10) is shown in Fig. 1. Such a shape implies that the set of prices satisfying (8) can be unconnected for certain $x^{\min}$. A situation of this kind is illustrated in Fig. 2.

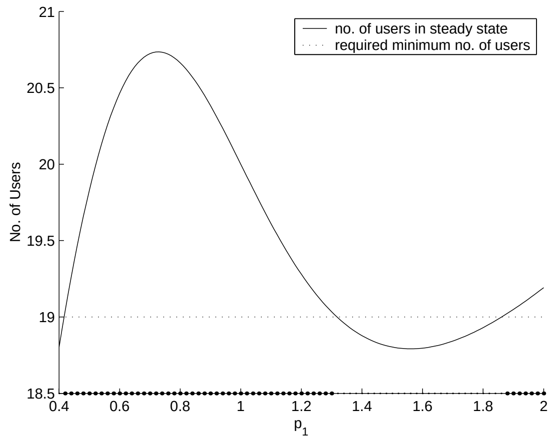

Fig. 2: Graph of the number of users x_1 in the steady state vs. some constraint (8), and the resulting search domain D. Here, $x^{\min} = 19$

Therefore, applying widely recognised modelling formulas, one can get involved in price optimisation with troublesome search domain even with a simple model as presented above. It is worth noticing that the curvature of the graph in Fig. 2 has its source rather in the three power terms ($p_1^{-\beta}$, p_1^{β}, $p_1^{\beta_1^{\text{usage}}}$) that dominate for varying p_1 than in the model internal switching. For the whole range of p_1, the quality coefficient q_1 never reaches neither 0 nor 1.

Optimisation in Transient State

Optimisation of prices for short-term profit maximisation is an alternative for the steady state optimisation. Here, one is interested in profit in relatively short period (3 to 6 months) after changing of prices. Such strategy is not uncommon on the market of new technologies. Merciless exploiting of current customers takes place, with hope to allure them (or others), after several months with completely new products of technological advance.

Minor modifications of the exemplary model given above provide a profile of the minimum number of customers w.r.t. price as the one drawn in Fig. 3 with thick line. Again, the domain D determined by (8) can be not connected for certain $x^{\min}$. This time it is the result of model switching executed when q_1 reaches the value of 1. The decrease of the number of customers w.r.t. the initial value happens for the following reasons. In the central and right part of the graph (i.e. for $p_1 > 0.39$) q_1 is less than 1, and ISP experiences fewer customers in initial months because of small number of new subscribers (when p_1 is high), and in final months because of high churn rate (when p_1 is low, which causes q_1 to grow). In the left part of the graph (for $p_1 < 0.37$), $q_1 = 1$ for all the time considered, and the initial number of customers, aug-

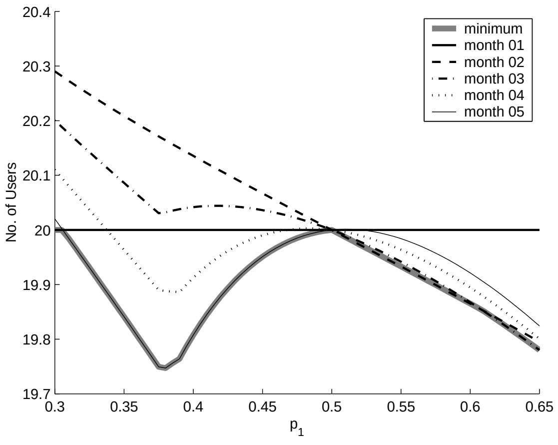

Fig. 3: Graphs of the number of users x_1 in the first 5 months w.r.t. the changed price. The minimum number of users, $\min_{t\in\{1,..,5\}} x_1(t)$ is represented by a thick gray line. Model settings are: $x_1(0) = 20$, $q_1(0) = 0.89$, $\beta_{11}^{\text{in}} = -1$, $\beta_{11}^{\text{out}} = 2$, $\beta_1^{\text{usage}} = -0.2$, $a_1 = -1$, $b_{11} = 0.082$. The initial value of p_1 was 0.5

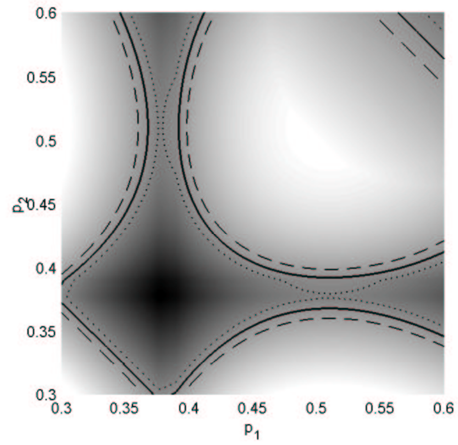

Fig. 4: Contour plots of the minimum number of the total of users an ISP had in the first 5 months. Lighter areas denote higher number of users. The solid line encloses the search domain, which will be used for the optimisation experiments. The dashed and the dotted lines represent alternative shapes of D for different $x^{\min}$

mented with those attracted in the first month by low prices, is dropping steadily with the rate $p_1^{\beta_1^{\text{out}}}$, forming a slope.

