Service Provider Revenue Dependence on Offered Number of Service Classes

Vesna Radonjić and Vladanka Aćimović-Raspopović

Abstract — In this paper possible applications of responsive pricing scheme and Stackelberg game for pricing telecommunication services with service provider as a leader and users acting as followers are analyzed. We have classified users according to an elasticity criterion into inelastic, partially elastic and elastic users. Their preferences are modelled through utility functions, which describe users' sensitivity to changes in the quality of service and price. In the proposed algorithm a bandwidth management server is responsible for performing automatic optimal bandwidth allocation to each user's session while maximizing its expected utility and the overall service provider's revenue. The pricing algorithm is used for congestion control and more efficient network capacity utilization. We have analyzed different scenarios of the proposed usage-based pricing algorithm. Particularly, the influence of the number of service classes on price setting in terms of service provider's revenue and total users' utility maximization are discussed. The model is verified through numerous simulations performed by software that we have developed for that purpose.

Keywords — service class, user's elasticity, Stackelberg equilibrium, responsive pricing.

I. INTRODUCTION

In terms of an efficient service provider (SP), a pricing scheme should enable optimization of available network capacity and effective network resource utilization. Therefore, there is a need for shifting from simple charging schemes such as duration based or flat rate based charging (i.e. static pricing schemes) towards the usage based charging (i.e. dynamic pricing schemes) [1]. Prices, thus become effective means of traffic engineering in terms of short-term (seconds, minutes) and/or mid-term (hours, days, weeks) control of congestion in the network. On the other hand, customers' requirements for quality of service (QoS) within a network can be very different. Charging can provide efficient service differentiation according to QoS criteria, in a way that traffic flows with similar characteristics join the appropriate class of service, depending on the QoS requirements. This means that users pay an appropriate fee for additional QoS guarantees.

In this paper we use a responsive pricing scheme for which we have developed an algorithm based on the Stackelberg game model [2]. In the proposed model users are classified in accordance with the elasticity criterion. The model can be performed for scenarios with different numbers of service classes and it provides dynamic allocation of required bandwidth in order to avoid network congestion. Calculations are performed by bandwidth management server (BMS), which is specially developed for that purpose. BMS automatically performs allocation of available bandwidth simulating users' responses to price per bandwidth unit imposed by the SP according to their individual utility functions without further users' involvement. In that way the proper network operation and fulfillment of service level agreement (SLA) between SP and user can be ensured. We analyze the dependence of SP's revenue and total users' utility on the service price and the number of offered service classes.

The paper is organized in the following way. In Section 2 we briefly discuss related work in the field of network pricing which motivated our research. Users and service provider optimization problems are explained in Section 3. The proposed algorithm is given in Section 4. Simulation results are presented and analyzed in Section 5 and concluding remarks are given in Section 6.

II. RELATED WORK

The research of network pricing methods has been introduced in the early 1990s with basic idea to use a pricing policy as an efficient, feasible and fair way to adjust demands to network resources [3]-[8].

There are numerous references treating the problem of eliminating congestion via a proper bandwidth assignment considering the capacity constraint. In [9] a decentralized scheme with congestion prices as means for achieving a fair and efficient sharing of bandwidth is proposed. Knowing the history of the pricing process, users attempt to maximize their individual utility by choosing actions based on prices prediction.

The responsive pricing scheme is proposed with an aim to incorporate feedback between telecommunication provider and its users. It describes a dynamic strategy imposed by a service provider, illustrating a manner of its exploitation the adaptive nature of users for increasing economic and network efficiency [10]. Price is emphasized as an alternative means for congestion control to ensure proper network operation and in particular to guarantee different QoS levels. Responsive pricing is based on the assumption that users are adaptive and...
respond to price levels [11]. In the case of high network utilization, resources are stressed and the SP increases the prices for the resources. Adaptive users then reduce the traffic offered to the network. Similarly, in the case of low network utilization, the SP decreases the price and adaptive users increase their offered traffic.

In [12] responsive pricing with resource flexibility is considered. Effects of demands variability and correlation, assuming normally distributed demand curve are described. In [13] a responsive pricing scheme is used in the context of mutual payments between SPs. In [14] a model for incorporating pricing in next generation networks with users sharing bandwidth under a fixed charge per bandwidth amount is presented.

On the other hand, game theory provides a good basis for defining the pricing system. Various game models are proposed for pricing telecommunication networks. For modelling the interaction between SPs and users some important facts should be considered, such as the expectations and aspiration of SPs to maximize their revenues, the fact that SPs and users interests are opposed and that they are unequal participants in the game. This context corresponds to the idea underlying the Stackelberg game. The Stackelberg model is considered as a leader-follower competition model. It is a strategic game that can be played with either price leadership or quantity leadership.

