❖ Scientific
Research

Analysis of a Pricing Method for Elastic Services with Guaranteed GoS

Marcos Postigo-Boix, José L. Melús-Moreno

Department of Telematics Engineering, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain.
Email: marcos.postigo@entel.upc.edu, teljmm@entel.upc.edu

ABSTRACT

Service Providers (SPs), which offer services based on elastic reservations with a guaranteed Grade of Service (GoS), should know how to price these services and how to quantify the benefits in different scenarios. This paper analyzes a method for evaluating the price of a service based on elastic reservations with a guaranteed Grade of Service. The method works as follows: First, the SP determines the requirements of the service that wants to offer; Second, the SP evaluates the average rate of the accepted elastic reservations of the service with a guaranteed GoS; Third, the SP calculates the price that guarantees the GoS with an aggregate demand function that depends on a demand modulation factor of the elastic reservations that is the mean reserved bandwidth, B_{res}; and Finally, the SP obtains the optimum value of the elasticity of the reservations that gives the maximum revenue, and the required access bandwidth in this case. The paper not only applies the method to a class i of elastic reservations when a linear-based demand and a revenue function are selected, but it also analyzes the influence of each one of the considered parameters. This method could be extended to the case of multiple classes of independent and guaranteed elastic services, applying the method to each service with its estimated demand and revenue functions.

Keywords: Elastic Reservations; Streaming; GoS; Mean Reserved Bandwidth per Accepted Request; Aggregate Demand Function; Pricing; Revenue

1. Introduction

Service Providers (SPs) want to estimate the revenue of the services that they provide that usually depends on the applied price to the offered services. Currently, the services that the SPs present treat to cover a wide spectrum of profiles in the aim to adjust them to the preferences of their different users. In that sense there exist users that request elastic services that could be delivered with variable bandwidth, that is, they assume that not always they could receive the same bandwidth for the requested service (the bandwidth reservation for the service is elastic and it fluctuates between a minimum and a maximum values). One example of this type of elastic service is the delivery of streaming video flows with different compression levels. Users want to get high-quality for their reservations, but also they could accept some tolerable degradation in the quality of their reservations if the reduction of the price for this service is significant.

Elastic reservations require the support of new signaling mechanisms other than the most commonly used today, the resource ReSerVation Protocol (RSVP) [1]. As an alternative, the Next Steps in Signaling (NSIS) [2] protocol family allows to reserve bandwidth in a specific range. The Internet Engineering Task Force (IETF) created the NSIS Working Group in 2001 to solve new signaling needs for reservations. Since then, several Internet RFCs and papers have been published [3,4], including the QoS NSIS Signaling Layer Protocol (QoS-NSLP) that describes the procedures to signal QoS reservations between a *Desired QoS* and a *Minimum QoS*. In our scenario each one of them will respectively represent the bandwidth that the user wants to reserve and the minimum bandwidth that the user needs to work properly.

Figure 1 shows the proposed scenario for the reservation of elastic services using the QoS-NSLP-based signaling mechanism. According to **Figure 1**, when the user wants to watch a video he access to the website where the SP lists their offered SLAs. In **Figure 1**, the SLA is defined by the Grade of Service (GoS) and other parameters such as the *Desired* and the *Minimum QoS*, which are respectively the highest (H) and lowest (L) bandwidth reservations for the service, and the elasticity of the reservations (ξ). In this paper this parameter, for a class i, is defined according to (1). Thus, if the elasticity of the reservations is 0 the *Desired* and the *Minimum QoS* have the same bandwidth and the elasticity is 1 if
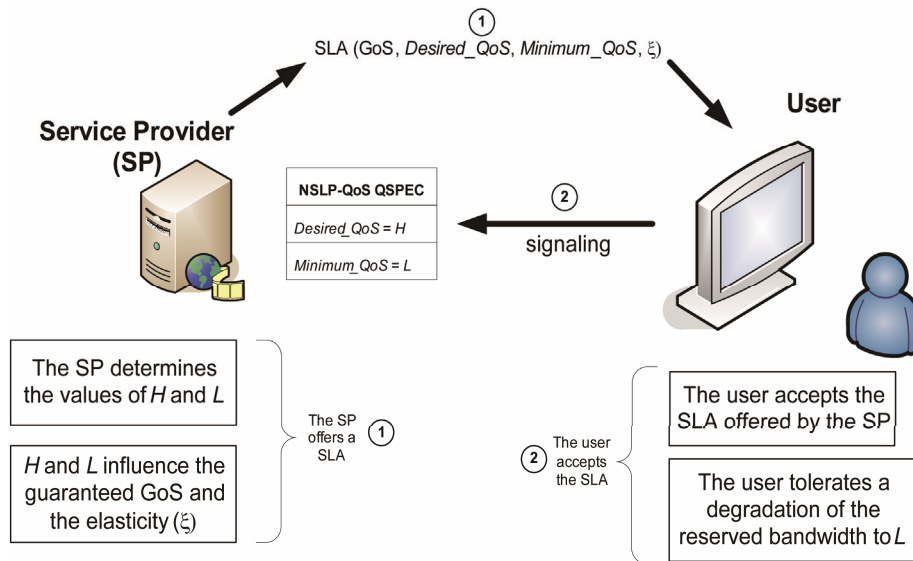
Figure 1. Scenario of the reservation of elastic services based on the QoS-NSLP signaling mechanism.

the *Minimum QoS* bandwidth is 0. In the scenario of video content distribution of **Figure 1** the SP determines both values, the highest-quality (H) and the lowest-quality (L). Thus the elasticity for the reservations of class i would be:

$$\xi_i = 1 - L_i / H_i \qquad (1)$$

In this scenario, another important parameter that helps users to qualify and to differentiate among SPs is the reserved bandwidth per accepted request $B_{res,i}$. It represents the effectively reserved bandwidth of class i for each user within its specified range, that is, between the H_i and L_i. In addition, the metric $\overline{B}_{res,i}$ defines the mean reserved bandwidth per accepted request, which establishes the mean size of the reservations of class i in the requested range. This metric represents a demand modulation factor of the accepted reservations in the sense that users would desire that the SP offered the value of this metric closer to H_i.

According to **Figure 1**, the SP allows their clients to request elastic reservations with the same GoS$_i$ in an established bandwidth range for each reservation of class i. In this paper, some previous considerations should be done from the scenario described in **Figure 1**: First, the parameters GoS and $\overline{B}_{res,i}$ are related to the reservations of the class of service i. The reservations are represented by their length and their reserved bandwidth of the generated session. Thus, each established session is characterized by the time since the user asks for the reservation until the session ends up, and does not take into account the features of the packets transmitted during the session; Second, although many definitions have been used to evaluate the GoS$_i$ for a class i of service, the evaluation of this parameter here is based on the probability of ob-

taining an accepted reservation within the requested range; And third, all the reservations of class i, have the same priority. **Figure 2** shows the entities involved in the scenario described in **Figure 1**. Thus the SPs, which may also act as Content Providers (CPs), offer for each class of service, class i, a guaranteed GoS$_i$. In that sense the SPs should establish the appropriated agreements with the Network Providers (NPs) to buy the necessary access bandwidth that allow them to have the appropriated access bandwidth (B_i) in order to guarantee the offered GoS$_i$.

Therefore, in this paper, it is proposed and analyzed a method that evaluates the price of a class i of guaranteed elastic reservations related to some of the described parameters such as: its elasticity, its guaranteed Grade of Service, its mean reserved bandwidth and the available access bandwidth for the reservations. Qualitatively speaking it works as follows: First, the SP establish the characteristics of the service that wants to offer; second, the SP obtains the average rate of the accepted elastic reservations of this class, class i, with a guaranteed GoS$_i$; third, the SP calculates the price of these reservations that guarantee the GoS$_i$ with a demand function that also depends on a demand modulation factor, $\overline{B}_{res,i}$. This last parameter could be justified by the desire of users of paying more for the reservation when the value of $\overline{B}_{res,i}$, is closer to H_i. And finally, the SP obtains the value of the elasticity of the reservations that gives the maximum revenue and the optimal bandwidth that maximizes the revenue for this elasticity. The pricing method for a class of guaranteed elastic service also could be extended to the evaluation of multiple classes of independent and guaranteed elastic services, applying the obtained expressions in this paper to each considered class. However
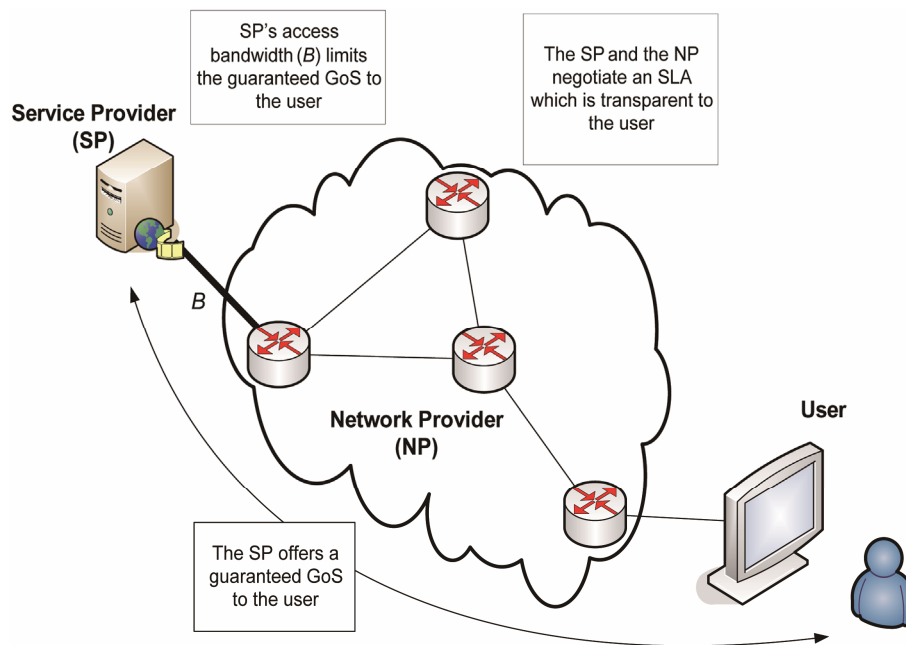
Figure 2. Entities considered.

this method is unable to evaluate and to analyze the case of dependent services, since to deduce the appropriate demand functions or to establish the relations between the variables that are involved in is very difficult task with the analytical tools used here.

