Pricing for IP Networks and Services

Piotr Arabas, Mariusz Kamola, Krzysztof Malinowski and Marek Małowidzki[*]

Warsaw University of Technology, Institute of Control and Computation Engineering, Warszawa, Poland

ABSTRACT: The issue of pricing for Internet Protocol (IP) networks and services is addressed. The subject of pricing can be viewed from operational and from marketing perspectives. At the operational (network) level pricing aims at congestion prevention and flow optimization. At the marketing level pricing aims at attaining network operator objectives (profit and market share). The paper presents authors' experience in market modeling leading to marketing level price optimization that was the main goal of the project "Quality of Service and Pricing Differentiation for IP Services" (QOSIPS). One of the major conclusions from QOSIPS is that those two pricing perspectives are now perceived separately, while the key element binding them more and more often is the quality of service guarantees, usually defined in form of service level agreements. The paper focuses on techniques of market modeling, product structure, sales and network utilization. Algorithms related to price-based network flow control employing the notion of utility function, are shortly presented in the part dedicated to operational level pricing, so as to complete the picture.

INTRODUCTION

Pricing issues for computer networks became recognized during the last decade as important, perhaps crucial, for further development and exploitation of those networks. After several years of exponential growth of network traffic and vast sums invested in networking business it is now apparent that new instruments are required for recovering expansion charges and as a means of congestion control. Also, as more and more network operators (NO) and network services providers (NSP), in particular Internet Services Providers (ISP), are competing to attract custom, it becomes important to provide means for better client procurement and management. Obviously, pricing of network services is essential in that respect. Further major incentives to consider pricing issues for networks come from emerging differentiation of services and their quality. In fact, it seems obvious that diversified quality of service (QoS) cannot be, for practical economic reasons, introduced before adequate pricing rules are established.

There are several perspectives from which to address pricing for networks. In broad terms it is useful to distinguish between the operational and the marketing perspectives.

The work oriented towards the operational aspect addresses the relationship between prices and traffic management functions, comprising resource provisioning, congestion control (CC) and call (or connection) admission control (CAC). The objective can be, in particular, to make the best possible use of the existing network by finding such prices that result in the desired behavior of the users and in fair allocation of the resources. The time scale at which the decisions concerning prices are made and are repeated can range between seconds or minutes (in the case of responsive pricing for congestion control) and months (in the case of re-adjusting tariffs for usage based billing of different services). The objectives

---

[*] Correspondence: Warsaw University of Technology, Institute of Control and Computation Engineering, ul. Nowowiejska 15/19, 00-665 Warszawa, Poland, Tel.: +48 22 660 7397, Fax: +48 22 825 3719, E-mail: {P.Arabas, M.Kamola, K.Malinowski, M.Malowidzki}@ia.pw.edu.pl

IKS 56

are to maximize revenue or social welfare function subject to limited network resources. It should be noted that most of the existing literature on network pricing is concerned with this operational perspective. The related work is surveyed, in particular, in recent papers by DaSilva (DaSilva, 2000) and Falkner (Falkner et al., 2000).

The marketing perspective is different. Here one recognizes, as the basic issue relevant to a network operator, the need to procure more customers, to maintain them and, eventually, to make them use — and pay for — as many services as possible. The suite of tariffs — including prices of hardware and services necessary to connect a customer to the network, monthly fees (access charges) and usage dependent services – can be then seen as major instruments of client procurement and management. Further, in a longer time frame it can be assumed that the network capacity may be expanded so as to accommodate any foreseeable demand for services from the existing and the new customers. The main concern is then to acquire new customers in a competitive environment, to make them use and pay for multiple services of different type and quality and, finally, not to lose the current customers, i.e. to avoid them migrating to other network operators and service providers. The assumption that network capacity can be increased to accommodate as much traffic as possibly required is presently shared by many specialists (Odlyzko, 2001). They argue that the advances in networking technologies — now in particular the emergence of optical Dense Wavelength Division Multiplexing (DWDM) networks of enormous capacity — make the congestion and call admission control as being not a very relevant subject. It should be noted, however, that this view is not commonly shared; in fact it is considered by many network specialists and researchers as largely over optimistic. The argument, amongst others, is that while the power of computers and computer systems doubles every eighteen months and the overall network capacity doubles every nine months, the demand for network services — in terms of traffic carried through — doubles every four months or so. Also, it is claimed that although capacities of the backbone networks can be vastly expanded this may not be so in case of local access networks and utility servers — leading to constraining overall network capacity.

In the case of a given NO, when marketing decisions have to made in view of available limited network resources, the marketing and operational perspectives merge and should be addressed jointly. In particular, the resource constraints have to be taken into account — either directly or through penalty functions expressing the costs of not meeting service level agreements – when the future profit of a NO is evaluated. It has to be accounted for that by acquiring new customers, as well as by changing the prices of usage-based charged network services, the NO commits the available resources and may then eventually not be able to meet all QoS obligations.

This paper is focused mainly on the marketing perspective by means of QOSIPS tools for long-term network services market modeling. The goal to express products offered by an NSP and their sales not merely verbally but as mathematical formulas leads to systematization of one's perception of both product and market structure. First, one has to state clearly the product structure: which goods are sold separately or bundled, how contract scale and duration affect unit price etc. Second, the market must be properly segmented to group customers of similar behavior. Next, relations must be defined that couple sales with appropriate prices in the model. Finally, formulas not only for sales of the products but also describing their utilization must be specified.

An explicit demand for facts and numbers resulting from mathematical market model makes the NO think over its product structure, and adjust the modeling accuracy to the available data. If the model is such that the data suffice to tune it, then it can be utilized for simulation of future client procurement and optimization of pricing decisions. Details of modeling at market level applied in QOSIPS system are described in Section "Market Level Pricing". Yet, to present the broader view, pricing concepts and mechanisms for operational network management should be also discussed. The basic reason is that, if introduced, such pricing mechanisms may largely influence the client decisions and hence impact marketing of differentiated network services. A few selected possible operational network pricing mechanisms are discussed in Section "Network Operation Level Pricing".
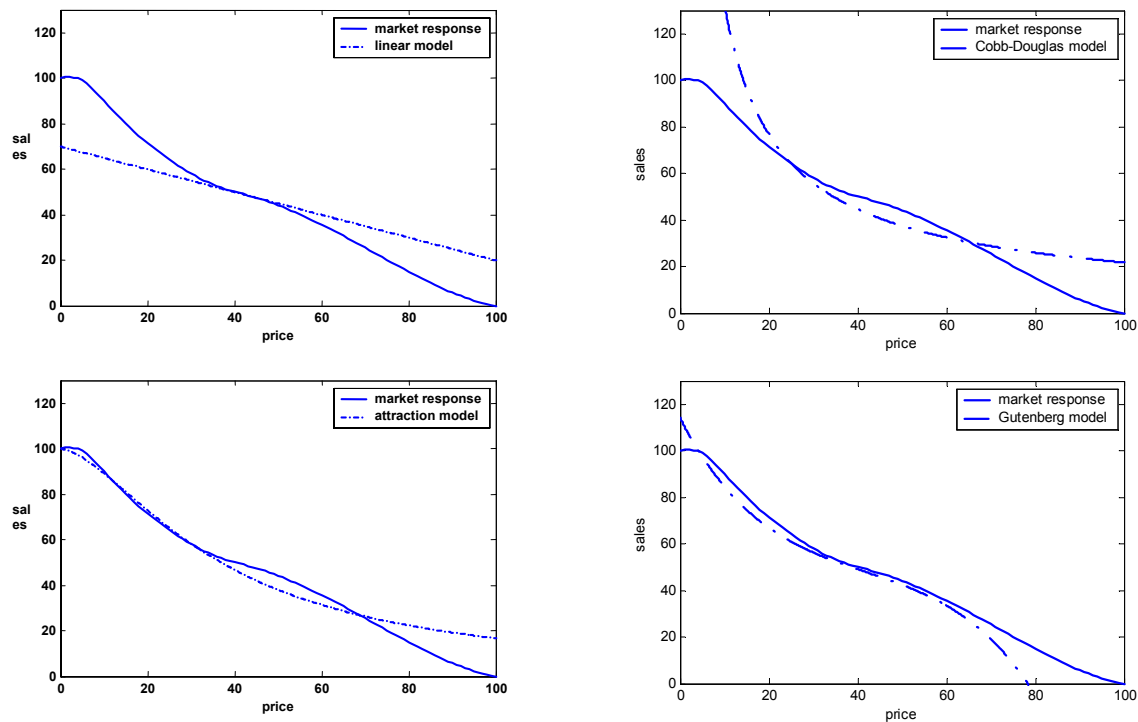
Fig. 1. Market response curve approximated by various modeling functions.