Stackelberg game is originally proposed in [15]. It is a game theory model based on monopoly as an economic phenomenon. In recent years, a dynamic version of the Stackelberg game is frequently used in the field of pricing telecommunications services, where a typical application of this game includes SP as a leader, which takes the first move and users as followers, who choose their strategies according to SP’s strategic choice. In [16] Nash equilibrium is determined as a solution of Stackelberg game with the network (the leader) and a large number of users (followers). A similar model is applied to connection oriented networks [17], where an algorithm for determining the optimal service prices in next generation networks is developed. The authors of [18] consider two cases: the Stackelberg game with independent followers and the combination of Nash and Stackelberg game where the followers’ moves are mutually dependent. Pricing algorithms based on the Stackelberg game, assuming one or more providers (leaders) and one or more users (followers), with different requirements in terms of traffic routing are proposed in [19]. Application of Stackelberg game in the process of negotiation between providers and users of a wireless network is proposed in [20], where users are allowed to choose the type and characteristics of the service, as well as the duration of the service level agreement. The authors of [21] have discussed the network and information requirements of Stackelberg game in which the role of leader was assigned to the network manager, who can also access the network as a user, while users act as followers. For this model, Stackelberg equilibrium is determined instead of Nash equilibrium, which is presented as a common solution of Stackelberg game in previous literature.

III. TWO-STAGE OPTIMIZATION

In our model, the problem of determining service prices is divided into optimization with respect to users’ utility and the optimization of service provider’s revenue.

A. Users’ Optimization Problem

According to the elasticity criterion, users can be classified into three categories: inelastic, partially elastic and elastic users. Users’ preferences may be modelled with utility functions, which describe users’ sensitivity to changes of QoS. A user’s utility function varies in accordance with the elasticity criterion of a user. For all users’ types, QoS is defined by bandwidth, \( \theta \) obtained from the SP.

Inelastic users are users who have strict requirements in terms of delay but can tolerate losses to some extent. Their bandwidth demands vary at a specified interval (between \( \gamma \) and \( \pi \)). Inelastic user’s utility has been most commonly described by a sigmoid function (Fig. 1) [22]:

\[
U(\theta) = \frac{m}{1 + e^{\frac{m}{\gamma} - \theta}}, \quad \gamma < \theta \leq \pi.
\]

It is considered that the user is willing to pay a maximum price \( m \) per bandwidth unit.

![Fig. 1. Inelastic user’s utility function.](image)

A utility function which best models elastic and partially elastic users’ behaviour is a generalization of the logarithmic function [14], [23] - [26]. Elastic users do not tolerate losses but can accept a delay to some extent. Partially elastic users are also not tolerant of losses but they have stronger requirements in respect to delay [2].

Depending upon the QoS requested, each partially elastic user would require a minimum bandwidth \( \gamma \). Less bandwidth than \( \gamma \) on average is of no utility to the user [17]. The law of diminishing marginal utility ensures that a user derives the same amount of satisfaction from any bandwidth more than the maximum \( \pi \) (Fig. 2). Partially elastic user’s utility function can be expressed as follows [2]:

\[
U(\theta) = m\gamma\left(\log\left(\frac{\theta}{\gamma}\right) + 1\right), \quad \gamma < \theta \leq \pi.
\]
For elastic users (Fig. 3) only the maximum required bandwidth, $\pi$ is defined [17]. As opposed to inelastic and partially elastic users, bottom bandwidth limit doesn’t have to be defined. Elastic user’s utility function can be expressed as follows:

$$U(\theta) = m \cdot k_i \cdot \log(1 + \theta), \quad 0 < \theta < \pi,$$

(3)

where $k_i = 1/\log(1 + \pi)$.

The following differential equation has to be solved:

$$\frac{d}{d\theta}(U(\theta) - M\theta) = 0.$$

(5)

Net benefit for an elastic user is as follows:

$$U(\theta) - M\theta = m \cdot k_i \cdot \log(1 + \theta) - M\theta.$$

(6)

Hence, the bandwidth that maximizes the net benefit of an elastic user is:

$$\theta^* = \frac{\gamma + \pi}{2} - \ln \left( \frac{m - 2M \pm \sqrt{m^2 - 4aM}}{2M} \right).$$

(7)

Based on (7) it is obvious that this problem has no unique solution. However, intuitively it is clear that a user always selects the solution that provides him more bandwidth for the same price. It is in this case the solution with the minus sign.