In the examples presented so far, the search domain generated by the implicit constraints on $\mathbf{x}$ consisted of two separated intervals. However, for more complex product structures, D is likely to consist of a greater number of unconnected sets. As an example, one can consider two products twin to that described in Fig. 3, offered simultaneously by an ISP to the market. Then, the contour plot of

$$\min_{t\in\{1,..,5\}}(x_1(t) + x_2(t))$$

looks as in Fig. 4 — four separate feasible sets of $\mathbf{p}$ can be generated.

Applied Optimisation Algorithms

The practical market model that QOSIPS dealt with did not present such problems like the unconnected domain shown in Fig. 4. Nevertheless, some of QOSIPS optimisation routines have been adopted to cope with problems of this type. They are CRS and COMPLEX. Both are direct search methods that maintain a pool of trial points, which is updated, to some extent randomly, with new trial solutions. The detailed rules for this random update distinguish CRS from COMPLEX and from a number of similar routines working in this fashion. CRS, in comparison to COMPLEX, is more exploratory, but its convergence rate decreases in the optimum proximity. Therefore, it is suitable for preliminary optimisation. On the other hand, COMPLEX often gets stuck in local extrema, but it has far better support for implicit constraints. (See [8] and [4] for the generic versions of CRS and COMPLEX).

The third routine developed, or rather adopted, for QOSIPS was a local, gradient-based deterministic algorithm that in each step computes the solution of the linearised original problem, and takes the new solution approximation to lie on the line between the LP solution and the solution computed in the former step. This is why it is named SLR, Sequential Linearisation with Relaxation.

CRS, COMPLEX and SLR support implicit constraints and unconnected domains either by the application of penalty functions for infeasible solutions, or by direct modifications of their code, allowing to skip over infeasible regions. They have been run for the system with twin products, with the purpose to maximise profit that was constituted mostly by usage-based income. The contour lines of the profit are presented in Fig. 5. The same figure illustrates solutions found by the three solvers. CRS and COMPLEX were always starting from the corner opposite to the solution. CRS approached the optimal point in all cases, but COMPLEX usually got closer to it, if only managed to get out of the three unpromising regions. Such results comply with the common opinion on the two methods.

SLR has never performed that good. It is suited to handle numerous nonlinear explicit constraints and in this case it exhausted computation budget without significant improvements. It was started 4 times, for $p_1, p_2 \in \{0.3, .5\}$, and its solutions are marked in Fig. 5 with circles that form a square. The fifth circle, much

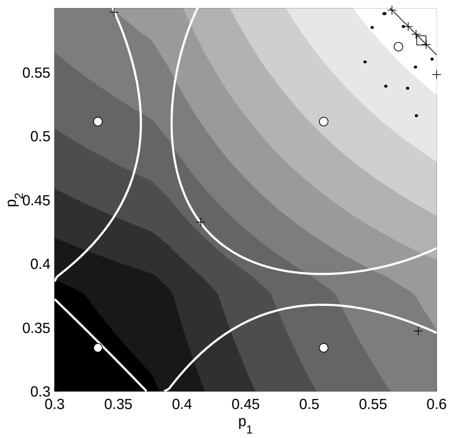

Fig. 5: Contour plots of the revenue function with domain borders indicated (brighter areas denote higher revenue). The optimal point, lying on the upper right border, is indicated by a transparent square. Exemplary CRS solutions are marked with dots; COMPLEX solutions are marked with crosses.

closer to the optimum, was the result of SLR run with much bigger budget (i.e. the number of function evaluations) allowed.

The results are generally satisfactory. However, one instant conclusion appears that performing the initial phase of optimisation with CRS, and then switching to COMPLEX would be the best strategy in this case. Utilisation of other optimisation algorithms was also considered in QOSIPS project. Methods based on the evolutionary strategies seemed good substitutes for the direct search routines, and a suite of gradient-based algorithms also existed that could compete with SLR routine. However, the chosen methods were previously widely and successfully applied by the authors and/or they did not require any extra work on the model, like expressing implicit constraints explicitly. The assumption was made that simulation is an impenetrable process, and the optimisation routines must do with what is available on the simulation output. This approach allows to apply the same optimisation routines to various problems and various simulators but excludes application of Constraint Logic Programming.