MARKET LEVEL PRICING

Pricing has been considered in QOSIPS system as major instrument for maximizing NO profit at management (i.e. strategic) level. Its importance originates from the fact that price has a very strong psychological impact and therefore is a basic factor taken into account during the process of comparing the competitive offers. To set prices successfully one needs to understand the behavior of consumers that depends on various, often difficult to measure, factors. When constructing QOSIPS model for optimal decision making the authors must choose carefully those factors which influence customers' actions in the most visible way and omit those which would bring nothing more than unacceptable complexity and would make calibration of the model too burdensome or even impossible. As the application of such overcomplicated models might not only slow down the computations significantly but also generate very noisy and impossible to analyze responses, the efforts were concentrated on relatively simple formulas often derived from elementary market models.

Available Models

Modeling of customers procurement and network utilization as a result of price changes is the basic action performed in each step (i.e. time unit, e.g. month) by QOSIPS system for further profit and QoS parameters estimation. As the underlying nature of modeled processes cannot be acutely defined, several models of various scales of complication and precision were applied to allow for the best possible description of the market system. These models reflect some important features of a real market in a way

dependent on accessible data and an appraisal of the marketing situation. They are presented in the order of growing complexity:
1.   linear model,
2.   multiplicative model,
3.   attraction model (not implemented in QOSIPS, described here only for comparison and completeness),
4.   models realising Gutenberg hypothesis.

Comparison of the above models for the task of approximating some real-life market response curve is given in Figure 1.

A linear model employs the simplest mathematical formula, stating that the rate at which the change of price influences the change of sales is constant. For one-dimensional case (i.e. for the situation when the input variable is only price of considered product) it can be denoted as:

$$q = b - ap,$$ (1)

where: q represents volume of sales, p is price and a and b are coefficients. The model is very simple and thanks to that easy to tune by a variety of methods (ranging from expert assessment to well established algorithms such as linear regression). The disadvantage of such a model is that it can only be applicable in a relatively narrow range of input variables — forecasting 10% rise of sales after 5% reduction of prices seems reasonable, but a 100% percent rise after reducing price by half will occur only in some cases, while in others it can result in a greater change, or in a much smaller increase. Another disadvantage is the model symmetry — it treats negative and positive change of prices in the same way, which is usually not true, as higher prices are often more acutely perceived by the clients.

The multiplicative or Cobb-Douglas model is derived directly from the price elasticity definition widely used to describe influence of price changes on various economic factors. The elasticity β is a term used to describe dependence between relative change of two variables p and q, and can be defined as:

$$\beta = \frac{\partial q}{\partial p}\frac{p}{q}$$ (2)

Negative elasticities describe dependence between the volume of sales of the product and its price (lowering the price effects in raise of sales) while positive refer to competitors' prices and are called cross elasticities. The modeling function has a form of the following product:

$$q = \alpha \prod_{i=1}^{N} p_i^{\beta_i}$$ (3)

where: α is scaling factor and β_i are direct and cross elasticities. The great advantage of such a function over a linear function is that it is asymmetric, and thus may reflect the customers' behavior more precisely. Nonlinearity of the function allows its application in a wider region than the linear formula. Identification of Cobb-Douglas functions can be done by linear regression after linearization attained by computing the logarithm of the formula (3):

$$\log(q) = \log\left(\alpha \prod_{i=1}^{N} p_i^{\beta_i}\right) = \log(\alpha) + \sum_{i=1}^{N} \beta_i \log(p_i)$$ (4)

As the method is analytical it produces no additional errors.

The attraction model (Simon, 1989) is based on the hypothesis that the sales are proportional to the attraction of the modeled product, versus the aggregate attraction of all products available on the market, which can be written as:

$$q_i = \alpha_i \, \frac{a_i}{\displaystyle\sum_{j=1}^{N} a_j} \tag{5}$$

where a_i represents attraction of product i and α_i is scaling factor. The attraction can be defined as a way the product is perceived by the clients and is usually linked with utility or value, the last one being determined by the product price. One of possible ways of describing so defined attraction is using multiplicative function. The advantage of such approach is that it allows for modeling of impact of the price on the market share of the product. The s-shape of the function allows for application in a broader range than the Cobb-Douglas formula — it better describes the limitation of ability to attract customers by changing the price beyond some range. Additional limitations are connected mainly with the identification, which has to be carried out by nonlinear methods, as there is no simple way of linearization.

The Gutenberg function is a mathematical implementation of Gutenberg's hypothesis (Simon, 1989) that customers are not likely to react on small changes of prices due to e.g. a kind of loyalty or lack of interest in earning small (unimportant) amounts of money. A possible approach is to modify the modeling function by adding a hyperbolical sine element introducing the "insensitivity area" whose argument is usually the difference between the current product price and aggregate (average) competitors' price. Selecting such arguments allows modeling of the customer's decision process — it is not the price itself but the differences between competing product prices that drives customers. The simplest example of such a function is combination of linear model (1) and hyperbolical sine (Simon, 1989):

$$q = a - bp - c_1 \sinh\big(c_2(p - \bar{p})\big) \tag{6}$$

where: a, b, c_1, c_2 are model parameters and $\bar{p}$ is competitor's price. As hyperbolical sine is symmetric function the model proposed will inherit disadvantage of linear function. To avoid this and attain more flexibility one can combine hyperbolical sine with multiplicative function (3):

$$q = \alpha \prod_{i=1}^{N} p_i^{\beta_i} + c_1 \sinh\big(c_2(p_j - \bar{p})\big) \tag{7}$$

where p_j is price selected for comparison with competitors' price. The main advantage of these models is that they try to imitate customers' behavior in a broader range of input variables (prices). The limitations are connected mainly with the identification procedures, which not only must use more complicated nonlinear methods, but also need more data as the model becomes more complex and has more parameters. Additionally, combining a number of convex, non-convex and s-shape functions may lead to non-convex or multiple minima optimization problems faced by the model identification, and even more often during decision making. A one-dimensional example (Simon, 1989) presented in Figure 2 shows two local maxima of the revenue function. It is obvious that in case of a multidimensional function, the number of local extrema may be greater.