It is straight to deduce qualitatively speaking some conclusions, such as, the value of the established GoS for the reservations determines the accepted demand of the requested services. So, qualitatively speaking, if the guaranteed GoS is high the accepted requests will be less than if the guaranteed GoS is low and consequentially, the price of the service would increase for the considered access bandwidth. However, the SPs not only want to get qualitative results, but they also need to establish procedures to know how to quantify the price of these services and to create the appropriated scenario to offer them.

Many questions could appear about the utility of this method for the SPs. Thus, the first one could be for them to try to identify the convenience of its implementation, that is, when this method could be appropriated to implement for elastic services? Other, without any specific order, maybe when the SPs don't have enough resources (limited access bandwidth) and also, their users could accept some changes in the type of service they have requested. In this case, does the obtained revenue allow them to get what they want? Or even, will be the obtained revenue for the elastic reservations not far away or even better than for inelastic ones? Or, the same question could be formulated in other words, what should be the size of the resources (access bandwidth) that prioritizes the use of the elastic against inelastic reservations? How many users could access to the service using elastic res-

ervations in comparison to the case of using inelastic reservations? What is the value of the elasticity of the reservations that gets the maximum revenue? etc. The method must allow answering these questions to the SPs with the aim of getting the solutions that best fit their requirements to price elastic reservations.

In essence, this paper presents four main contributions:

- An evaluation of the proposed pricing method that allows to assign price to multiple classes of independent and guaranteed elastic streaming services with the same priority and, according to some parameters such as the elasticity of their reservations, the available resources (access bandwidth), as well as, other parameters that define and establish the offered services and their demand functions. Each one of the independent classes could be evaluated following the same procedure as for a class i.

- The analysis highlights the importance of the elasticity when the access bandwidth is limited, as well as the importance of its appropriated dimensioning, in order to maximize the revenue of the SP.

- The analytical evaluation of the price assignation for a single class of elastic reservations, class i, is based on a closed-form expression that reduces the computational complexity of the Markov models.

- The demand function for each service will depend not only on its price (€/reservation), p_i, as usually is considered, but also on a demand modulation factor, $\bar{B}_{res,i}$ that also depends on the access bandwidth. Although the selected demand function in the paper is a linear-based function that depends on the price and the access bandwidth, it could be considered another

one. In any case, the price will be finally obtained inverting this demand function.

The remainder of the paper is organized as follows. Section 2 presents some research in this field and it is labeled as related work. Section 3 describes the proposed pricing method for services based on elastic reservations. Section 4 analyzes and applies the proposed method to evaluate the price assignation for a single class of elastic reservations, class i, when a linear-based demand function D_i and a revenue function are selected, quantifying and highlighting the importance of the elasticity of the reservations and the available resources (access bandwidth) in the evaluation of the SPs revenue. Section 5 summarizes the main conclusions. Also this paper includes two appendices. In Appendix A, the GoS_i and the mean reserved bandwidth per accepted request $\bar{B}_{res,i}$ are deduced by means a Markov-chain based model (based on quadratic computational complexity) and an approximate model (based on constant complexity). As the approximate model is very close to the Markov-chain based, the analysis of the method is done based on that because it allows understanding more clearly the relations among the parameters that use this method in the process of assigning prices in Section 4. In Appendix B, the price for a class i of elastic reservations is analytically deduced, taking into account the selected linear-based demand function D_i.

2. Related Work

This paper describes and analyzes a method that helps SPs to price elastic services with guaranteed GoS, selecting an aggregate demand function, D, that establishes the relation between the number of users that are willing to get the service and the price they pay for it. The price of each class of these services is based on: the average rate of the accepted class of elastic reservations with guaranteed GoS and their mean reserved bandwidth per accepted request, $\bar{B}_{res,i}$. In [5] the parameters $B_{res,i}$ and $\bar{B}_{res,i}$ were introduced and was also analyzed how is the influence of the value of the GoS_i in their evaluation In that sense, the parameter $\bar{B}_{res,i}$ is considered as a demand modulation factor for the price of the elastic reservations. This paper introduces the parameter $\bar{B}_{res,i}$ as a new component in the determination of the price of the elastic reservations and also carries out the analysis and calculation of the elasticity of the reservations that maximizes a chosen revenue function. Although many methods to price services have been proposed only a few are focused on elastic reservations but neither of them has jointly tackled the issues treated in this paper. For example, no papers assume the GoS as a constraint that affects the price of the services based on elastic reservations. Thus, in this section some papers that share part of the issues related in this paper have been revised.

Reference [6] analyses the quantitative influence of the guaranteed GoS_i in the evaluation of the mean reserved bandwidth for each reservation $\bar{B}_{res,i}$. Additionally the paper proposes a method to establish the prices of two classes of elastic services, but differs from the paper presented here since the calculated price there didn't have into account the influence of the user's demand, that is, the price of the service always depend on the considered aggregate demand function. The proposal presented in this paper is totally different from the presented there, since here there is selected an estimated (linear-based) demand function for the service that establishes a relation with the price of the service and the $\bar{B}_{res,i}$. Further, is also analyzed how the elasticity and the bandwidth affect the SP's revenue. In [7] the same authors of this paper described a method to price substitute guaranteed services. There, it was selected an exponential aggregate demand function. The prices were found inverting their demand functions and knowing that in equilibrium it is accomplished that the value of the average rate of accepted reservations for each class of service, that maximizes the chosen revenue function, is equal to its aggregate demand function D_i. Besides, the considered method, N classes of substitute services, was only graphically analyzed for the case of two substitute services, the access bandwidth and the elasticity of the reservations were fixed, the determination of the demand functions of the two substitute services, each one depending on the price of the other service, and the attainment of the pairs of the accepted demand (which were obtained by trial and error until they match an expression) that accomplished for both services the guaranteed GoS and maximized the revenue. Although in this paper the used method to deduce the price of the elastic services seems to be the proposed in [7], there are many substantial differences in its application. Thus, in [7] wasn't presented any kind of analysis of the obtained results, due to the difficulty of getting them from the use of a Markov-chain based model. However, in this paper, the approximate analytical solution (closed form solution) allows to establish in a clear way the relationships between the parameters that intervene in the method in the process of pricing elastic services. This paper differs from the presented in [7] at least in three aspects: First, it only treats with one class of service, although the considered dependencies are more complex than there. In fact, the graphical results depend among other parameters on the access bandwidth and the elasticity of the reservations, what allows getting the elasticity and the access bandwidth that optimize the selected revenue function. Second, the demand function of the services depends not only on the price, but on a demand modulation factor, $\bar{B}_{res,i}$. Third, the utilization of the method is based on an approximate analytical model (very close to the simulation model) used in the evalua-

tion of the GoS and $\overline{B}_{res,i}$, and different from the model used in [7].

Some papers present different methods to price elastic services but the proposed solutions that are described there are clearly separated to what is presented in this paper. Thus in [8], the authors present a combined study of price competition and traffic control in a congested network where the SPs set the prices in the aim to maximize their profits. In [9] the authors present a State Estimation based Internet traffic flow control system where the objective is maximize the aggregate bandwidth utility of network sources over their transmission rates. In [10], the paper focuses on the provider competition aspect, in a game theoretic setting, the traffic considered is elastic and there are multiple types of it and each type of traffic is sensitive to a different degree to Quality of Service (QoS). In [11] the authors design a framework that is composed of feedback signals and the corresponding source adaptation scheme to provide differentiated bandwidth service for elastic and inelastic applications. In [12], authors propose an appropriate prioritization pricing structure where users are provided with incentives and are able to choose between two service classes. In [13], authors present an integrated solution (integrating pricing into QoS routing) for enabling the next generation Internet to achieve the differentiated service and availability guarantee. Reference [14] describes in a wireless scenario an admission control algorithm that optimizes the revenue when the QoS is guaranteed. The price depends on the holding prices (bandwidth reserve), the usage price (average usage, the elasticity of the traffic) and the congestion price. References [15-17] present the issue of pricing related to different scenarios and further features for the offered services. Thus in [15] the authors briefly review the state of the art and technological growth of congestion control for integrated service networks since pricing is a proper tool to manage congestion, encourage network growth, and allocate resource to users in a fair manner. Reference [16] is one of the first books that treat conjunctly technology and pricing and, in reference [17] the authors present a recent classification of the proposed pricing methods in wireless networks. In references [18-20] different methods are presented to determine the user utilization function. Thus in [18], the authors propose a solution for bridging the gap between the existing theoretical work on optimal pricing and the unavailability of precise user utility information in real networks. In [19] users specify the utility or value they attach to different quantities of resource using a utility function, so the resource allocator knows the utility function of users at the time of resource allocation and then allocates resources based on the objective of maximizing the aggregate average utility obtained by unit time. In [20] each user is assumed to have a utility function which is a concave increasing function of the rate at which she sends data through the network. The problem is to find the vector of users' rates such that the sum of all users' utility functions is maximized, subject to resource capacity constraints. Other references evaluate how admission control affects the obtained GoS (considered as a technical constraint) of the services. These papers analyze not only the case of a class of service, but for multiple service classes. Thus in [21] authors pay their attention to the interrelation between pricing and admission control in QoS-enabled networks and propose a tariff-based architecture framework that flexibly integrates pricing and admission control for multi-domain Diffserv networks. In [22] a comprehensive survey about Call admission control in wireless networks is shown. In [23] authors say that traditional CAC schemes mainly focus on the trade-offs between new call blocking probability and handoff call blocking probability. Therefore, they introduce the pricing as an additional dimension of call admission control process in order to efficiently and effectively control the use of wireless network resources. In [24] authors investigate the conditions where both BE traffic and traffic explicitly requiring QoS (Guaranteed Performance, GP) are present and they propose three CAC rules for the GP traffic. In [25] authors utilize admission control algorithms designed for revenue optimization with QoS guarantees to derive optimal pricing of multiple service classes in wireless cellular networks. Other authors analyze price assignation and propose solutions that work in different behavior. Thus in [26], it is described a scalable connection management strategy for QoS-enabled networks to tackle the problem of appropriately provisioning and allocating connections. In [27] is introduced a service model that provides per-flow bandwidth guarantees, where users subscribe for a guaranteed rate. In [28] authors consider the problem of pricing for bandwidth provisioning over a single link. The network administrator controls the resource allocation by setting a price at every epoch, and each user's response to the price governed by a demand function. In [29] authors investigate the sensitivity of resource allocation and the resulting QoS to resource prices in a reservation-based QoS architecture that provides guaranteed bounds on packet loss and end-to-end delay for real applications. In [30] is considered the pricing and allocation issues of distributing digital contents via Web and P2P channels. Utilizing a game theoretic model, the allocation equilibrium with respect to various business goals is examined. In [31] is established a method to assign prices based on-packet queues sizes in the networks.