UNCERTAINTY MODELLING

In the preceding description of market models the stress has been put on the shape of the optimisation domain, and the uncertainty of the modelled values was apparently suppressed by treating them as crisp, instead of random, variables. However, constructing the model in terms of mean values only is delusive. Consider, for example, the number of customers $\mathbf{x}$ and the QoS coefficient $\mathbf{q}$ influenced by $\mathbf{x}$. Had the relation of $\mathbf{x}$ and $\mathbf{q}$ been linear, it would not matter if $\mathbf{x}$ were deterministic or stochastic. However, $\mathbf{q}$ saturates at 0 and 1 and, strictly saying, treating $\mathbf{x}$ as a mean and forgetting its distribution, can be misleading. The same could be said e.g. of computation of an income based on usage $\mathbf{u}$. The formula for calculating the usage-based income from a single customer is

$$j_i^{\text{income}} = f_i(u_i) \ . \tag{11}$$

If $f(\cdot)$ in (11) is linear then everything is fine — otherwise the mapping changes the type of u_i distribution.

The conclusion can be made that treating the modelled variables deterministically is incorrect. If so, then where to take the distributions from? All one has is the output from Cobb-Douglas-like formulas that contains no additional data as to the modelled variable distribution. Those data can be estimated from historical records, if they are available. As for model parameter tuning, in the branch of networking solutions one can collect many usage-related and few sales-related data. Therefore, it is easier to deduce the distribution of $\mathbf{u}$ than of $\mathbf{x}$.

The modelling rules in QOSIPS have been the result of some compromise between the demand for modelling accuracy and the amount of data available. Therefore, for the computation of usage-based income, $\mathbf{u}$ is assumed to be a random variable with some known distribution. All the other model variables have been assumed to be crisp. Unfortunately, no real-life usage data were available at the time of designing the model, and the design team had to assume some distribution of $\mathbf{u}$, preferably in analytic form, to speed up the calculations. In such conditions the principle of maximum entropy was applied that proposes distribution maximising the uncertainty about the missing information. This rule leads to the usage of exponential distribution.

Supporting the model with as much information as possible has always a good effect. In the case of usage-based costs, the performance function is based on expected values of profit, is much smoother, and is better supported by profit optimisation algorithms. On the contrary, depriving the model of information causes problems. QOSIPS team experienced this in case when there was pressure to deprive the number of customers of its fractional part, for presentation purposes. What the graphs of the total number of customers would look like is presented in Fig. 6. No one would dare to perform price optimisation in such conditions. Neither there would be any sense to do so because such model is severely crippled.

At the point when QOSIPS ended, capabilities of its market modelling were as presented above. However, the research on modelling improvement is still in progress. It seems that when it is impossible to infer about variable distribution from the past data (because there are none), then the principle of maximum

a) b)

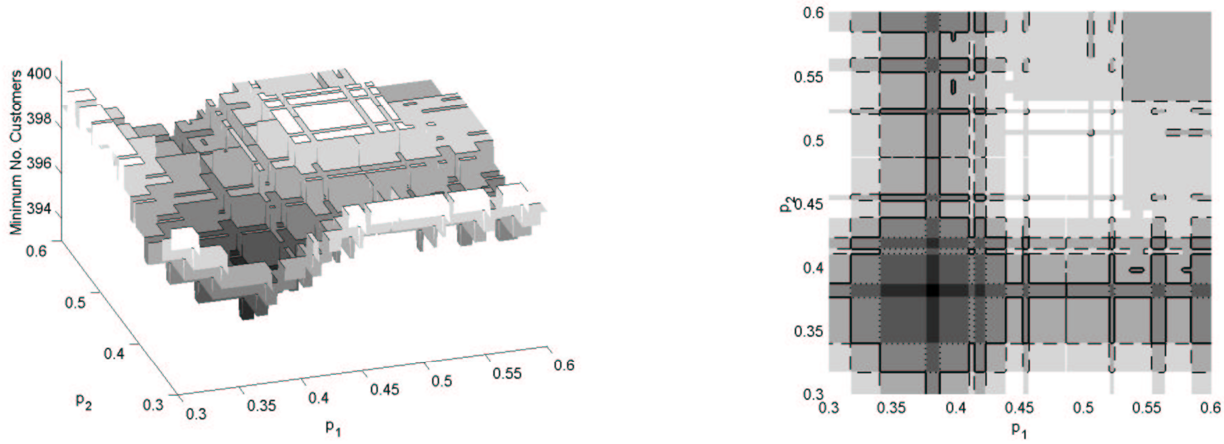

Fig. 6: Graph (a) of the minimum number of customers in the first 5 months (analogously to Fig. 4) and the corresponding search domain (b), for the model with customer number rounding switched on

entropy should be applied — as for **u**. In case of **u**, this emerging indeterminism did not propagate within the model, but in general the whole modelling will become stochastic. There are many routines suitable for solving this kind of problems — for their overview, see [6] and references therein. What differs such model from others is that it is simultaneously stochastic and has unconnected search domain.