Applicability to Modeling the ISP Market

Modeling the sales of network hardware and services differs significantly from the case of, e.g. grocery products. Network services offered to small and medium enterprises are sold in relatively small number of instances, which impairs the possibility of collecting large amounts of data, making statistical analyses difficult or even impossible. An additional problem is that there is virtually no historic data as ISP market is relatively young, often rapidly expanding, with new products being introduced every year. All this means that the market models should incorporate some knowledge about customer behavior prior to their identification, which is facing a difficult task, as the number of samples may often be insufficient.
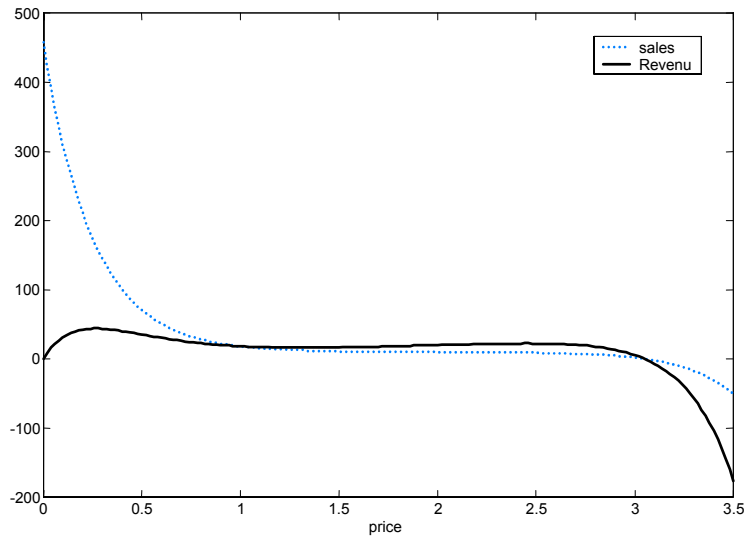
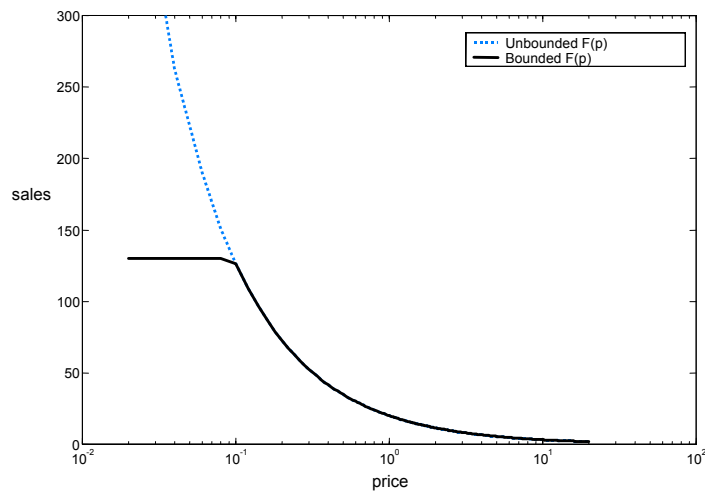Fig. 2. Graph of exemplary revenue function.

Fig. 3. Graph of bounded and unbounded modeling functions.

Another difficulty is connected with narrow validity range of models in situations like that describing the introduction of a new product, which may result in large and even unexpected changes of the market variables. Unfortunately, while introducing new products there is not enough data to tune even the simplest models so another solution must have been applied in the QOSIPS market model. It was the truncation of the output of a modeling function to provide for acceptable (although rough) values of sales for very small prices and thus to overcome the well-known and undesirable behavior of many modeling functions for very small prices without the need of developing a more complicated model. An example graph of such a truncated function is presented in Figure 3.

One can eventually consider expanding of the model applicability range by using functions of more complex shape, allowing for the modeling of more complicated effects like, e.g. super-saturation (Lilien et al., 1992). One of the situations when it can be needed is to account for a possible loss of high standing (image) caused by over-aggressive prices lowering. Unfortunately the behavior of the function outside the surveyed range may be unacceptable as the function is unbounded and usually it must be truncated in a way similar to the one presented above.

Modeling General Trends and Preparing Identification Data

The functions presented above serve in QOSIPS to forecast acquisition of the customers in subsequent time periods, with regard to prices — cf. (QOSIPS, 2001). Moreover, when observing historical records of sales with fixed prices it was easily noticed that the number of new customers differed from month to month. The reasons of such behavior, considered in the market model, are:
- Seasonality i.e. the rise or fall of sales in a specific period of the year (e.g. sales are usually lower during the summer, due to holidays);
- General market growth — through several past years one can observe growing demand on network products as more and more companies require new or extended network services.

Unfortunately, the presented modeling functions are stationary (i.e. they do not depend on time) and using them for the description of time dependent behavior would require preparing a new set of model parameters for every time period. Since such procedure would render the use of many valid historical data impossible, another approach, allowing for use of the same sales model for several time periods, had to be applied. The basic idea was to model an overall customer behavior as a superposition of two sub-models. The first one describes customer reactions to prices using one of the functions presented above and the second is a scenario of seasonality and market growth constructed after consulting an expert.

Complex Products and Typical Pricing Policies of the Providers

Selling the network services involves selling a variety of elements. Some of them may be hardware items like leased lines or routers while others belong to class of software or legal issues (like e.g. selling licenses to use specific application programs etc.). A customer buys a set of items forming a product. Pricing and modeling of demand for such products depends on the way in which they are presented to customers. Two most popular strategies are:
1. pick and mix,
2. predefined packages.

The idea of pick and mix approach is to allow the customer to compose a product by means of choosing a number of appropriate items. This straightforward scheme involves modeling the sales of every individual item — which is hardly manageable due to their large number. Such a strategy deprives NO of an opportunity to perceive products as given entities (i.e. sets of items) and makes customer classification less natural than in cases where those customers are offered the complete solutions. Classifying the customers and products allows for simplification of not only sales models but also for more precise network utilization and QoS modeling.

The second strategy is to prepare a number of complete packages addressed to customers, each associated with a single price. Modeling of such a strategy is simpler as the number of products is much smaller and the construction of models and the collecting of data needed for their identification are less complicated. Unfortunately, such inflexible structure of products is not very attractive for customers who prefer to tailor the services to suit their needs. For that reason, the service providers and network operators usually offer sets of options that can be chosen to customize each package. Precise modeling of these options involves modeling the distribution of customers of a given package across such options,

which poses difficulties — especially during the identification, as the data available for any given package must be spread among the option distribution models. In the case of insufficient data (which is a rule in the case of modeling of newly introduced packages) the best solution seems to be to assume (after consulting the experts) some stationary (i.e. independent on external variable) distributions.

Product structuring applied in QOSIPS is flexible enough to support both abovementioned strategies. Product elements are treated uniformly, regardless off their type (e.g. appliance, link, service, application etc.), and can be arranged freely in hierarchical order, thus forming a tree-like structure. Those elements that are attached directly to the root are subject to direct modeling, like in pick and mix solution. The others that have their parent elements are subject of partial modeling, i.e. only some of their attributes are modeled, and the rest just depends on the parent element attributes.

Impact of Prices on Network Usage

It is obvious that prices may affect network usage by present customers in a similar way as they influence new customers procurement. This fact creates a possibility of using prices to control network utilization without introduction of pricing at the operational level. Carefully designed pricing policies maximize the profit of a provider on one hand (through maximizing the network throughput) and the utility of users (by better utilization of the network and avoiding congestion) on the other hand. The control is achieved by designing such pricing polices that stimulate the users to refrain from overloading the network in the periods of congestion as well as to utilize it more during the off-peak hours.

The simplest policy is flat fees when each user is monthly charged a constant fee independently of the usage. Such a scheme, used often for leased lines, provides no incentive for the users to control their traffic and therefore is useless for congestion control which then must be achieved by other means. Another scheme is usage-based pricing, when the user is charged for a specific amount of connection time (usually for commuted lines), or per specific amount of information transferred. Although means of limiting users activity are provided in this way there is no incentive causing them to differentiate their usage depending upon the state of the network. As measuring the link loads is relatively difficult there are many attempts to construct policy by using a substitute of the network state derived from methods known from traditional (i.e. voice) telecommunication. One of them is applying progressive usage-based pricing schemes in which, in contrast with the previous one, the unit price depends on the amount of transferred data (or utilized time). Such technique allows setting prices such that users do not overload links and prevents discrimination of small users who send small amount of data and are not willing to pay huge bills. A sample relation between usage and fee for two kinds of usage-based pricing is presented in Figure 4. Since progressive usage-based pricing is more universal of the above policies, it has been chosen as standard in QOSIPS system for product elements that can be attributed with usage (as links, applications, and alike).