3. A Pricing Method for Services Based on Elastic Reservations

This section describes the pricing method for multiple classes of independent and guaranteed elastic reservations. However, before going on with the method, it is

convenient to take into account the difficulty of determining the aggregate demand function D_i, which is a similar problem in many proposals of pricing services. The knowledge of this function in advance is always, as many researchers have pointed out, a very difficult task that the SP needs to solve. As it is known, the aggregate demand function usually represents the sum of individual demands of each user that have different willingness to pay for the service. It is hard to identify this behavior and therefore, the curve that shows the desire of the users to pay a price for the requested services. Thus, the SP has to estimate by whatever means it deems adequate (analytically, by simulation, heuristically, etc.) the demand function for each service. For simplicity, this paper assumes that in the evaluation of a service class of elastic reservations, class i in Section 4, the chosen aggregate demand function is linear-based. Of course, if the demand function changes, the quantitative results that the method obtains should be different.

The outcomes of this method are the price p_i and the value of the elasticity of each service ξ_i and the optimal bandwidth that maximize the SP's revenue. The method consists of the following steps:

The SP determines the service requirements that limit its feasibility. **Figures 1** and **2** illustrate the first requirement: the SP wants to offer guaranteed and independent services based on elastic reservations. This implies that each service has to guarantee a particular GoS_i, for the service of each elastic reservation with elasticity ξ_i that is determined by the SP in order to optimize its revenue. Also, the service requires a *Desired QoS* equal to H_i Mb/s and therefore, from (1), the *Minimum QoS* will be equal to $H_i \cdot (1 - \xi_i)$. Other requirement could appear from the available resources of the SP, that is, the access bandwidth for each independent service (B_i Mb/s). As each service allocates B_i, the sum of the reservations of all classes should be below this access bandwidth B_i. B_i may be limited due to several circumstances, such as the network access technology used by the SP to offer the service.

The SP evaluates the maximum demand (in terms of requests per unit time) that can be allowed for the service in order to guarantee the requirements of step 1, the value of the GoS_i for every offered class of independent elastic reservations. In this paper, we present in Appendix A.2 an analytical expression that roughly approximates the GoS_i for the elastic services that are offered using an access bandwidth B_i. This expression has been validated by simulation and using the loss system model also included in Appendix A.2.

The SP obtains the price of the service i that guarantees the GoS_i. In order to obtain the price for the service, the SP needs to estimate the demand function D_i by whatever means it deems adequate. This paper assumes a

linear-based demand function, explained in Appendix B, that depends on the price and the mean reserved bandwidth $\overline{B}_{res,i}$, (a demand modulation factor that is calculated in Appendix A.3). Since $\overline{B}_{res,i}$ is also dependent on B_i, the demand function is also defined in terms of the price of the service and B_i.

The SP evaluates the revenue R_i using the obtained price in step 3 and finds out the elasticity of the reservations that maximizes the revenue. In addition, if bandwidth was not limited, the SP could obtain the bandwidth that optimally determines the service access bandwidth B_i. The revenue function, used in this paper, assumes for simplicity that only depends on the price, the rate of accepted requests and the cost of the access bandwidth. However, it is well known that more complex expressions, which may express part of the SP's business model, could also be used.

4. Analysis of the Method: The Importance of the Elasticity of the Reservations and the Bandwidth on the SP's Revenue

In this section we apply the pricing method, described in Section 3, to analyze quantitatively how the price of a class i of elastic reservations and the revenue change depending on the elasticity of the reservations and the access bandwidth of the service.

4.1. Step 1: Determining the Requirements for the Service

Before the SP applies this method, it should define the suitable parameters for each class of elastic service. Thus, in this paper it is assumed a guaranteed $GoS_i = g_i$, for all reservations, the same elasticity, and the highest bandwidth of the requested reservation range is H_i that is equal to the maximum required bandwidth to deliver the content ($B_{top,i}$). Specifically, it is supposed a guaranteed GoS_i of 0.95, $H_i = 1$ Mb/s, and 3 minutes for the mean reservation holding time. Also, other parameters for the linear-based demand function as it is presented in Appendix B are: $D_{max,i} = 120$ reservations/minute, $B_{top,i} = H_i = 1$ Mb/s and $P_{max,i} = 10€$.

Step 2: Evaluating the maximum demand that guarantees the GoS.

The SP calculates the average rate of the accepted reservations, λ_i, that guarantees the GoS_i. Using the approximation (12) in Appendix A.2, the expression for the maximum accepted demand (2) is obtained.

$$\lambda_i \big|_{(GoS_i = g_i)} = B_i \cdot \left(\frac{1}{\mu_i} H_i (1 - \xi_i) g_i \right)^{-1} \tag{2}$$

Figure 3 shows graphically how the maximum accepted demand that guarantees a $GoS_i = 0.95$ changes for different values of elasticity and bandwidth. If the elas-

ticity tends to 1, the demand tends to $+\infty$, since the reservation requests always are accepted. On the other hand, if the elasticity is zero, the demand tends to a minimum value if bandwidth is fixed, since a reservation with elasticity 0 requires no less than H_i Mb/s. For the rest of combinations of bandwidth and elasticity, λ_i increases slightly for low-medium elasticity values and for high elasticity values λ_i increases sharply approaching to a vertical asymptote to $+\infty$ for elasticity equal to 1. As B_i increases, the maximum accepted demand also increases in a linear way with a higher slope as elasticity approaches to 1. Also it is worth to mention that the values of the guaranteed GoS$_i$, H_i and the holding time $\frac{1}{\mu_i}$ have an impact on the maximum accepted demand. Thus, the lower they are, the higher can be the maximum accepted demand.

4.2. Step 3: Calculating the Price

Appendix B describes the demand function for the analyzed service that depends on the users' willingness to pay and a demand modulation factor (*i.e.*, the mean reserved bandwidth per accepted request that is described in Appendix A). In this step, the SP obtains the price of the accepted elastic reservations that guarantee a GoS$_i$ = g_i. This price can be obtained using expressions (2) and (19) when $B_{top,I} = H_i$.

Figure 4 shows graphically how the price p_i changes for different values of elasticity and bandwidth. The price decreases to 0€ when the elasticity ξ_i approaches to

$$1 - \sqrt{B_i \mu_i / \left(D_{max,i} H_i g\right)} = 1 - \sqrt{B_i / 342 \text{ Mb/s}} ,$$

or similarly, when B_i approaches to

$$\frac{1}{\mu_i} D_{max,i} H_i \left(1 - \xi_i\right)^2 g_i = 342 \left(1 - \xi_i\right)^2 \text{ Mb/s} .$$

A price of 0€ means that the GoS$_i$ is guaranteed even

with the maximum accepted demand. Regarding the guaranteed GoS$_i$ and according to (3) the price tends to be 0 as lower is g_i. This is because the SP can guarantee a GoS$_i$ that tends to 0, even if the maximum accepted demand is considered. The *Desired QoS,H$_i$*, and the mean holding time of the reservations, $\frac{1}{\mu_i}$, make the price to increase if bandwidth is limited, since the required resources (bandwidth) also increases. Finally, the increasing of the maximum accepted demand $D_{max,i}$, implies that the price augments in the aim to guarantee the GoS$_i$ and therefore, the average rate of accepted reservations decreases. On the other hand, an augment of the maximum price $P_{max,i}$ implies that the price increases since the willingness to pay of users also increases.

4.3. Step 4: Establishing the Elasticity of Reservations and the Bandwidth in Order to Maximize the Revenue

The SP calculates the revenue from the deduced price of the service that guarantees the GoS for this class of service, class i. In this paper, an intuitive revenue function R_i(4) is considered that have three terms: the accepted service's demand $D_i \cdot$GoS$_i$, the price paid for the service p_i and the cost of the allocated resources, which is supposed proportional to the access bandwidth. Alternatively, other more complex revenue functions [32] could be applied in order to include other special characteristics of each SP.

$$\begin{aligned} R_i &= D_i \cdot \text{GoS}_i \cdot p_i - \text{Cost}\left(B_i\right) \\ &= \lambda_i \cdot \text{GoS}_i \cdot p_i - \alpha_i B_i \ \left(\text{€/minute}\right) \end{aligned} \tag{4}$$

The SP evaluates the revenue substituting in expression (4) the value of the price that guarantees the GoS$_i$ (3), the value of λ_i for $B_{top,i} = H_i$ (32), and the value of the GoS$_i$ (12), as it is shown in expression (5).

$$p\big|_{\text{GoS}_i = g_i} = \begin{cases} P_{max,i}\left(1 - \dfrac{B_i}{\dfrac{1}{\mu_i} H \left(1 - \xi_i\right)^2 D_{max,i}} \dfrac{1}{g_i}\right) & B_i \leq \dfrac{1}{\mu_i} D_{max,i} H_i \left(1 - \xi_i\right)^2 g_i \\[4ex] 0 & B_i \geq \dfrac{1}{\mu_i} D_{max,i} H_i \left(1 - \xi_i\right)^2 g_i \end{cases} \tag{3}$$

$$R_i = \begin{cases} \dfrac{B_i}{\dfrac{1}{\mu_i} H_i \left(1 - \xi_i\right)} P_{max,i}\left(1 - \dfrac{B_i}{\dfrac{1}{\mu_i} H \left(1 - \xi_i\right)^2 D_{max,i}} \dfrac{1}{g}\right) - \alpha_i B_i & B_i \leq \dfrac{1}{\mu_i} D_{max,i} H_i \left(1 - \xi_i\right)^2 g \\[4ex] -\alpha_i B_i & B_i \geq \dfrac{1}{\mu_i} D_{max,i} H_i \left(1 - \xi_i\right)^2 g \end{cases} \tag{5}$$
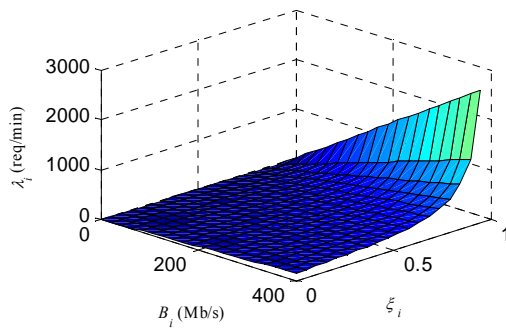
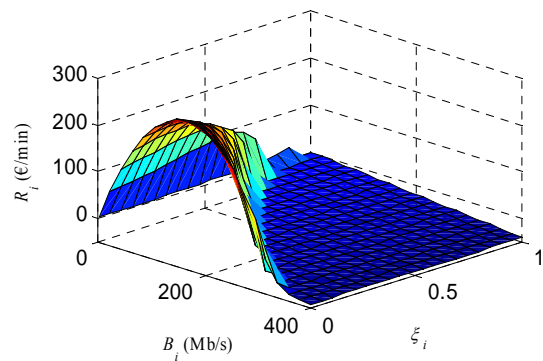**Figure 3. Maximum demand for a guaranteed GoS$_i$ = 0.95 and H_i = 1 Mb/s.**