Net benefit for a partially elastic user is as follows:

$$U(\theta) - M\theta = m \cdot k_i \cdot \log(1 + \theta) + 1 - M\theta.$$

The value $\theta$ for which the partly elastic user’s net benefit attains a maximum is:

$$\theta^*(M) = \frac{m \cdot k_i}{M}.$$

(9)

Net benefit for an elastic user is as follows:

$$U(\theta) - M\theta = m \cdot k_i \cdot \log(1 + \theta) - M\theta.$$

(10)

The bandwidth maximizing the net benefit of an elastic user is:

$$\theta^* = \frac{m \cdot k_i}{2M} - 1.$$

(11)

Knowing the rules of pricing scheme, users have no motivation to give other than their true valuation of the maximum price per bandwidth unit.

For each user, bandwidth and price parameters are parts of the SLA. Each user gives his parameters of desired bandwidth and maximal price i.e. $\gamma$, $\pi$ and $m$, that can’t be changed during the agreement period. The bandwidth allotment to user $i$ is denoted as $\theta_i$ and that is his best response to the price a SP determined for half an hour period. If he sets a maximal price $m$ to be higher than he is willing to pay for bandwidth unit, his utility decreases. If he sets a maximal price $m$ to be lower than he valuates a bandwidth unit, he might not receive a desired amount of bandwidth, i.e. QoS he desired. Moreover he could lose his sessions during periods of network congestion. Therefore, the dominant strategy of each user is to give his true valuation of the maximum price per bandwidth unit.

Although it is common that the agreement between a SP and user for service consumption has to be settled for a long time period, in our model allocations are performed in short time scales. In practice, it would be impossible for each user to update his bandwidth in a short time scale. Instead of that, updates are performed by BMS, based on users’ parameters of desired bandwidth and maximal price, considering network congestion and network capacity utilization with the aim of maximizing a SP’s profit.

B. Service Provider’s Optimization Problem

We assume that there is only one critical link in the network, concerning high bandwidth utilization. It is a reasonable assumption in a properly designed network [27]. Such a link usually does not behave badly all the time, but only under certain worst-case conditions.

The SP’s revenue $T(M, \theta)$ is a function of the market price and the bandwidth allocated to various users:

$$T(M, \theta) = M \sum_{i=1}^{N} \theta_i(M).$$

(12)

This function is assumed to be monotonically increasing and strictly concave. The appropriate market price is achieved by solving the optimization problem, which involves maximizing SP’s revenue as a function of price, considering the bandwidth constraint:

$$\max_u T(M, \theta^*) = \max_u M \sum_{i=1}^{N} \theta_i(M), \sum_{i=1}^{N} \theta_i \leq C, M \geq 0$$

(13)

$C$ is the total capacity of the critical resource link.
The optimization is performed on a single critical link for each service class using the responsive pricing scheme and Stackelberg game model [2], [28] - [30]. Our research is based upon Stackelberg price leadership model. In our model, players are a service provider acting as a leader and users acting as followers. Both SP and users behaviour is supposed to be rational.

The SP initializes his algorithm by assigning an initial price $M^*$ either randomly or based on historical data. The price offered by the SP, which maximizes its profit, along with the best response bandwidth of each user constitutes the Stackelberg equilibrium - that is:

$$M^* = \arg \max_{M} T(M, \theta(M)). \quad (14)$$

IV. PRICING ALGORITHM

We developed a usage-based pricing algorithm for the responsive pricing scheme where users are charged according to bandwidth consumption. The algorithm can be performed for scenarios with different numbers of service classes in order to optimize bandwidth usage, price per bandwidth unit and a total capacity of critical link in the network. We assumed that each user uses a fixed amount of bandwidth that is associated with the class of service he chose, according to his needs and network price. However, in each time unit, the service class user belongs to can be changeable because of dynamic changes of price that depend on a network congestion level. On the observed link, the total number of users is $N$ (Fig. 4).

There are $S$ rounds in total and in each round $s$, $l_s$ iterations are performed, where $s = 1, 2, \ldots S$. Each round $s$ consists of the following iterative steps:

Step 1: For the critical link capacity, $C_s$ and the fixed bandwidth $\theta_j$ provided to every user belonging to service class $j$, network price is proposed by the SP: $M_j^{os}$ for $s = 1$ and $M_j^{os} = k_s M_j^{min(s-1)}$, where $M_j^{min(s-1)}$ is a minimal considered price for a service class $j$ in a round $s - 1$ for $s = 2, 3, \ldots S$, $k_s = rk_{s-1}, 0.8 \leq r < 1$ and $M_j^{os} > 0$.