The drawback of usage-based pricing is that it does not stimulate the users in any way to use network in off-peak time. This can be achieved by providing different pricing schemes for different times of the day. Differentiating pricing schemes w.r.t. to the time of the day can act relatively well (as it is in telephony) provided that the operator can precisely estimate the network capacity and predict the behavior of the users. Unfortunately, both of these tasks are not so easy for the Internet as for telephony. Network capacity in connectionless networks is not so easy to assess as packets may follow various routes and encounter problems inside or outside the provider network. Behavior of the users is also more complex as worldwide companies have departments in various time zones and automated systems generate loads during off-work hours. On short time base the rate of transfer becomes even more chaotic due to rapid changes generated by rate–control algorithms within the Transmission Control Protocol (TCP). All these arguments suggest that an efficient pricing algorithm should react to changes in network load by moving the task of pricing from management to operational level. This problem — lying outside the scope of
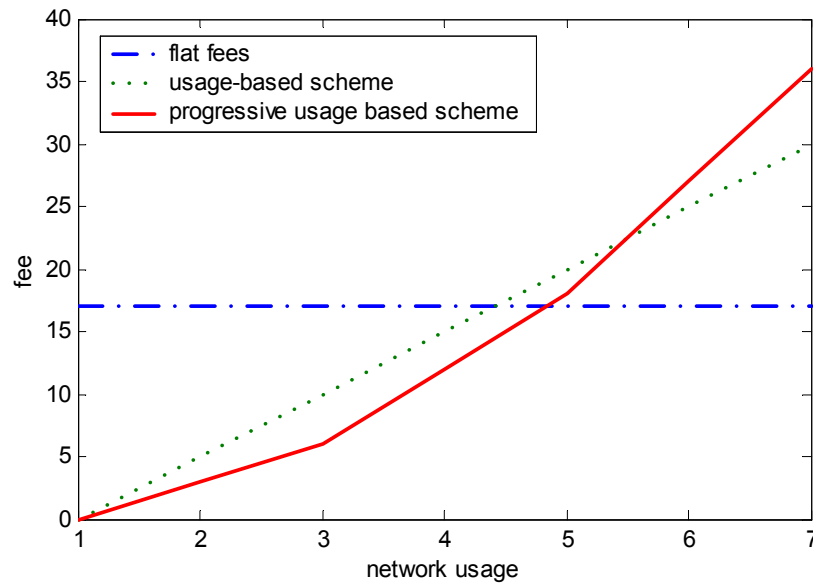
Fig. 4. Pricing schemes.

QOSIPS — urgently demands to be solved. Currently proposed algorithms, closely connected with providing users with quality guarantees and so far definitely disconnected from marketing level perspective, are presented in Section "Network Operation Level Pricing".

Providing and Modeling Quality of Service

Providing stable parameters for a connection service between any two applications running somewhere in the web is not an easy task considering the IP network partition among many operators, limitations of IPv4 and network nodes, security issues, economic issues etc. This section briefly describes the current state of art and the way it affects QoS modeling that was implemented in QOSIPS.

Desired service level is described in the Service Level Agreement (SLA) between a customer and a network provider. SLA may contain many clauses defining various QoS parameters, as there are many ways in which QoS can be defined. Usually, for the convenience of both parties, the values of QoS parameters specified in SLA come from a standard set. This makes network engineering easier and gives to a customer an initial idea what actually can be chosen.

According to Verma (Verma, 1999), connection performance can be specified and measured at network level and at application level. At both levels, two major kinds of QoS metrics can be defined:

- availability — related to for how much time the service (network/application) is available to the customer (e.g. total time that a destination/database is reachable/operable or number of corrupted packets/failed transactions);
- responsiveness — related to how quickly the response is obtained after the request has been made (e.g. roundtrip latency/transaction time).

QoS metrics are the basic bricks for further QoS measuring and modeling. To make QoS modeling easier, it is wise to derive and use some aggregates of basic QoS measurements. These aggregates can model the overall network behavior as well as performance of a specified service in a time window. QoS modeling algorithms running within QOSIPS utilize the notion of QoS aggregates.

Three common situations are distinguished in the literature, and they are reflected in three kinds of SLAs:

- ISP delivers just network connectivity between two or more Network Access Points (NAPs);
- ISP provides a client application with desirable parameters of an access to the server;
- ISP guarantees only QoS parameters of the server end of a connection.

SLAs are supported by QOSIPS system uniformly, without distinction as to their kind. It is assumed that SLA consists of a number of clauses whose violation by an ISP can be described quantitatively by a scalar value $Q \in [0,1]$ — the value of 0 standing for meeting QoS requirement completely, the value of 1 for failing altogether. As clauses concern certain elements of a package, it is assumed that every element may be attributed with several such values, each for the corresponding SLA clause. In the SLA there may exist numerous distinct QoS clauses depending on the same, single metrics. Let us consider the system uptime, for example. The first QoS clause may require the total monthly down time to be less than three hours. Another may require the total daily down time in business hours to be less than 10 minutes. Yet another one may demand the system to be unconditionally operational from 6 pm to 6.30 pm. In this example, QoS clauses are based on different time scales.

Failing to meet a given SLA has explicit and implicit adverse impact on ISP finances. The explicit one comes in the form of compensation that must be paid to the customer. The implicit one is the growing notion of ISP's unreliability that scares off the existing as well as potential customers.

To predict exactly the impact of ISP decisions (connecting new customers, enlarging the bandwidth, restructuring the network) on QoS-related parameters, one would need to have a network model, either mathematical or simulation-driven. Mathematical, open-form modeling requires very strong assumptions about the network structure and the traffic nature, and in fact it is used only to yield rough estimates of the network performance. In practice network configuration is so complicated that open-form models become useless. Various rate control algorithms, diverse router behaviors and, last but not least, the network size would render mathematical models huge and intractable. They may also be incomplete, due to lack of the exact information on the distribution of traffic for each network user.

Therefore, one might turn towards the latter, i.e. simulation-driven, model. A number of commercial (OPNET, 2002) and open-source (Fall & Vardhan, 2002) network simulators exist, most of them up to date with the current IP technology. Running the simulation for long time improves the accuracy of results.

As network simulators are expensive and frequently over-fitted to the goals, a heuristic model seems to be a good compromise. Such model can be based on the knowledge of an experienced network administrator. The administrator is presented with a set of QoS-related metrics, and prompted to assess the values of other QoS parameters. Such approach has been applied in QOSIPS — a linear function with saturation is used to compute the value of Q, aggregated and pre-processed traffic parameters being the model input. The advantages of such a model are its simplicity (only two parameters are needed) and approximate similarity to network behavior.

In the Small and Medium Enerprises (SME) market sector it is hard to think about compensation for QoS failure other than money. Network services in most cases are supposed to be available round the clock, except during a little off-peak time needed for the system maintenance. Therefore, there is no way of compensating the customers with, say, extra hours on-line, as it would be in individual customers sector.

The money paid back in QOSIPS market model depends on the severity of the QoS failure. This overall severity can be reflected by several traffic metrics. For example, if the SLA states that one-way delay may not exceed 100 ms, then the severity is a combination of the number of delayed packets, the amount of delayed data, and the extra delay over 100 ms. Usually, for non time-critical applications (like POS terminals), traffic parameters are not sharp and a minor QoS decrease is accepted (a customer can well wait 1 or 1.1 seconds for the transaction to be completed at the cash desk). There also exist very QoS-sensitive customers that wish to penalize even the slightest QoS failure. Confront Figure 5 for possible penalization schemes.