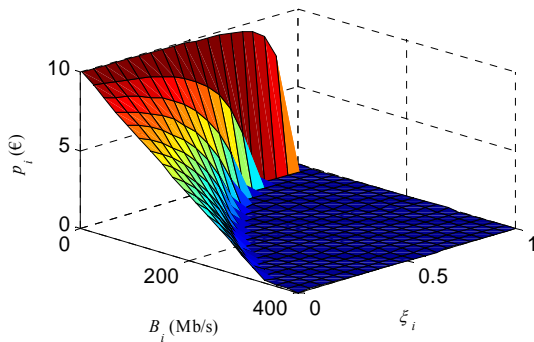**Figure 4. Price for a Guaranteed GoS$_i$ = 0.95 and H_i = 1 Mb/s.**

In the case that the SP has a limited bandwidth the elasticity of the reservations that maximizes the revenue, ξ_i^* is:

$$R_i = \begin{cases} 1 - \sqrt{\dfrac{3B_i\mu_i}{H_i D_{\max,i} g}} & B_i \leq \dfrac{1}{3\mu_i} D_{\max,i} H_i \cdot g \\[2ex] 0 & B_i \geq \dfrac{1}{3\mu_i} D_{\max,i} H_i \cdot g \end{cases} \quad (6)$$

Figure 5 shows graphically how the revenue R_i(5) changes for different values of elasticity and bandwidth and **Figure 6** shows how the elasticity that maximizes the revenue for a Guaranteed GoS$_i$ depends on bandwidth. Analyzing expressions (5) and (6), it can be deduced that the revenue increases until the value of elasticity given by $\xi_i = 1 - \sqrt{B_i/114 \text{ Mb/s}}$, if $B_i \leq 114$ Mb/s. If bandwidth is higher than this value, the elasticity of the reservations will not provide any significant benefit in comparison with an inelastic reservation. The optimum value of the elasticity is different for each considered B_i and decreases as B_i increases. This behavior is due to the fact that an increment of B_i implies more resources that allow accepting more users with less elasticity in their reservations. On the other hand, if the SP is forced to use a pre-established price, the use of a lower g_i can increase its revenue. The increase of H_i, and $1/\mu_i$ implies that the optimum elasticity augments. Finally, an increase of the maximum accepted

Figure 5. Revenue for a Guaranteed GoS$_i$ = 0.95 and H_i = 1 Mb/s.

Figure 6. Values of the elasticity of the reservations that maximize the revenue for a Guaranteed GoS$_i$ = 0.95 and H_i = 1 Mb/s.

demand $D_{\max,i}$, makes the revenue and the elasticity that maximize the revenue to increase and an increment of the maximum price $P_{\max,i}$, forces the revenue to increase, since the willingness to pay of the users also increases, but this effect has no impact on the optimum elasticity.

The revenue for the optimum elasticity of the reservations is the following:

$$R_i\Big|_{\xi_i=\xi_i^*} = \frac{2}{9} P_{\max,i} \sqrt{\frac{3B_i\mu_i}{H_i} D_{\max,i} g} - \alpha_i B_i$$

when $B_i \leq \dfrac{1}{3\mu_i} D_{\max,i} H_i \cdot g$

$$R_i\Big|_{\xi_i=\xi_i^*} = \frac{B_i}{\frac{1}{\mu_i} H_i} P_{\max,i} \left(1 - \frac{B_i}{\frac{1}{\mu_i} H \cdot D_{\max,i}} \frac{1}{g}\right) - \alpha_i B_i \quad (7)$$

when $\dfrac{1}{3\mu_i} D_{\max,i} H_i \cdot g \leq B_i \leq \dfrac{1}{\mu_i} D_{\max,i} H_i g$

$$R_i\Big|_{\xi_i=\xi_i^*} = -\alpha_i B_i$$

when $B_i \geq \dfrac{1}{3\mu_i} D_{\max,i} H_i \cdot g$.

In the case that the SP has not limited bandwidth the value of the bandwidth that gives the best revenue, when

is applied the elasticity of the reservations that maximizes the revenue. It further allows an optimal dimensioning of the access bandwidth B_i, is B_i^* and is expressed according to (8).

$$B_i^*\big|_{\xi_i=\xi_i^*} = \begin{cases} 0 & \dfrac{\mu_i P_{\max,i}}{H_i} \le \alpha_i \\[2mm] \dfrac{1}{27}\dfrac{\mu_i P_{\max,i}^2 D_{\max,i} g}{\alpha_i^2 H_i} & \alpha_i \le \dfrac{\mu_i P_{\max,i}}{H_i} \le 3\alpha_i \\[2mm] \dfrac{1}{2\mu_i} D_{\max,i} g\left(1 - \dfrac{\alpha_i H_i}{\mu_i P_{\max,i}}\right) & \dfrac{\mu_i P_{\max,i}}{H_i} \ge 3\alpha_i \end{cases}$$

(8)

Figure 7 shows how the revenue $R_i\big|_{\xi_i=\xi_i^*}$ (7) changes for different values of bandwidth. As it can be seen, the revenue has a maximum equal to 268.16€/min for $B_i^* = 165.87$ Mb/s that corresponds to $\xi_i^* = 0$. The value of the optimum bandwidth B_i^* decreases with the increment of the price of the access bandwidth (α_i).

5. Conclusions

This paper proposes a pricing method that helps SPs to assign suitable prices to multiple classes of independent streaming elastic services with guaranteed GoS. The paper determines the price of one class i of elastic streaming services taking into account the accepted average rate of reservations λ_i, with a guaranteed GoS_i and, assuming a linear-based function as aggregate demand function D_i.

The SPs that want to offer elastic services should have to calculate their prices. In this process this method could help them in calculating them by means of defining or estimating in advance for these services some of the parameters that best could match their needs for the available resources. Some of them are: the value of the offered elasticity of the reservations ξ_i (it could be what offers the maxim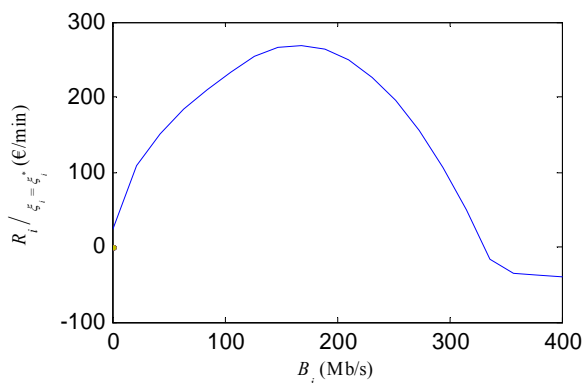um revenue), the highest value of the reservation H_i, this parameter is related to the maximum quality of the delivered content that the SP expects to give their users, the value of the guaranteed GoS_i (this value could be set according to the value offered by other SPs or totally different) and the available resources (*i.e.*, the access bandwidth, B_i).

On the other hand an issue that could be difficult to determine is the aggregate demand function. However, it is known that the SPs have the appropriate tools to approximately estimate this function and to overcome this situation. The accuracy of this estimation is crucial to evaluate the price using this method.

Currently we are developing simulation tools for estimating and establishing the profile of the users that could access to this type of services what would make possible to deduce appropriated aggregate demand functions in scenarios where the SPs could offer these elastic services. Also, we are also working on extending this method to the case of substitute services and to include different priorities to the reservations based on their elasticity. Finally it is our challenge to analyze the case of multiple classes of elastic services that are not independent, in this sense simulation tools are under investigation since analytic models to deduce the derived aggregated demand functions in this case and the relations between the parameters that intervene are really hard to find out. Future work should also include a deep revision and proposal of new revenue functions and consequently in the their evaluation, the value of the elasticity that maximizes their revenues and what should be the relation in each case among the elasticity, the access bandwidth and other parameters involved in the aim to get the maximum revenue.

6. Acknowledgements

Figure 7. Optimum bandwidth for the elasticity of the reservations that maximizes the revenue for a Guaranteed GoS$_i$ = 0.95 and H$_i$ = 1 Mb/s.

REFERENCES

[1] B. Braden, *et al*., "Resource ReSerVation Protocol (RSVP), Version 1, Functional Specification," RFC 2205, 1997. http://www.rfc-editor.org/rfc/pdfrfc/rfc2205.txt.pdf

[2] R. Hancock, G. Karagiannis, J. Loughney and S. Van den Bosch, "Next Steps in Signaling (NSIS): Framework," RFC 4080, 2005. http://www.rfc-editor.org/rfc/pdfrfc/rfc4080.txt.pdf

[3] J. Manner, G. Karagiannis and A. MacDonald, "NSIS Signaling Layer Protocol (NLSP) for Quality-of-Service Signaling," RFC 5974, 2010. http://www.rfc-editor.org/rfc/pdfrfc/rfc5974.txt.pdf

[4] J. Ash, A. Bader and C. Kappler, "QoS-NSLP QSPEC

Template", RFC 5975, 2010.
http://www.rfc-editor.org/rfc/pdfrfc/rfc5975.txt.pdf

[5] M. Postigo-Boix and J. L. Melús-Moreno, "Influence of the Grade of Service in the Evaluation of the Mean Reserved Bandwidth for Elastic Services," *IEEE Communication Letters*, Vol. 12, No. 2, 2008, pp. 143-145. doi:10.1109/LCOMM.2008.071659.

[6] M. Postigo-Boix and J. L. Melús, "Utilization-Based Prices for Elastic Services", *Computer Communications*, Vol. 31, No. 14, 2008, pp. 3503-3509. doi:10.1016/j.comcom.2008.06.006.

[7] M. Postigo-Boix and J. L. Melús, "A Proposal for Pricing Substitute Guaranteed Services," *IEEE Communications Letters*, Vol. 15, No. 1, 2011, pp. 100-102. doi:10.1109/LCOMM.2010.01.101336.