CONCLUSIONS

This paper focused on two problems that even a relatively simple market model can cause: unconnected search domain due to implicit constraints *and* indeterminism of modelled variables. The observation can be made that the two problems are likely to appear together, thus creating difficult optimisation problem, rarely addressed in the literature. Most authors suggest to build an explicit model so that the optimisation domain is connected (and, preferably, convex). Unfortunately, there are cases when such modification is impossible, for example due to inaccessibility of modelling algorithm details or model internal variables.

The proposed approach to such problems is to develop an optimisation environment capable of dealing with unconnectedness of domain and with indeterminism simultaneously. Such environment is the goal of the authors' current research. It seems that it ought to contain several generic algorithms already suitable for stochastic problems (e.g. based on evolutionary strategies scheme, controlled random search, and more effective local optimisation routines), adapted for operation on unconnected domains. The crucial part of such environment would be a routine for proper choosing of the sequence of algorithms to attack the problem, and for passing the solution from one algorithm as a starting point for the next algorithm in the sequence.

The basic condition for such environment is to develop a standard of the interface between the model (i.e. the simulator) and the optimisation solver. Firstly, such interface would allow the model to communicate its characteristics and the intermediate simulation values to the solver. Secondly, the solver would be able to intervene and cut unworthy simulations. Such requests for the interface have been already expressed by others in [3].

ACKNOWLEDGMENTS

The work reported in this paper was carried out within the IST Project 1999-20003 "Quality of Service and Pricing Differentiation for IP Services" of the 5^{th} Framework Programme of the European Union. The work reported in the paper was also supported by KBN Polish grant No. 7 T11A 022 20.

REFERENCES

[1] Arabas, P., M. Kamola, and K. Malinowski. 2003. "IP Services Market: Modelling, Research and Reality". Edited by P. Burakowski, B. Koch, and A. Bęben, *Proceedings of Workshop on Architectures for Quality of Services in the Internet jointly held with the Final AQUILA Seminar, Art-QoS.* Warszawa, Poland, 64–75.

[2] Arabas, P., M. Kamola, K. Malinowski, and M. Małowidzki. 2003. "Pricing for IP Network and Services". *Information Knowledge Systems Management* 3, no. 2 (March).

[3] Boesel, J., R. O. Bowden Jr., F. Glover, J. P. Kelly, and R. Westwig. 2001. "Future of Simulation Optimization". Edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, *Proceedings of the 2001 Winter Simulation Conference.* 1466–1469.

[4] Box, M. J. 1965. “A new method of constrained optimization and a comparison with other methods”. *The Computer Journal* 8 (1): 42–52 (April). www3.oup.co.uk.

[5] Fall, K. and K. Varadhan. 2003. *The* ns *Manual*. UC Berkeley, LBL, USC/ISI, Xerox PARC. http://www.isi.edu/nsnam/ns/doc/ns_doc.pdf.

[6] Kamola, M. 2003. “Simulation Optimization - A Survey of Methodologies”. *Applied Mathematics and Computer Science*. to appear.

[7] Lilien, G.L., P. Kotler, and K.S. Moorthy. 1992. *Marketing Models*. Upper Saddle River: Prentice Hall.

[8] Price, W. L. 1987. “Global Optimization Algorithms for a CAD Workstation”. *Journal of Optimization Theory and Applications* 55 (1): 133–146 (October).

[9] Simon, H. 1989. *Price Management*. Amsterdam: Elsevier Science Publishers.

BIOGRAPHIES

Mariusz Kamola, 32, graduated in the Institute of Control and Computation Engineering (ICCE) in 1997. His MSc thesis concerned various aspects of the repetitive control scheme. Currently he is teaching assistant in ICCE and research assistant in NASK. The subject of his dissertation will be “Simulation Optimisation with Ill-Posed Constraints”. His teaching activities concern parallel objective programming, distributed simulation, and IP network technology.

Piotr Arabas, 31, graduated in ICCE in 1996. His MSc thesis concerned parallel multiscenario predictive control. Currently he is teaching assistant in ICCE and research assistant in NASK. Main fields of interest of his research activity are hierarchical control structures, management of computer networks and predictive control.

Krzysztof Malinowski, 54, Prof. of techn. sciences, DSc, PhD, MEng, professor and head of the Research Division of NASK, head of the Network Control and Management Group. He is the author or co-author of four books and over 130 journal and conference papers. For many years he was involved in research on hierarchical control and management methods. His current interests are control, simulation and optimization of communication networks. He was a visiting professor at the University of Minnesota in 1976 and in 1978/1979, and then, from 1981 till 2001, served as a consultant to the Decision Technologies Group of UMIST in Manchester, UK. From 2000 to 2002 he was the project manager, on the Polish side, of the QOSIPS project. K. Malinowski is also an ordinary professor of control and information engineering at the Warsaw University of Technology and a member of the Polish Academy of Sciences.