To verify whether the traffic parameters are conformant to SLA, traffic measurement mechanisms must be available. Generally, the QoS may be measured at network level and at the application level.

As regards network level measurements, most network nodes can currently be configured to interact with the domain administrator via Simple Network Management Protocol (SNMP). Using SNMP network devices can be configured remotely, and queried for their parameters. Available parameters vary depending on the type of network node — for a terminal with an SNMP agent they may be only network adapter and IP stack settings; for more advanced device like router they can also comprise traffic statistics for the internal queues (bytes and packets transmitted/dropped, queue length, packet delays incurred at the router etc.) SNMP, as well as proprietary management protocols, allow to monitor bandwidth, network connectivity, error and loss rates and delays caused by the device.

However, the matter of most interest and causing most trouble are delays and latencies. The simplest tool, `ping` command, measures the time it takes a single Internet Control Message Protocol (ICMP) packet to reach the destination and then to return. Unfortunately, ICMP is subject to additional processing by routers; hence it cannot give an accurate roundtrip latency estimate. Better results can be obtained by utilizing Network Time Protocol (NTP), and even more accuracy is given by specialized tools, synchronized by Global Positioning System (GPS), that inject their own probe packets into the network. This allows also for one-way delay measuring.

QOSIPS, unlike the above solutions, has reached very accurate (1 to 10 μs) delay measures without the injection of probe packets that cause additional network traffic (QOSIPS, 2002). The measurements provided by Ipanema are based only on tracking of the existing IP packets.

As regards the application level measurements, it is difficult to indicate universal QoS monitoring tools due to a vast diversity of applications. The most common practice is to analyze server logs, if the object of interest is a server. Prevalently, application servers record in the logs such events as the system startup, transaction processing time and transaction failures. Next, this information, extracted by an auditor, is processed and presented in the form of periodic reports to the ISP and to the customer.

Even more detailed application behavior statistics are or may be collected by system agents. System agents are background applications that, hooked to the operating system, monitor the selected applications. Some implementations of system agents provide also Application Programming Interface (API) with entry points to monitoring functions. Vendors encourage business application developers to make calls to the monitoring functions as it allows for more accurate QoS measuring, resulting in better customer satisfaction. The negative side effect of a system agents operation is an additional computational load (typically about 5%) that may become the last nail to the coffin of heavily loaded systems. Application of the technique of agents to monitor QoS at the application level was outside the scope of QOSIPS.

Provider Objectives

The short-term goal of any ISP is certainly to profit from the business. Only by satisfying this condition, a company is able to consider any further investments. Except that, every enterprise wants to minimize the risk of being struck by a sudden change of market trends, or by new, fashionable competitors' products. The system being developed within QOSIPS project supports ISPs in decision making that maximizes income, guarantees stability of income, and allows testing of the impact of new products being introduced.

Revenue can be maximized either by increasing incomes or by reducing costs. The incomes and costs in QOSIPS market modeling consist of the constant part (determined by the number of customers and the recurrent fees) and the usage-dependent part. To maximize the revenue, the ISP must encourage the turnover (i.e. the traffic volume) to be high, but not so high that users demand paybacks for QoS deterioration. Such encouragement is accomplished by setting product prices appropriately, and by
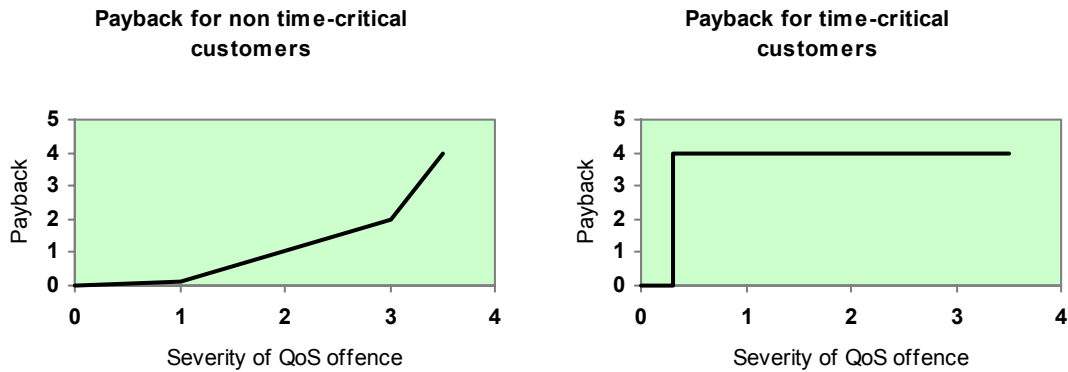
Fig. 5. QoS-based penalization schemes.

promotions. As regards the pricing, it is automated by running optimization routines coupled with the market and the network model. However, promotions must be handled manually. During the promotions, prices are lowered to cause the increased network usage. Customers, suddenly able to afford themselves more services or just more online time, get used to a new lifestyle and keep on transferring more data even after the promotion ends.

Stability of income means in practice many loyal customers, whose number is controlled by prices offered and by the quality and diversity of the ISP products. Therefore, QOSIPS automatic pricing system provides a user with a feature of setting the prices so that the number of customers is always greater than some specified minimum value.

Other means of attracting customers are:

- Introduction of new products — new packages and options must be intermittently introduced and advertised to stimulate customers and make them aware of the ISP presence on the market;
- Dumping prices — that would make the customers come in droves, but causes increased need for network reengineering rather than provide for income stability. ISP's goal is to attract *good* customers, i.e. those who will remain for long;
- Contracts are *a priori* activities — usually signed when new customers are acquired. Customers are given numerous discounts if they declare that they will stay with the ISP for a given period of time, typically 1, 2 or 3 years;
- Loyalty programs are *a priori* activities — targeted at the users that might leave the ISP, rewarding them with bonuses for long or fruitful cooperation.

Out of the above activities, broadening ISP's offer with new products is particularly important, as it keeps the ISP up with the pace of technology. Tools like the QOSIPS system model market behavior and help the sales managers to forecast the effect a new product would have on sales. There are two important rules while introducing a new product:

- it should not compete with the ISP own products that have not yet paid off;
- it should not be identical to the competitor products, thus not allowing the price to be the only means of comparison between them.

As mentioned above, mathematical market modeling in QOSIPS is done mainly for two purposes: i) adjusting prices of the existing products so that certain objectives are reached and ii) forecasting sales and impact of new products on the market. As for the former purpose, the operator is supposed to make quantitative specification of the ISP criteria (e.g. which market segments are crucial, what is the minimum acceptable market share) Then, the optimization routine tries to find the set of the decision

variables (mostly prices) for which the simulated (predicted) market behavior is most promising. Due to the fact that these requirements are sometimes contradictory and that the market model is complex, the optimization routines may easily get stuck without finding a satisfactory solution. Therefore, assistance of an expert to operate the model is advisable. The latter purpose cannot be easily automated. Usually, the sales model for a to-be product is based on that of the already existing that is most similar. However, a new product can differ so much as to render an adoption of some existing sales model impossible. In such cases, querying a set of test customers and supplying them with the new product to obtain an estimate of how it is received seems a good idea.

NETWORK OPERATION LEVEL PRICING

The idea of applying price to network operation has been a matter of much broader interest than market level pricing. Falkner et al. (Falkner et al., 2000) address a broad spectrum of pricing concepts for IP networks. They review the notions of flat pricing, priority pricing, Paris-Metro pricing, smart-market pricing, responsive pricing, expected capacity pricing, proportional fairness pricing as well as other pricing schemes. They also state a number of questions arising from both economic and social welfare points of view, and also from a technical point of view. From their survey as well as from that by DaSilva (DaSilva, 2000) and from other related works (Paschalidis & Tsitsiklis, 2000; Cao & Shen, 2000; Kelly et al., 1998; Low & Lapsley, 1999) quite complex picture does indeed emerge. It looks like that there are no easy solutions and no easy answers. Stahl et al. (Stahl et al., 1998) point out that the pricing problem is far from being solved due to the complexity of the Internet and multiple engineering, economic and social issues involved.