Step 2: For the price $M_j^{os}$ BMS calculates a desired bandwidth for each user $i$ applying for a desired class of service $j$, $\theta_i^{os}$ ($j = 1, J$, $i = 1, N$) based on equations (7), (9) and (11), for inelastic, partially elastic and elastic users, respectively.

Step 3: BMS simulates user behaviour in the following fashion: After his needs, user $i$ chooses one class $j$ and is willing to pay $M_j^{os}$ for a service of chosen class if $\theta_i^{os} \leq \theta_j$; user $i$ is not willing to pay $M_j^{os}$ for the same service if $\theta_i^{os} > \theta_j$ and he applies for a service of class $j'$ such that $\theta_i^{os} \leq \theta_j$ and $\theta_i^{os} = \max\{\theta_i^{os}, \ldots \theta_i^{os}\}$; $j' = 1, J, j' \neq j$.

Step 4: For each class $j$ calculations $N_{ij}^0 \theta_j$ are performed, where $N_{ij}^0$ is the number of users such that $\theta_i^{os} \leq \theta_j$, $j = 1, J$.

Step 5: If $\sum_{j=1}^{J} N_{ij}^0 \theta_j < k_s C_s$ new prices for each service class $j$, $M_j^{is}$ are calculated:

$$M_j^{is} = (1 - k_s) M_j^{os}, \quad 0 < k_s < 0.1$$

and it crosses over to a new iteration with new prices $M_j^{is} < M_j^{os}$. Factor $k_s$ points to a high level of capacity $C_s$ utilization (we choose $0.95 \leq k_s \leq 1$).

Step 6: If $\sum_{j=1}^{J} N_{ij}^0 \theta_j > C_s$ the SP sets new prices $M_j^{is}$ for each service class $j$, $M_j^{is} = (1 + k_s) M_j^{os}$, $0 < k_s < 0.1$ and it crosses over to a new iteration with new prices $M_j^{is} > M_j^{os}$.

Step 7: If $k_s C_s \leq \sum_{j=1}^{J} N_{ij}^0 \theta_j \leq C_s$ total sums $T^{0s} M_{ij}^{os}$ and $\sum_{j=1}^{J} M_{ij}^{os} N_{ij}^{os}$ are calculated and it crosses over to a new round $s + 1$.

```
\text{Fig. 4. Allocation of available bandwidth between } N \text{ users that belong to } J \text{ service classes.}
```

The simplified illustration of this algorithm is shown in Fig. 5. In each round, iterations are performed with different values of the critical link capacity, in such a way that $C_1 < C_2 < \ldots C_s$. The SP initializes his algorithm by assigning initial prices based on historical data. In each next step initial prices are decreased by factor $k_{s+1}$ so in the round $(s + 1)$: $k_{s+1} = rk_s, 0.8 \leq r < 1$. After $S$ rounds, graphs $T$ and $U$ for values $\theta$ and $M$ satisfying condition $\sum_{j=1}^{J} N_{ij}^0 \theta_j \leq C_s$ are created. The point $M_l = M_{los}$ for $\max_{M} T(M, \theta^{0s})$ gives the Stackelberg equilibrium price for the lowest QoS class (i.e. service class $J$). For each higher class equilibrium prices are obtained by multiplication with a certain factor denoted as $p_l$.

We have demonstrated in [2] that the algorithm converges quickly after a few tens of rounds.
For each service class $j$ SP proposes price $M_i^j$ for $i=1$ and $M_i^{15} = M_i^{15+1}$ for $i=2,3,\ldots,5$.

For $M_i^j$, BMS, for each user computes desired bandwidth ($j=1,2,\ldots,J$)

A user $i$ chooses the highest service class $j$ within condition $J_j^i = \max_j$ and for each class $j$, $N_j^i$ is computed.

What is the utilization of the critical link?

Insufficient

High

Excessive

For $i=S$, the end

Fig. 5. Pricing algorithm.

V. SIMULATION RESULTS ANALYSIS

For the purpose of carrying out simulations of the pricing algorithm, we developed software in C#. The software is performed by BMS that automatically allocates the available bandwidth according to the state of network capacity utilization and their individual utility functions.