[8] A. Ozdaglar, "Price Competition with Elastic Traffic," *Journal on Networks*, Vol. 52, No. 3, 2008, pp. 141-155. doi:10.1002/net.20239.

[9] G. Abbas, A. K. Nagar, H. Tawfik and J. Y. Goulermas, "Estimation Based Distributed QoS Pricing and Scheduling for Elastic Internet Services," *2nd International Conference on Developments in eSystems Engineering* (*DESE*), Abu Dhabi, 2009, pp. 89-98. doi:10.1109/DeSE.2009.15.

[10] P. Dube and R. Jain, "Diffserv Pricing Games in Multi-Class Queuing Network Models," *22nd International Teletraffic Congress* (*ITC*), 7-9 September 2010, Amsterdam, pp. 1-8. doi:10.1109/ITC.2010.5608737.

[11] Z. G. Li, C. Chen and Y. C. Soh, "Pricing Based Differentiated Bandwidth Service," *4th IEEE Conference on Industrial Electronics and Applications ICIEA* 2009, Xi'an 25-27 May 2009, pp. 780-787. doi:10.1109/ICIEA.2009.5138310.

[12] J. L. Van Den Berg, M. R. H. Mandjes and R. Nuñez Quejia, "Pricing and Distributed QoS Control for Elastic Network Traffic," *Operations Research Letters*, Vol. 35, No. 3, 2007, pp. 297-307. doi:10.1016/j.orl.2006.03.018.

[13] G. Cheng, N. Ansari and S. Papavassiliou, "Adaptive QoS Provisioning by Pricing Incentive QoS Routing for Next Generation Networks," *Journal of Computer Communications*, Vol. 31, No. 10, 2008, pp. 2308-2318. doi:10.1016/j.comcom.2008.02.019.

[14] X. Wang and H. Schulzrine, "Pricing Network Resources for Adaptive Applications," *IEEE/ACM Transactions on Networking*, Vol. 14, No. 3, 2006, pp. 506-519. doi:10.1109/TNET.2006.872574.

[15] X. Chang and D. W. Petr, "A Survey of Pricing for Integrated Service Networks," *Journal on Computer Communications*, Vol. 24, No. 18, 2001, pp. 1808-1818. doi:10.1016/S0140-3664(01)00327-9.

[16] C. Courcoubetis and R. Weber, "Pricing Communication Networks Economics, Technology and Modelling," Wiley, New York, 2003. doi:10.1002/0470867175.

[17] C. A. Gizelis and D. D. Vergados, "A Survey of Pricing Schemes in Wireless Networks," *IEEE Communications Surveys & Tutorials*, Vol. 13, No. 1, 2011, pp. 126-145. doi:10.1109/SURV.2011.060710.00028.

[18] S. Park, S. Han and S. W. Cho, "User Utility Discovery

[19] S. Kalyanasundaram, E. K. P. Chong and N. B. Shroff, "Optimal Resource Allocation in Multi-Class Networks with User-Specified Utility Functions," *Computer Networks*, Vol. 38, No. 5, 2002, pp. 613-630. doi:10.1016/S1389-1286(01)00275-4.

[20] J. Kuri and S. Roy, "Pricing Network Resources: A New Perspective," *International Conference on Wireless Communications*, *Networking and Mobile Computing*, Shanghai, 21-23 September 2007, pp. 1937-1940. doi:10.1109/WICOM.2007.485.

[21] T. Li, D. Y. Iraqi and R. Boutaba "Pricing and Admission Control for QoS-Enabled Internet," *Computer Networks*, Vol. 46, No. 1, 2004, pp. 87-110. doi:10.1016/j.comnet.2004.03.020.

[22] M. H. Ahmed, "Call Admission Control in Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys & Tutorials*, Vol. 7, No. 1, 2005, pp. 50-69. doi:10.1109/COMST.2005.1423334.

[23] J. Hou, J. Yang and S. Papavassiliou, "Integration of Pricing and Call Admission Control to Meet QoS Requirements in Cellular Networks," *IEEE Transactions on Parallel and Distributed Systems*, Vol. 13, No. 9, 2002, pp. 898-910. doi:10.1109/TPDS.2002.1036064.

[24] M. Baglietto, R. Bolla, F. Davoli, M. Marchese and M. Mongelli, "A Proposal of New Price-Based Call Admission Control Rules for Guaranteed Performance Services Multiplexed with Best Effort Traffic," *Computer Communications*, Vol. 26, No. 13, 2003, pp. 1470-1483. doi:10.1016/S0140-3664(03)00032-X.

[25] O. Yilmaz and Ing-Ray Chen, "Utilizing Call Admission Control for Pricing Optimization of Multiple Service Classes in Wireless Cellular Networks," *Computer Communications*, Vol. 32, No. 2, 2009, pp. 317-323. doi:10.1016/j.comcom.2008.11.001.

[26] E. W. Fulp and D. S. Reeves, "Bandwidth Provisioning and Pricing for Networks with Multiple Classes of Service," *Computer Networks*, Vol. 46, No. 1, 2004, pp. 41-52. doi:10.1016/j.comnet.2004.03.018.

[27] J. Elias, F. Martignon, A. Capone and G. Pujolle, "A New Approach to Dynamic Bandwidth Allocation in Quality of Service Networks: Performance and Bounds," *Computer Networks*, Vol. 51, No. 10, 2007, pp. 2833-2853. doi:10.1016/j.comnet.2006.12.003.

[28] U. Savagaonkar, E. K. P. Chong and R. L. Givan, "Online Pricing for Bandwidth Provisioning in Multi-Class Networks," *Computer Networks*, Vol. 44, No. 6, 2004, pp. 835-853. doi:10.1016/j.comnet.2003.12.011.

[29] N. Jin and S. Jordan, "The Effect of Bandwidth and Buffer Pricing on Resource Allocation and QoS," *Computer Networks*, Vol. 46, No. 1, 2004, pp. 53-71. doi:10.1016/j.comnet.2004.03.023.

[30] Y.-M. Li, "Pricing Digital Content Distribution over Heterogeneous Channels," *Decision Support Systems*, Vol. 50, No. 1, 2010, pp. 243-257. doi:10.1016/j.dss.2010.08.027.

[31] R. Hassin and M. Haviv, "To Queue or Not to Queue,"

Springer, Boston, 2003.

[32] K. Talluri and G. Van Ryzin, "The Theory and Practice of Revenue Management," Kluwer Academic Publishers, Dordrecht, 2005.

[33] L. Kleinrock, "Queueing Systems, Vol. I: Theory," John Wiley and Sons, New York, 1975.

[34] M. Abramowitz and I. A. Stegun, "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables," Dover, New York, 1972.

Appendix A. Determining the GoS_i and the mean Reserved Bandwidth per Accepted Request, $\bar{B}_{res,i}$

In this Appendix, we present a Markov-chain based model to evaluate the GoS_i and the mean reserved bandwidth per accepted request, $\bar{B}_{res,i}$. In addition, it is also presented an approximate model that notably simplifies the computational complexity of this evaluation and it is very close to the simulation model. This analytical approximation has been used to get the quantitative results of Section 4, as well as to obtain the demand function in terms of the price and the bandwidth of Appendix B.

A.1. Establishing Assumptions

The following assumptions are about the allocation of B_i among the requested elastic reservations:

The reservations requests follow a Poisson process with an average rate of λ_i (reservations/minute).

The reservation's holding time—that is, the time needed to deliver the content—is exponentially distributed with mean $1/\mu_i$ (minutes). The mean delivery time for a particular request remains the same regardless of the value of the reserved bandwidth ($B_{res,i}$). This is the case for the delivery of audio or video streaming using two different reserved bandwidths $\left(B_1 < B_2\right)$; the reservation's holding time for B_2 is $d = C_2/B_2$ (minutes), where C_2 (bits) represents the data to be delivered. When the reserved bandwidth is B_1, the delivery time remains the same, since the content size is reduced to $C_1 = (B_1/B_2) \cdot C_2$ (bits), and the quality of the received content is reduced accordingly.

The reserved bandwidth for an accepted reservation request is according to (1) within the interval $\left[L_i, H_i\right] = \left[H_i \cdot (1 - \xi_i), H_i\right]$.

The server rejects an elastic reservation if no access bandwidth is available to hold the required minimum bandwidth L_i for this reservation.

The content server assigns a bandwidth B_i for reservations of class i that must share it in competition. The maximum number of reservations of class i (N_i) that can share the access bandwidth is calculated as follows:

$$N_i = \left\lfloor B_i/L_i \right\rfloor = \left\lfloor B_i \Big/ \left(H_i \cdot (1 - \xi_i)\right) \right\rfloor \qquad (9)$$

where $\lfloor . \rfloor$ represents the standard floor operator. N_i represents the access number of elastic reservations of class i with an access bandwidth B_i for each class i.

According to these assumptions, the model that better fits to perform the evaluation of the GoS_i and $\bar{B}_{res,i}$ of the class i of elastic reservations is the M/M/N/N loss queuing model [33]. **Figure 8** represents the state-transition rate diagram. The associated analysis of this loss queuing model is based on a well-known expression [33]. Besides the simplicity of the **Figure 8**, the difficulty in this type of analysis is to identify the appropriate queuing model that performs the evaluation of the GoS_i and $\bar{B}_{res,i}$ for each class i of elastic reservations. Thus the state-transition diagram of **Figure 8** represents the number of elastic reservations of class i (N_i) that shares the access bandwidth B_i.

Each state of the Markov chain shown in **Figure 8** represents the probability of having $k \in \left[0, N_i\right]$ accepted elastic reservations of class i, where the state (k,i) represents that k bandwidth reservations are sharing the access bandwidth.