It seems, however, possible to arrive at the following basic conclusions:

- Different concepts of pricing the network services will most likely be adopted for various purposes, various types of users and applications. In particular, small end users, e. g. individual households, will mostly prefer to use flat pricing schemes but may be also enticed to accept simple usage-based pricing, defined as a tariff component — most likely as an easily implementable time-of-day pricing of differentiated services. Clarity and predictability of pricing is and will remain essential to them. On the other hand bigger network players like company networks and service providers (e.g. video servers) may, in their dealing with long haul network operators, accept other types of pricing. They could be willing, in particular to use expected capacity pricing and/or congestion pricing based upon the proportional fairness concept. They will also have to choose and pay for differentiated classes of network transporting services;
- Assuming that usage-based pricing will become more widely used, it seems that network operators will be charging the traffic sources — as it would be otherwise very difficult to recognize at the network level who should pay for traffic carried during a particular, say TCP, connection or for User Datagram Protocol (UDP) transmission. If the actual beneficiary of a given traffic stream is at the receiving end, then the traffic source should be able to recover the transmission costs by charging them — through the billing scheme — to the receiver.

Flow Optimization Algorithms with Distributed Price Coordination

Let us now present an exemplary approach to network level pricing — namely, a suite of algorithms for network flow optimization with distributed price coordination.

For the purpose of presenting pricing mechanisms that could be used for congestion control in a backbone network, consider the basic flow optimization problem that can be related to Internet pricing and to so called socially fair pricing. Recently an optimization approach to flow control has been

proposed, where the control algorithm is derived as a projection method to solve the dual of a global optimization problem (Low & Lapsley, 1999). They propose both synchronous and asynchronous versions of this algorithm and provide the convergence results. In (Low, 2000) two important extensions to the basic algorithm are considered. One of them consists of the modification whereby the link prices are computed using only local information (the local buffer queue length). This modification can be, in fact, related to the concept of coordination by price instruments with feedback, i.e. to the Interaction Balance Method with Feedback (Findeisen et al., 1980; Malinowski, 1977). The second modification is concerned with the possibility of having multiple paths between a source-destination pair. Price instruments are also used in the approaches proposed in (Courcoubetis et al., 1996) and in (Kelly, 1997; Kelly et al., 1998). The basic differences between the proposed mechanisms consist of how and by whom the prices for traffic rates are computed and then who decides upon the transmission rates. The flow optimization problems are essentially the same. Likewise, in (La & Anantharam, 2000) such optimization problem is considered and an interesting decentralized pricing scheme is introduced. In this scheme network users propose prices they are willing to pay per time unit. Special, (p,1)-proportionally fair algorithm (Mo & Walrand, 1998) is then used for TCP under which user transmission window sizes converge. Then the users can estimate the price per unit flow and to compute the new prices they are willing to pay, and so on. The advantage of this scheme when compared to those proposed by Low and Lapsley or by Kelly is that no extra signaling between traffic sources and network routers is necessary. Also, transmission window size control is applied rather than direct transmission rate control. On the other side, the distributed flow control is by a two-level mechanism and thus likely to converge more slowly.

Basic Flow Optimization Problem and Price Algorithm

Let us present now briefly the formulation of the flow optimization problem and the distributed pricing algorithm as proposed in (Low & Lapsley, 1999).

A network consisting of a set of $L=\{1,...,L_n\}$ unidirectional links is considered. In the basic model each link has capacity c_l, $l \in L$. The network is shared by a set $S=\{1,...,S_m\}$ of traffic sources; source s is defined by a four-tuple ($L(s)$, $U_s(\cdot)$, $x_{s,min}$, $x_{s,max}$). $L(s) \subseteq L$ is a subset of links that source s uses to transmit information to one or more destinations at the egress points of the network. $U_s(\cdot)$ is source utility function defined over interval $I_s = [x_{s,min}, x_{s,max}] \subseteq R_+$, with the values in R, where $x_{s,min}$, $x_{s,max}$, with $x_{s,min} < x_{s,max}$, are, respectively, minimum and maximum transmission rates that source s may wish to transmit and $U_s(x_s)$, for $x_s \in I_s$, is utility attained when this source transmits at rate x_s. For each link l let $S(l) = \{s \in S: l \in L(s)\}$ be the set of sources that use this link. Observe that $l \in L(s)$ if and only if $s \in S(l)$. Let $I = I_1 \times ... \times I_{Sm}$. A simple network structure with three paths of data sent by S_1 and S_2 is shown in Figures 6a and 6b.

The basic Flow Optimization Problem can be then formulated (Low & Lapsley, 1999) with the objective to choose source rates vector $x = (x_s, s \in S)$ so as maximize the sum of source utilities:

FOP:
$$\max_{x_s \in I_s} \sum_s U_s(x_s)$$
$$\text{subject to} \quad \sum_{s \in S(l)} x_s \leq c_l, \quad l \in L \tag{8}$$

If the feasible set is nonempty (i.e. if $\sum_{s \in S(l)} x_{s,min} \leq c_l$) and the performance function is strictly concave — each $U_s(\cdot)$ is strictly concave over I_s — then the unique maximizing solution $\hat{x}$, called the primal optimal solution, exists.
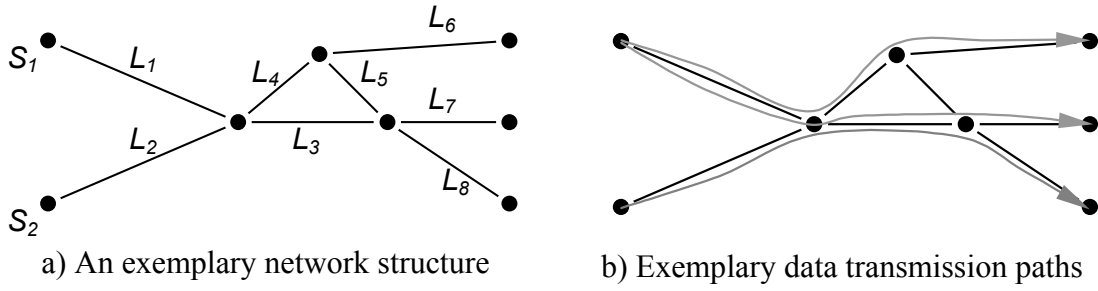
a) An exemplary network structure   b) Exemplary data transmission paths

Fig. 6. Sample network.

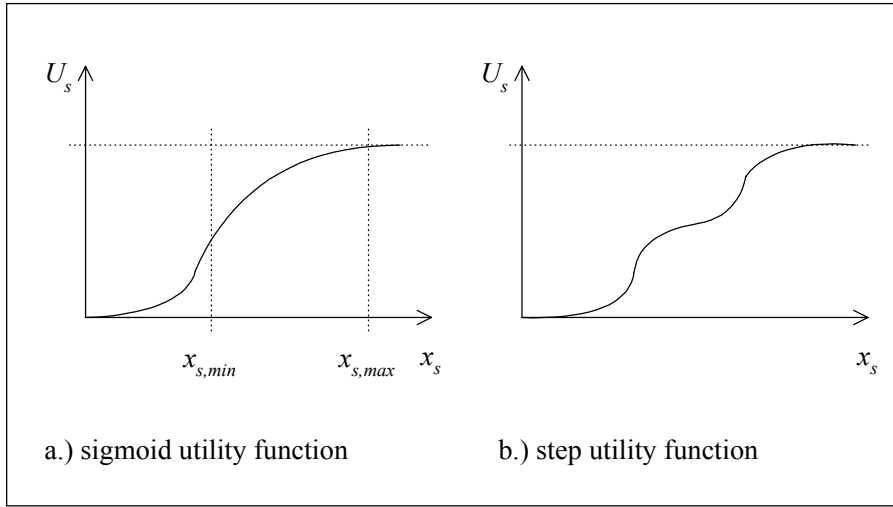a.) sigmoid utility function   b.) step utility function

Fig. 7. Exemplary utility functions.