Algorithm and network parameters can be varied. We chose the following values: $k = 0.05, k_i = 0.85, k_v = 0.95$, $N = 100$ and $C_s = 100000$MB. Initial price and users’ price parameters differ depending on service class and users’ elasticity. In this paper we present simulation results for four scenarios of algorithm 2, i.e. for $J = 2, 3, 4, 5$. We chose the following values for initial price for the lowest service class: $M_i^1 = 0.5 + 1.5MU$ and prices are multiplied by factor $p_1$ for each higher service class: $M_i^j = p_1M_i^{j-1}, j = 1,2,\ldots,J$. Similarly, we chose maximal prices users are willing to pay for the lowest service class to be $m_1 = 1 + 1.5MU$ and prices are multiplied by factor $p_2$ for each higher service class: $m_i = p_2m_{i-1}, j = 1,2,\ldots,J$. Values $p_1$ and $p_2$ differ depending on the number of service classes SP offers to his users. Those values are shown in Table 1.

<table>
<thead>
<tr>
<th>$J$</th>
<th>$p_1$</th>
<th>$p_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.46</td>
<td>1.36</td>
</tr>
<tr>
<td>3</td>
<td>1.21</td>
<td>1.17</td>
</tr>
<tr>
<td>4</td>
<td>1.14</td>
<td>1.11</td>
</tr>
<tr>
<td>5</td>
<td>1.10</td>
<td>1.08</td>
</tr>
</tbody>
</table>

SP’s revenue and total users’ utility graphs for simulations of the proposed algorithm are presented in Figs. 6 and 7, respectively. From Fig. 6 it can be noticed that the optimal price per bandwidth unit increases with the number of service classes SP provides to his users: $M_{opt}^2 = 0.9, M_{opt}^3 = 0.9, M_{opt}^4 = 0.97, M_{opt}^5 = 0.97$, for the lowest QoS class and scenario with 2, 3, 4 and 5 classes, respectively. Optimal prices per bandwidth unit for each higher service class are slightly higher which is obtained by multiplication with a value of $p_1$ in the presented results of simulations. That results from the fact that with the higher service classes, users’ valuation of parameters $\gamma, \pi$ and $m$ is higher.

We can perceive that the highest value of the SP’s revenue is obtained for the case with four service classes (SP’s revenues in that case are: $T_2^* = 12251.71, T_3^* = 12037.80, T_4^* = 13692.78, T_5^* = 12078.87$, for the scenario with 2, 3, 4 and 5 classes, respectively). Total users’ utility as a function of price and number of different service classes is presented in Fig. 7. It can be observed that the values of total users’ utilities that correspond to the optimal price are similar in different scenarios (i.e. $U_2^* = 5512.26, U_3^* = 5511.98, U_4^* = 5211.80, U_5^* = 5719.44$, for scenario with 2, 3, 4 and 5 classes, respectively). Thus, in terms of total users’ utility, it is best to use a scenario with five service classes. Bandwidth distributions among users for different scenarios are presented in Table 2.

In all performed simulations of this algorithm, bandwidth utilization is very high (over 90 percent). The total optimal capacity in each scenario is very close to the maximal value (i.e. $C_{opt} = 9800$ in all cases). Better results can be achieved with a higher value of maximal capacity meaning that SP needs more bandwidth to satisfy his users’ demands. However, complexity of the algorithm enlarges in case with more than 3 service classes with capacity optimization which dilates the time of algorithm execution. Therefore the possibility of price and capacity optimization performed separately should be considered.
according to the elasticity criterion. Within each user type responsive pricing scheme in which users are classified uses the Stackelberg game model that is adjusted to the proposed pricing algorithm is applied. The algorithm utility on the price and number of service classes in case telecommunications provider’s revenues and total users’ can be concluded that the SP is achieving the highest offered service classes. Based on the performed analysis, it can be concluded that the SP is achieving the highest revenue if the scenario with four service classes is offered to his users. In terms of total users’ utility, it is best to use a scenario with five service classes.

In addition to the price and bandwidth optimization, the critical link capacity optimization is performed. However, the complexity of the algorithm indicates that in future research the possibility of performing capacity optimization independently of price optimization should be considered.

VI. CONCLUSION

This paper considers the dependence of telecommunications provider’s revenues and total users’ utility on the price and number of service classes in case the proposed pricing algorithm is applied. The algorithm uses the Stackelberg game model that is adjusted to the responsive pricing scheme in which users are classified according to the elasticity criterion. Within each user type (inelastic, partially elastic and elastic users), classification referring to QoS differentiation is derived. We analyzed results for several scenarios with different numbers of offered service classes. Based on the performed analysis, it can be concluded that the SP is achieving the highest revenue if the scenario with four service classes is offered to his users. In terms of total users’ utility, it is best to use a scenario with five service classes.

REFERENCES