The probability of each state is well known (10) and it is represented by the vector $\prod_i = \left(\pi_{(0,i)}, \pi_{(1,i)}, \cdots, \pi_{(N,i)}\right)$ and each component of this vector is represented by the well-known formula for loss queuing models [33].

$$\pi_{(k,i)} = \frac{\left(\lambda_i/\mu_i\right)^k}{k!}\left(\sum_{l=0}^{N_i} \frac{\left(\lambda_i/\mu_i\right)^l}{l!}\right)^{-1} \qquad (10)$$

A.2. Evaluation of the GoS

In this paper, the GoS_i for the elastic reservations is defined as the probability of accepting a reservation of class i. Equation(11) shows the value of the GoS_i where $\Gamma(a,b)$ represents the incomplete Gamma function [34].

$$
\begin{aligned}
GoS_i &= 1 - \pi_{(N_i,i)} = 1 - \frac{\left(\lambda_i/\mu_i\right)^{N_i}}{N_i!}\left(\sum_{l=0}^{N_i} \frac{\left(\lambda_i/\mu_i\right)^l}{l!}\right)^{-1} \\
&= 1 - \frac{\left(\lambda_i/\mu_i\right)^{N_i}}{e^{\left(\lambda_i/\mu_i\right)} \cdot \Gamma\left(N_i + 1, \left(\lambda_i/\mu_i\right)\right)}
\end{aligned} \qquad (11)
$$

Figures 9 and **10** show the type of relations between the GoS_i and the parameters λ_i, and B_i for ξ_i equal to 0
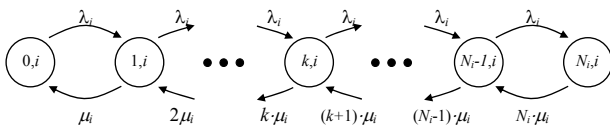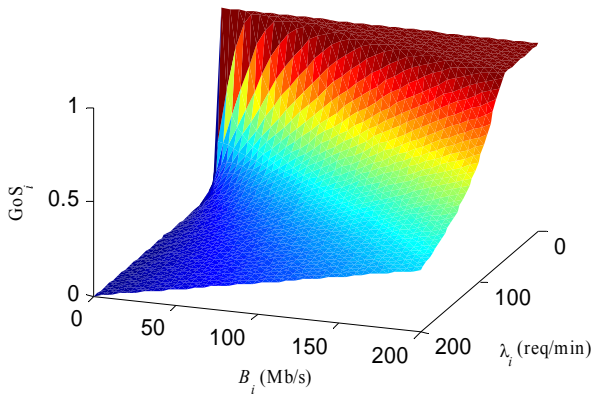
Figure 8. Transition diagram.

Figure 9. GoS$_i$ vs. λ_i and B_i, for H_i = 1 Mb/s and ξ_i = 0.
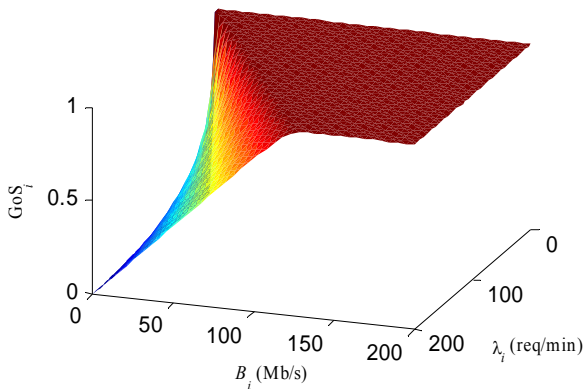

Figure 10. GoS$_i$ vs. λ_i and B_i, for H_i = 1 Mb/s and ξ_i = 0.8.

and 0.8 respectively, where the mean holding time is exponential with a mean of $1/\mu_i$ = 3 minutes and H_i = 1 Mb/s. Although the model has been validated by simulation, here these results are not presented since them exactly match the evaluation results of (11). The observed behavior in **Figures 9** and **10** for the GoS$_i$ allows deducing some interesting general properties for the GoS, which is defined in the interval $\Omega_{\text{GoS}} = [0,1]$:

The GoS decreases with λ, that is,

$$\exists \lambda^* > \lambda \left| \text{GoS}(\lambda, B) > \text{GoS}(\lambda^*, B) \right. .$$

This is because more requests arrive at the system for the same shared resources (bandwidth B_i).

The GoS increases with B, or similarly,

$$\exists B^* > B \left| \text{GoS}(\lambda, B) < \text{GoS}(\lambda, B^*) \right. .$$

This is because the resources increase and the average rate of reservation requests (λ_i) remains constant.

The GoS increases with ξ. Equivalently,

$$\exists \xi^* > \xi \left| \text{GoS}(\lambda, B, \xi) < \text{GoS}(\lambda, B, \xi^*) \right. .$$

The elasticity of the reservations reduces their bandwidth consumption requirements. Thus, if the reservation is more elastic, more requests can be accepted for the same GoS$_i$, and for the same average rate of reservation requests the GoS$_i$ increases.

For specific B and ξ parameters,

$$\exists \lambda \in \left[0, \lambda_g \right] \left| \text{GoS}(\lambda, B, \xi) \geq g \right. .$$

The g parameter is the guaranteed GoS. This is a consequence of the first property and implies that to guarantee a particular GoS, g_i, is necessary to keep the average rate of reservation requests below λ_{g_i}.

In terms of computational complexity, the use of this model presents a quadratic complexity with respect to N, and consequently, the needed number of mathematical operations to obtain the GoS$_i$ increases quadratically with B_i and ξ_i, and decreases quadratically with H_i. Therefore, in order to decrease this complexity, our suggestion is to roughly approximate the value of the GoS$_i$, taking in account that GoS$_i$ is a measure of the probability of using the available resources to get the service. According to this, the value of the GoS$_i$ can be approximated by the ratio between the used and requested bandwidth:

$$\text{GoS}_i \approx \begin{cases} 1 & \text{if } \dfrac{\lambda_i}{\mu_i} H_i (1 - \xi_i) \leq B_i \\[3ex] \dfrac{B_i}{\dfrac{\lambda_i}{\mu_i} H_i (1 - \xi_i)} & \text{if } \dfrac{\lambda_i}{\mu_i} H_i (1 - \xi_i) > B_i \end{cases} \qquad (12)$$

This approximation has constant computational complexity and allows clarifying, by means of a closed-form expression, how the different parameters affect the evaluation of the GoS$_i$. The relative error of this approximation is less than the 3% for values of B_i far away from the condition

$$\frac{\lambda_i}{\mu_i} H_i (1 - \xi_i) = B_i$$

in which the error is significant. Nevertheless, in this work, this approximation offers a more intuitive expression that allows knowing better and quicker its quantitative evaluation and to clarify its qualitative behavior. **Figure 11** shows the relative error (higher than 3%) for the calculation of the GoS$_i$ using the approximation in (12). The relative error is higher when

$$\frac{\lambda_i}{\mu_i} H_i (1 - \xi_i) = B_i ,$$
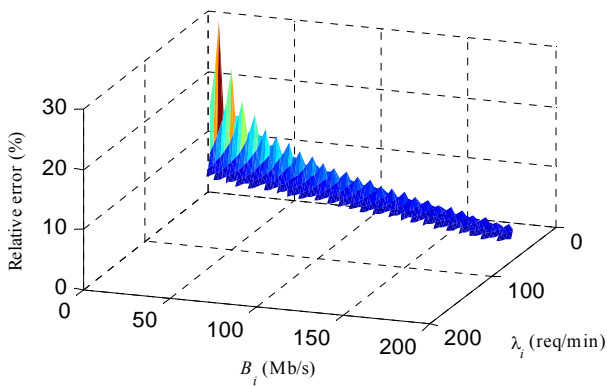
and the value of GoS tends to 0.

Figure 11. Relative error (higher than 3%) for the calculation of the GoS of Figure 9 using the approximation in (12).
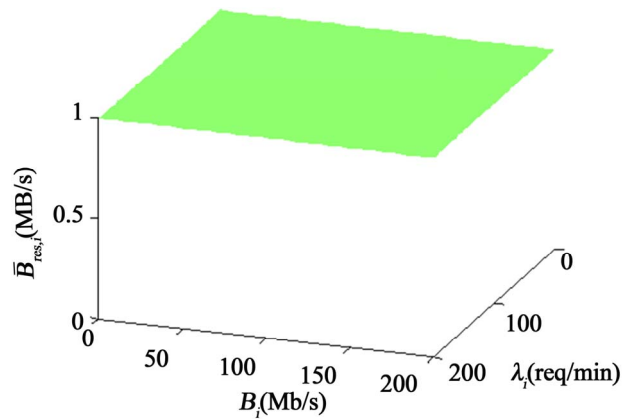
A.3. Evaluation of the Mean Reserved Bandwidth per Accepted Request $\overline{B}_{res,i}$

The mean reserved bandwidth per accepted reservation request for class I [5], is expressed as follow:

$$\overline{B}_{res,i} = \text{mean\_value}\{B_{res,i}\} \quad (\text{Mb/s}) \tag{13}$$

The evaluation of the $\overline{B}_{res,i}$ includes the value of each $B_{res,i}$ for each state $k(B_{res,i})_k$. Thus, expression (14) evaluates $\overline{B}_{res,i}$:

$$\overline{B}_{res,i} = \sum_{k>0}(B_{res,i})_k \cdot \pi_k \tag{14}$$

When k reservations of class i are accepted in the access link and $k \cdot H_i \le B_i$, the value of $(B_{res,i})_k$ is H_i. Otherwise, all k reservations will share the bandwidth B_i, and $(B_{res,i})_k$ is B_i/k.

$$(B_{res,i})_k = \begin{cases} H_i & 0 < k \le B_i/H_i \\ B_i/k & k > B_i/H_i \end{cases} \tag{15}$$

Figures 12 and **13** show the relationship among the parameters $\overline{B}_{res,i}$, λ_i, and B_i, with $\xi_i = 0$ and $\xi_i = 0.8$ respectively, the mean holding time is $1/\mu_i = 3$ minutes (exponential) and $H_i = 1$ Mb/s. The showed curves in these figures allow deducing some features of the parameter $\overline{B}_{res,i}$ and its value is defined in the interval

$$\Omega_{\overline{B}_{res,i}} = [L_i, H_i] = [H_i \cdot (1-\xi_i), H_i].$$

The value of $\overline{B}_{res,i}$ tends to H_i when $\frac{\lambda_i}{\mu_i}H_i < B_i$. According to this condition, it is required that the mean reserved bandwidth less than the total bandwidth B_i to offer to all the arriving requests a reservation equal to H_i.