It should be observed at this point that most common utility functions will be simply increasing, possibly having a sigmoid shape as shown in Figure 7a. However, a utility function also might show a step pattern, as shown in Figure 7b. This may happen if a user would like to use services with different quality-of-service requirements (in this case, QoS factor is capacity). Note that the two cases correspond somehow to the notion of QoS paybacks in the marketing perspective of pricing (cf. Figure 5).

Still, the utility functions may often be treated as strictly concave within the $[x_{s,min}, x_{s,max}]$ range, assuming that $x_{s,min}$ is correctly chosen. The minimum rate, $x_{s,min}$, should be guaranteed in the SLA made between a user and a network operator. Moreover, a user should pay a flat fee for traffic within $[0, x_{s,min}]$ range and make extra payments for extra traffic within range $[x_{s,min}, x_{s,max}]$.

The basic Flow Optimization Problem (8), with additively separable objective functions and capacity constraints, is a particular instance of a problem which can be solved by a dual method — using price coordination (Lasdon, 1970; Findeisen et al., 1980; Tamura & Yoshikawa, 1990).

The Local (Source) Problems are:

$$\text{LP}_s: \quad \begin{aligned} &\max_{x_s \in I_s} \quad U_s(x_s) - p^s x_s \\ &\text{where} \quad p^s = \sum_{l \in L(s)} p_l \end{aligned} \tag{9}$$

Each source can, independently from others, solve the above local problem for a given price p^s; it is important to note that the local utility function $U_s(\cdot)$ may not be known to other users and to the network operator as well. Solution of LP_s is denoted as $x_s(p^{ȝ})$ and the associated optimal value of LP_s objective as $B_s(p^s)$.

The dual problem to FOP, defined through the solutions of LP_s, $s = 1,...,S_m$, is the key to the distributed algorithm for adjusting prices p^s. The basic distributed synchronous link algorithm (Low & Lapsley, 1999), which, in fact, is just the basic descent algorithm for dual function minimization, is as follows:

$$A1_{link}: \qquad p_l(t+1) = [p_l(t) + \gamma(\sum_{s\in S(l)} \hat{x}_s(p^s(t)) - c_l)]_+ \qquad (10)$$

where $[y]_+ = \max(y,0)$. In the above Eq. (10) $p_l(t)$ denotes the value of the l-th link price at iteration instant t; the same notation is used for $p^s(t)$. Thus, in the presented synchronous version of the distributed price adjustment algorithm all sources receive, at a given time t, prices $p_l(t)$, then they compute respective source prices $p^s(t)$ and the solutions of LP_s, $s = 1, ..., S_m$. The obtained values of source rates $\hat{x}_s(p^s(t))$ are then signaled to links, where the new values of link prices $p_l(t+1)$, $l = 1, ..., L_m$, are computed according to link algorithm $A1_{link}$; the iteration index is advanced by one and so on. This basic algorithm requires full time synchronization — new values of source rates and link prices should be computed only after all information bearing the current iteration index (time index) is signaled through the network.

Convergence of the basic algorithm can be easily established using general results available for coordination by price instruments (Findeisen et al., 1980).

Modifications of Flow Optimization Problem and Price Algorithm

Source decision algorithms (solving LP_s for $p^s(t)$, $s = 1, ..., S_m$) and distributed pricing strategy may not be well suited for real-life applications in data networks. First, routing of traffic from a given source may be done through multiple paths and both these paths and the splitting of traffic stream x_s may vary over time. Second, current real traffic rate over a particular link, say x^{*l}, can be directly monitored at that link; this information is valuable and should be used to adjust the link price. Third, when the network becomes congested, distributed flow control rules of shaping and policing (Armitage, 2000) result in changes of flow rates and distribution of traffic within the network. Finally, transmission rate control through dynamic pricing as presented above must definitely be associated with a particular time scale — sudden traffic bursts, especially in data networks, should be allowed to happen and should not disturb other flow rates.

Low (2000) made an important contribution to the further development of dynamic pricing algorithm (10) by taking into account on-line observations of real traffic parameters and by considering traffic routing through multiple paths. However, as regards the routing, fixed possible paths are still assumed and it is the source that is expected to decide about splitting the transmitted packet stream amongst those paths.

Next, Modified Flow Optimization Problem was proposed in (Malinowski, 2002). In this problem it is explicitly recognized that the actual traffic flows can be different from those resulting from routing through fixed paths and also different from those that could be calculated by sources from the basic network model. This may, in particular, happen due to shaping and policing schemes used by routers and due to traffic fluctuations over observation time window within which the traffic rates are evaluated. It is assumed possible that due to routing operations such as tunneling, using Multiprotocol Label Switching (MPLS) or other tunneling protocol, traffic from source s can be divided over several paths. It is also assumed that the routing rules may be unknown both to sources and, due to the distributed nature of the network, also to routers at the network links.

In the Modified Flow Optimization Problem it is also proposed that the total capacity c_{*l} of each link is

divided between *steady-state flow capacity* c_l and *headroom* h_l equal to $c_{*l} - c_l$, where $c_l < c_{*l}$, $l \in L = \{1,...,L_n\}$. Leaving nonzero headroom capacity is well adopted — in fact unavoidable — practical means to deal with fluctuations of traffic and, eventually, with link failures. It allows also for using early congestion notification. It is assumed that steady-state link capacity constraint — equivalent to link capacity constraint in FOP (Eq. 8) — is allowed to be violated temporarily during network operation. The violations may result in packet drops or in marking and rerouting of some packets. Note again that $c_l = c_{*l} - h_l$, with headroom capacity $h_l > 0$.

CONCLUSIONS

In spite of optimistic forecasts (Odlyzko, 2001) Internet users experience, and probably will go on experiencing, problems with the network. Possibly most of the trouble is caused not by the throughput itself, but by improper web design and configuration, or by deliberately inefficient forwarding between competitive ISPs' domains. Regardless of the reason, the users demand fault free service.

Sections "Market Level Pricing" and "Network Operation Level Pricing" underlined how important and complex it is to elaborate appropriate pricing from the marketing and the network operation perspectives. The role of proper pricing structure is particularly important at the marketing level, as it forces the NO to rethink and to define properly product structure, scope of price/sales/usage interaction, and modeling resolution. However, the goals defined at marketing level (e.g. profitability, market share, income stability) differ from those at the network operational level (e.g. congestion avoidance).

Operational level pricing seems to be one of the best (if not the only one) ways to prevent network congestion and provide full utilization of available hardware. However, both NO and users demand simple, easy to manage mechanisms. Such a situation is caused by variety of reasons, two main being the conservativeness (or kind of momentum) and the will to operate in stable economic conditions. Although the telecommunication market is growing rapidly and new technologies arise every year it is often a huge problem to introduce them, or if eventually introduced to utilized fully their capabilities due to need of providing backward compatibility with older systems. This can be e.g. said of version 6 of Internet Protocol that would be helpful in implementing price driven mechanisms as well as providing better quality guarantees. Similarly many NOs and especially users prefer simple flat rates guaranteeing stable income for the provider and no need for difficult budget calculations for users. Usage based pricing is still considered difficult due to problems with providing correct measurements and quality which in this scheme becomes even more delicate issue (nobody wants to pay for e.g. retransmissions caused by poor network performance). Such a situation suggests little success in practical use of operational level pricing for congestion control as resulting fees are not only usage based but also varying frequently, making themselves difficult to control by standard (i.e. human oriented) methods. Some propositions concern providing users with software agents which can act on their behalf during transfer price negotiations basing on some easy to define policy like keeping link prices in some limits (e.g. refraining from transmitting data when price rises above 0.01 € per MB) or observing user's budget in a time window. It is also widely stated that prices paid by users should not be directly linked to prices proposed by congestion control mechanisms being treated rather like signals for rate control mechanisms. Such procedure will allow to avoid too low (usually zero) prices during proper functioning of the network and too high in case of untreatable congestion.