The value of $\overline{B}_{res,i}$ tends to L_i when $\frac{\lambda_i}{\mu_i}L_i > B_i$. According to this condition, it is required that the mean reserved bandwidth higher than the total bandwidth B_i to offer to all the arriving requests a reservation equal to L_i.
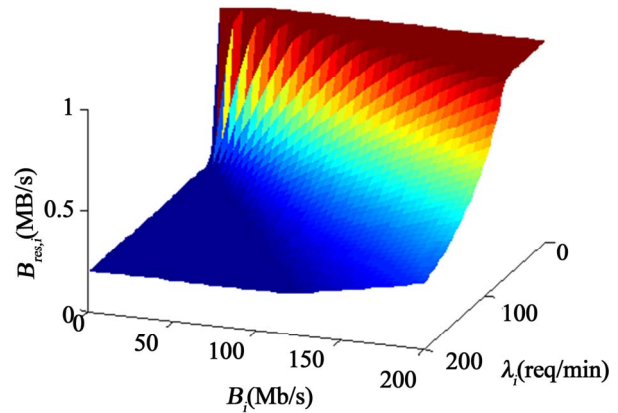

Figure 12. $\overline{B}_{res,i}$ vs. λ_i and B_i, for $H_i = 1$ Mb/s and $\xi_i = 0$.

Figure 13. $\overline{B}_{res,i}$ vs. λ_i and B_i, for $H_i = 1$ Mb/s and $\xi_I = 0.8$.

In terms of the elasticity of the reservations ξ_i, the value of $\overline{B}_{res,i}$ tends to $L_i = H_i(1-\xi_i)$ when

$$\frac{\lambda_i}{\mu_i}L_i = \frac{\lambda_i}{\mu_i}H_i(1-\xi_i) > B_i.$$

Similarly, if

$$\xi_i < 1 - \frac{B_i}{H_i}\frac{\mu_i}{\lambda_i},$$

the $\overline{B}_{res,i}$ tends to $L_i = H_i(1-\xi_i)$.

The value of $\overline{B}_{res,i}$ is placed in the range from H_i to L_i when

$$\frac{\lambda_i}{\mu_i}L_i \le B_i \le \frac{\lambda_i}{\mu_i}H_i.$$

In particular, the value of $\overline{B}_{res,i}$ can be approximated using the linear expression, $\overline{B}_{res,i} = \frac{\mu_i}{\lambda_i}B_i$.

Therefore, gathering the above results, the value of $\overline{B}_{res,i}$ is summarized in (16).

$$\bar{B}_{res,i} = \begin{cases} H_i & \text{if} & B_i > \dfrac{\lambda_i}{\mu_i} H_i \\[2ex] \dfrac{\mu_i}{\lambda_i} B_i & \text{if} & \dfrac{\lambda_i}{\mu_i} H_i \left(1 - \xi_i\right) \leq B_i \leq \dfrac{\lambda_i}{\mu_i} H_i \\[2ex] L_i & \text{if} & B_i < \dfrac{\lambda_i}{\mu_i} H_i \left(1 - \xi_i\right) \end{cases} \quad (16)$$

Figure 14 shows the relative error (higher than 3%) in the calculation of the $\bar{B}_{res,i}$ when its value is the deduced in expression (16). As it can be observed, the relative error is higher when $\dfrac{\lambda_i}{\mu_i} H_i \left(1 - \xi_i\right) = B_i$, and the value of $\bar{B}_{res,i}$ tends to L_i.

Appendix B. The Aggregate Demand Function

As other pricing methods, this method needs to know the service demand of the users in order to deduce the price of the service and therefore the optimal elasticity of the reservations. The aggregate demand function of each service is difficult to know for the SPs in advance as many researchers have pointed out. The aggregate demand function usually determines the quantity of the product that all customers want to buy at a given price. In other words, it represents the sum of the individual demands of every user, due to the fact that each user has a different willingness to pay for the product. The most common aggregate demand functions are: linear, exponential, constant demand elasticity and logit functions [32]. The aggregate demand is not always easy to obtain since it means to know the curb that shows the desire of the users to pay for the service. In any case, it must be always estimated by each SP using the appropriated tools. This demand function could also depend on different service characteristics other than price such as a quality factor of the elastic reservations as well as other marketing variables. In this paper, an equilibrium condition is assumed, that is, the average rate of the accepted reservation requests for each class of service should be equal to its aggregate demand function (17).
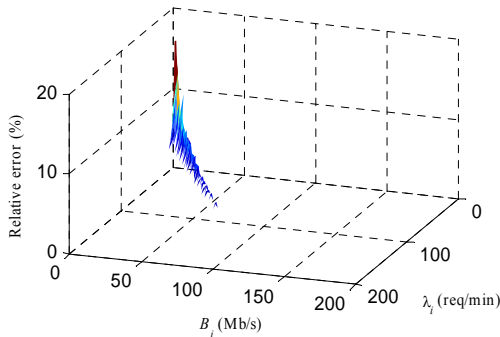

Figure 14. Relative error (higher than 3%) for the calculation of the $\bar{B}_{res,i}$ of Figure 13 using the approximation in (16).

This paper also assumes that the aggregate demand function for each service depends not only on its price (€/reservation), p_i, but on a demand modulation factor, $\bar{B}_{res,i}$ that acts as a quality factor of the elastic reservation. First are revised some of the general assumptions made for common aggregate demand functions [32], and then these assumptions are applied to a linear-based aggregate demand function, which is the selected demand function applied to the demand service in Section 4. Finally, the demand function can be expressed in terms of the price of the service and the used access bandwidth.

So, in equilibrium, the average rate of the accepted reservation requests, λ_i, is equal to the aggregate demand function D_i for each pair of values (p_i, $\bar{B}_{res,i}$).

$$\lambda_i = D_i \left(p_i, \bar{B}_{res,i} \right) \text{ where } p_i \geq 0 \text{ and } \bar{B}_{res,i} \geq 0 \quad (17)$$

In general, an aggregate demand function D must satisfy the following common regularity conditions [32].

1) D is continuously differentiable and strictly decreasing on p, $\dfrac{\partial}{\partial p} D < 0$, $\forall p, \bar{B}_{res} \geq 0$. Therefore, the demand for the service decreases with the price.

2) D is non-decreasing on

$$\bar{B}_{res}, \quad \dfrac{\partial}{\partial \bar{B}_{res}} D \geq 0,$$

$\forall p, \bar{B}_{res} \geq 0$. Therefore, the demand for the service increases or remains constant with the quality of the reservations.

3) D tends to a maximum demand when $\bar{B}_{res}$ is high enough for a fixed price. Therefore, the demand will remain constant if the price is fixed and $\bar{B}_{res}$ is higher than the maximum required bandwidth to deliver the content (B_{top}).

4) D is bounded below and thus tends to zero for high enough prices or when $\bar{B}_{res}$.

The selection of a linear-based function as the aggregate demand function notably simplifies the involved equations and their graphical representation. However, it should be clear that the selection of the aggregate demand function could be another one and, in any case, it will always depend on the information that the SP has about the service and the behavior of their users. In the aim to evaluate the price in this paper the general conditions previously deduced are applied to the selected linear-based function.

In (18) is defined the expression of the demand for a linear-based function that accomplishes condition 1) such as expression (19) shows.

$$D(p) = a_1 - a_2 p \quad a_1 > 0, \ a_2 > 0 \text{ and } p \geq 0 \quad (18)$$

Since

$$\dfrac{\partial}{\partial p} D(p) = -a_2 < 0 \text{ and } p \geq 0, \quad (19)$$

the dependence between D and $\bar{B}_{res}$ is defined as follows:

$$D\left(\bar{B}_{res}\right) = b_1 + b_2\bar{B}_{res} \quad \bar{B}_{res} \geq 0 \tag{20}$$

Condition 2) is accomplished as expression (21) shows:

$$\frac{\partial}{\partial \bar{B}_{res}} D\left(\bar{B}_{res}\right) = b_2 \geq 0, \ \bar{B}_{res} \geq 0, \tag{21}$$

Condition 3) states that the demand for the service will reach a maximum with respect to $\bar{B}_{res}$ for values equal to or greater than B_{top}. This maximum depends on the fixed value of p, expressions (22) to (25).

$$D\left(B_{top}\right) = D_{max}\left(p\right) = D\left(p\right) \Rightarrow b_1 + b_2 B_{top} = D\left(p\right)$$
$$\Rightarrow b_2 = \frac{a_1 - a_2 p}{B_{top}} - b_1 \quad \forall \bar{B}_{res} < B_{top} \tag{22}$$

Also considering that, for a given price, the demand will be constant when $\bar{B}_{res} \geq B_{top}$, we can deduce:

$$b_1 = D\left(p\right) = a_1 - a_2 p \quad \forall \bar{B}_{res} \geq B_{top} \tag{23}$$

$$b_2 = 0 \quad \forall \bar{B}_{res} \geq B_{top} \tag{24}$$

The absolute maximum demand $\left(D_{max}\right)$ is achieved when $p = 0$ and $\bar{B}_{res} \geq B_{top}$:

$$D_{max} = D\left(p = 0\right) = a_1 \tag{25}$$

The first part of condition 4) means that the demand will tend to zero as the price approaches some value P_{max}, (26).

$$D\left(P_{max}\right) = 0 \ \Rightarrow a_1 - a_2 P_{max} = 0 \ \Rightarrow P_{max} = \frac{a_1}{a_2} \tag{26}$$

The second part of condition 4) states (27):

$$D\left(\bar{B}_{res} = 0\right) = 0 \ \Rightarrow b_1 + b_2 \cdot 0 = 0 \ \Rightarrow b_1 = 0 \tag{27}$$

This paper assumes that in equilibrium the average rate of the accepted reservation requests for each class of service should be equal to its aggregate demand function as expression (17) shows. Therefore, the analytical demand function D_i can be expressed as in (28). **Figure 15** represents the linear function that specifies the sum of all reservations from all users that arrive at the SP's server, when $D_{max} = 120$ requests/minute, $P_{max} = 10€$ and $B_{top} = 1$ Mb/s.

$$\lambda_i = D_i\left(p_i, \bar{B}_{res,i}\right)$$
$$= \begin{cases} D_{max,i} \cdot \left(1 - \dfrac{p_i}{P_{max,i}}\right) \cdot \dfrac{\bar{B}_{res,i}}{B_{top,i}} \ p_i \leq P_{max,i}, & 0 \leq \bar{B}_{res,i} < B_{top,i} \\[3mm] D_{max,i} \cdot \left(1 - \dfrac{p_i}{P_{max,i}}\right) p_i \leq P_{max,i}, & \bar{B}_{res,i} \geq B_{top,i} \\[3mm] 0, & p_i > P_{max,i} \end{cases} \tag{28}$$

According to condition 1), the expression for D_i must be an invertible function. Therefore, by means of expression (28) the price p_i is determined and this value is what adapts the service demand to the value of the average rate of the accepted reservation requests, λ_i (29).

$$p_i\left(\lambda_i, \bar{B}_{res,i}\right)$$
$$= \begin{cases} \left(1 - \lambda_i \dfrac{B_{top,i}}{\bar{B}_{res,i} \cdot D_{max,i}}\right) \cdot P_{max,i} & 0 \leq \bar{B}_{res,i} < B_{top,i}, 0 \leq \lambda_i \leq D_{max,i} \dfrac{\bar{B}_{res,i}}{B_{top,i}} \\[3mm] \left(1 - \dfrac{\lambda_i}{D_{max,i}}\right) \cdot P_{max,i} & \bar{B}_{res,i} \geq B_{top,i}, 0 \leq \lambda_i \leq D_{max,i} \end{cases} \tag{29}$$

Figure 15 shows the demand function in terms of price and $\bar{B}_{res}$. The price is selected by the SP, but $\bar{B}_{res}$ is determined by the real demand (the average rate of the accepted reservation requests, λ_i), the bandwidth (B_i) and the elasticity of reservations $\left(\xi_i\right)$, as it is explained in Appendix A.2. Therefore, **Figure 15** shows the general users' aggregate behavior with respect to the price and $\bar{B}_{res}$, but cannot describe the real demand with respect to the price for a particular bandwidth and elasticity. Accordingly, we particularize the demand function of **Figure 15** to the particular service that the SP is offering that will determine the $\bar{B}_{res}$ perceived by the users. To make this particularization, we search a value of λ_i that for a given price makes $\bar{B}_{res}$ to be the same that the one obtained using the model of Appendix A.