The place where issues related to operational level behavior meet marketing level instruments is the construction of the SLA. The significance of SLA at marketing level strategies emerges from the fact that it defines the cost of the service for the user and the income for the operator. This means that all usage-based, quality related and operation pricing mechanisms should be covered by SLA.

According to its scope, the QOSIPS system implemented QoS control by means of much simpler instruments. First of all, precise statistics and measurements of quality parameters were provided with the help of Ipanema's system of non-intrusive devices. The QoS guarantees were covered by a simple scheme — by means of compensations paid to the users by NO in case of not meeting specified quality parameters. The only way the NO can influence the users behavior is through progressive pricing schemes which provide kind of substitute for direct state of the network feedback which is closer to common charging policies known from every day life (e.g. telephony) and thanks to it easier to accept by NO and its customers. The experience of QOSIPS project shows that even such simple solution should improve operation the network and thus provide NO higher revenue.

One may find inconvenient the fact that no universal algorithm exists that would charge users for the traffic and, simultaneously, prevent them (and NOs) from QoS deterioration. Current algorithms for congestion control seem to not care about regular usage based pricing — they impose prices only while in a congested state. Hence, there are prices of two kinds: the "real" one, for the data transmitted, that a user must really pay, and the shadow ones that serve only for transmission rate control. Binding the two perspectives, or rather making market level prices influence operational level prices and vice versa, still remains an open task for future research. Development of such unified pricing strategy was outside the scope of the QOSIPS project. Nevertheless one of project participants' major conclusions was that solving the task would be really a scientific big advance, and the authors of this paper felt strong need of communicating the problem to the public.
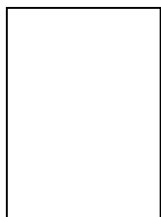
ACKNOWLEDGMENTS

REFERENCES

Armitage, G. (2000). *Quality of Service in IP Networks: Foundations for a Multi-Service Internet*. Indianapolis: Macmillan Technical Publishing.

Cao, X.-R. & Shen, H.-X. (2000). Internet Pricing: Comparison and Examples. *Proceedings of 39th IEEE Conference on Decision and Control* (CD ROM), 12–15 Dec. 2000. Sydney.

Courcoubetis, C., Siris, V.A. & Stamoulis, G.D. (1996). Integration of pricing and flow control for ABR services in ATM networks. *Proceedings of Globecom'96*, Nov. 1996 pp. 644–648.

Cross, R. (1997). *Revenue Management; Hard-Core Tactics for Profit-Making and Market Domination*. London: Orion Business Books.

DaSilva, L.A. (2000). Pricing for QoS-Enabled Networks: A Survey. IEEE Communication Surveys, http://www.comsoc.org/pubs/surveys.

Falkner, M., Devetsikiotis, M. & Lambadaris, I. (2000). An Overview of Pricing Concepts for Broadband IP Networks. *IEEE Communication Surveys*, http://www.comsoc.org/pubs/surveys.

Fall, K. & Varadhan, K. (Eds.) (2002). The *ns* manual. http://www.isi.edu/nsnam/ns/ns-documentation.html.

Findeisen, W., Brdyś, M., Malinowski, K., Tatjewski, P. & Woźniak, A. (1978). On-line hierarchical control for steady-state systems. *IEEE Transactions on Automatic Control*. AC-23, 189–209.

Findeisen, W., Bailey, F.N., Brdyś, M., Malinowski, K., Tatjewski, P. & Woźniak, A. (1980). *Control and Coordination in Hierarchical Systems*. London: J. Wiley.

Kelly, F.P. (1997). Charging and rate control for elastic traffic. *European Transactions on Telecommunications*. 8, 33–37, http://www.statslab.cam.ac.uk/~frank/elastic.html.

Kelly, F.P., Maulloo, A.K. & Tan, D.K.H. (1998). Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of Operations Research Society*. 49(3), 237–252, http://www.statslab.cam.ac.uk/~frank/rate.html.

La, R.J & Anantharam, V. (2000). *Charge-Sensitive TCP and Rate Control in the Internet*, http://www.eecs.berkeley.edu.

Lasdon, L.S. (1970). *Optimization Theory for Large Systems*. New York: MacMillan.

Lilien, G.L., Kotler, P. & Moorthy, K.S. (1992). *Marketing Models*. Upper Saddle River: Prentice Hall.

Low, S., & Lapsley, D.E. (1999). Optimization flow control, I: basic algorithm and convergence. *IEEE/ACM Transaction on Networking*. 7(6), 861–874.

Low, S. (2000). Optimization Flow Control with On-line Measurement or Multiple Paths, preliminary version entitled Optimization flow control with on-line measurement. *Proceedings of the International Traffic Conference*. vol. 16, June 1999.

Malinowski, K. & Ruszczyński, A. (1975). Application of interaction balance method to real process coordination. *Control and Cybernetics*. 4(2).

Malinowski, K. (1977). Applicability of the Lagrange multiplier method to optimization of quadratic systems. *IEEE Transactions on Automatic Control*. AC-22(4), 648–651.

Malinowski, K. (1987). Steady-state systems: on-line coordination. In: *Systems and Control Encyclopedia*. Oxford: Pergamon Press pp. 4566–4571.

Malinowski, K. (2002). Optimization Network Flow Control and Price Coordination with Feedback; Proposal of a New Distributed Algorithm. *Computer Communications*. 25(11–12), 1028–1036.

Mesarovic, M.D., Macko, D. & Takahara, Y. (1970). *Theory of Hierarchical, Multilevel, Systems*. New York: Academic Press.

Mo, J. & Walrand, J. (1998). *Fair end-to-end window-based congestion control*. Technical Report M988/46, UCB/ERL, http://www.path.berkeley.edu/~jhmo.

Odlyzko, A. (2001), *Internet growth: Is there a „Moore's Law" for data traffic?*, http://www.dtc.umn.edu/~odlyzko/doc/networks.html.

OPNET Modeler Main Page (2002). http://www.opnet.com/products/modeler/home.html.

Paschalidis, I.Ch. & Tsitsiklis, J.N. (2000). Congestion-Dependent Pricing of Network Service. *IEEE/ACM Transactions on Networking*. 8(2), 171–184.

QOSIPS Deliverable 3.2.1 (2001). *Research report on how pricing model, learning and optimization are to be implemented to satisfy requirements*.

QOSIPS Project Web Page (2002). http://qosips.ia.pw.edu.pl.

Simon, H. (1989). *Price Management*. Amsterdam: Elsevier Science Publishers.

Singh, M.G. & Bennavail, J.C. (1993). PriceStrat: a knowledge support system for profitable decision making during price wars. *J. Information and Decision Technologies*. 19, 227–246.

Stahl, D.O., Whinston, A.B. & Zhang, K. (1998). A Simulation Study of Competitive Internet Pricing: AOL Flat Rates versus GSW Usage Pricing. *Proceedings of ICE'98, Charleston* pp. 68–76.

Tamura, H. & Yoshikawa, T. (Eds.) (1990). *Large-Scale Systems Control and Decision Making*. New York: Marcel Dekker.

Verma, D. (1999). *Supporting Service Level Agreements on IP Networks*. Indianapolis: Macmillan Technical Publishing.

Photos + Biosketches?