Figures 16 and **17** show the particularized demand function with respect to price and bandwidth B_i, for $\xi_i = 0$ and $\xi_i = 0.8$ respectively, where the mean holding time is $1/\mu_i = 3$ minutes (exponential) and $H_i = 1$ Mb/s. In addition, we can approximate the particularized demand function by using (29) and (32), when $B_{top,i} < H_i\left(1 - \xi_i\right)$:

$$\lambda_i = D_i\left(p_i, B_i\right) = \begin{cases} D_{max,i} \cdot \left(1 - \dfrac{p_i}{P_{max,i}}\right) & p_i \leq P_{max,i} \\[3mm] 0 & p_i \geq P_{max,i} \end{cases} \tag{30}$$

If $H_i\left(1 - \xi_i\right) \leq B_{top,i} \leq H_i$:

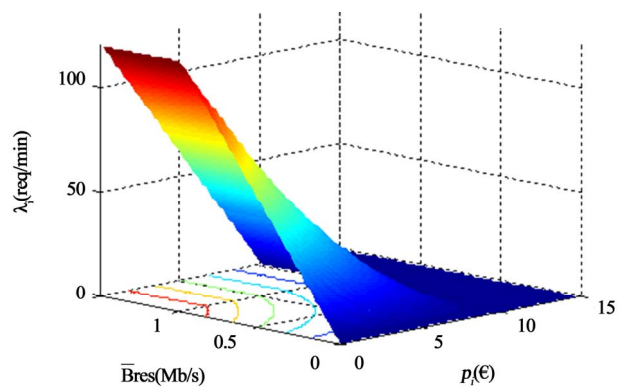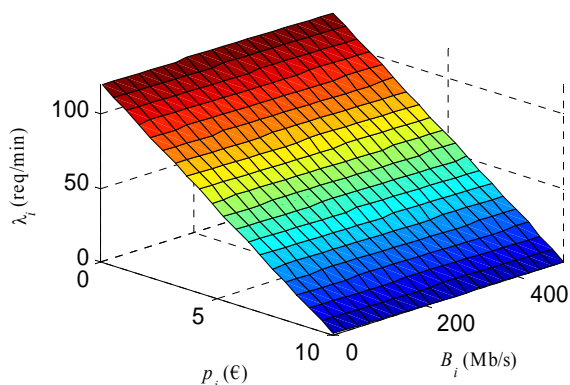**Figure 15. Demand function when D_{max} = 120 requests/minute, P_{max} = 10 €and B_{top} = 1 Mb/s.**

Figure 16. Particularized demand function for $\xi_i = 0$**,** D_{max} **= 120 requests/minute,** P_{max} **= 10 €and** B_{top} **= 1 Mb/s.**
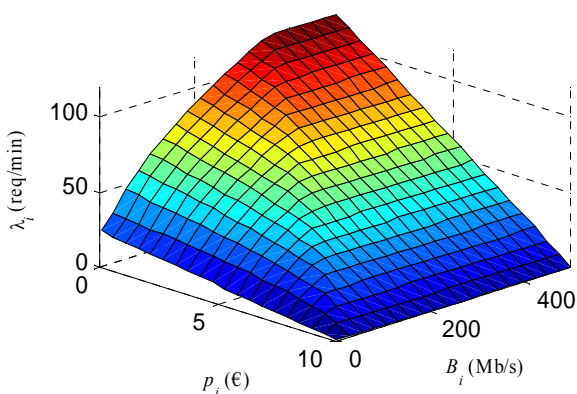

Figure 17. Particularized demand function for $\xi_i = 0.8$**,** D_{max} **= 120 requests/minute,** P_{max} **= 10€and** B_{top} **= 1 Mb/s.**

$$\lambda_i = D_i \left(p_i, B_i \right) = \frac{D_{\max,i}}{B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) H_i \left(1 - \xi_i \right)$$

when $p_i \leq P_{\max,i}$, $B_i \leq \frac{1}{\mu_i} \frac{D_{\max,i}}{B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) H_i^2 \left(1 - \xi_i \right)^2$,

$$D_i \left(p_i, B_i \right) = \sqrt{\frac{D_{\max,i}}{B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) \mu_i B_i}$$

when $p_i \leq P_{\max,i}$,

$$\frac{D_{\max,i}}{\mu_i B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) H_i^2 \left(1 - \xi_i \right)^2 \leq B_i$$

$$\leq \frac{D_{\max,i}}{\mu_i} \left(1 - \frac{p_i}{P_{\max,i}} \right) B_{top,i},$$

$$D_i \left(p_i, B_i \right) = D_{\max,i} \cdot \left(1 - \frac{p_i}{P_{\max,i}} \right)$$

when $p_i \leq P_{\max,i}$, $B_i \geq \frac{1}{\mu_i} D_{\max,i} \left(1 - \frac{p_i}{P_{\max,i}} \right) B_{top,i}$, (31)

$$D_i \left(p_i, B_i \right) = 0$$

when $p_i \geq P_{\max,i}$.

If $B_{top,i} \geq H_i$:

$$\lambda_i = D_i \left(p_i, B_i \right) = \frac{D_{\max,i}}{B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) H_i \left(1 - \xi_i \right)$$

when $p_i \leq P_{\max,i}$, $B_i \leq \frac{1}{\mu_i} \frac{D_{\max,i}}{B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) H_i^2 \left(1 - \xi_i \right)^2$,

$$D_i \left(p_i, B_i \right) = \sqrt{\frac{D_{\max,i}}{B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) \mu_i B_i}$$

when $p_i \leq P_{\max,i}$,

$$\frac{D_{\max,i}}{\mu_i B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) H_i^2 \left(1 - \xi_i \right)^2 \leq B_i \leq \frac{D_{\max,i}}{\mu_i} \left(1 - \frac{p_i}{P_{\max,i}} \right) H_i^2,$$

$$D_i \left(p_i, B_i \right) = \frac{D_{\max,i}}{B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) H_i$$

when $p_i \leq P_{\max,i}$, $B_i \geq \frac{1}{\mu_i} \frac{D_{\max,i}}{B_{top,i}} \left(1 - \frac{p_i}{P_{\max,i}} \right) H_i^2$,

$$D_i \left(p_i, B_i \right) = 0$$

when $p_i \geq P_{\max,i}$.

(32)

Also, we can express the value of price as following. If $B_{top,i} \leq H_i \left(1 - \xi_i \right)$:

$$p_i \left(\lambda_i, B_i \right) = \begin{cases} P_{\max,i} \cdot \left(1 - \frac{\lambda_i}{D_{\max,i}} \right) & \lambda_i \leq D_{\max,i} \\ 0 & \lambda_i \geq D_{\max,i} \end{cases} \quad (33)$$

If $H_i \left(1 - \xi_i \right) \leq B_{top,i} \leq H_i$:

$$p_i \left(\lambda_i, B_i \right) = P_{\max,i} \left(1 - \frac{\lambda_i}{D_{\max,i} H_i \left(1 - \xi_i \right)} B_{top,i} \right)$$

when $\lambda_i \leq \frac{D_{\max,i}}{B_{top,i}} H_i \left(1 - \xi_i \right)$, $B_i \leq \frac{\lambda_i}{\mu_i} H_i \left(1 - \xi_i \right)$,

$$p_i \left(\lambda_i, B_i \right) = P_{\max,i} \left(1 - \frac{1}{\mu_i} \frac{\lambda_i^2}{D_{\max,i}} \frac{B_{top,i}}{B_i} \right)$$

when $\lambda_i \leq \sqrt{\frac{D_{\max,i}}{B_{top,i}} \mu_i B_i}$, $\frac{\lambda_i}{\mu_i} H_i \left(1 - \xi_i \right) \leq B_i \leq \frac{\lambda_i}{\mu_i} B_{top,i}$,

$$p_i \left(\lambda_i, B_i \right) = P_{\max,i} \cdot \left(1 - \frac{\lambda_i}{D_{\max,i}} \right)$$

when $\lambda_i \leq D_{\max,i}$, $B_i \geq \frac{\lambda_i}{\mu_i} B_{top,i}$,

$$p_i \left(\lambda_i, B_i \right) = 0 \qquad \text{otherwise.}$$

(34)

If $B_{top,i} > H_i$:

$$p_i\left(\lambda_i, B_i\right) = P_{\max,i}\left(1 - \frac{\lambda_i}{D_{\max,i} H_i\left(1-\xi_i\right)} B_{top,i}\right)$$

when $\lambda_i \leq \dfrac{D_{\max,i}}{B_{top,i}} H_i\left(1-\xi_i\right), B_i \leq \dfrac{\lambda_i}{\mu_i} H_i\left(1-\xi_i\right),$

$$p_i\left(\lambda_i, B_i\right) = P_{\max,i}\left(1 - \frac{1}{\mu_i}\frac{\lambda_i^2}{D_{\max,i}}\frac{B_{top,i}}{B_i}\right)$$

when $\lambda_i \leq \sqrt{\dfrac{D_{\max,i}}{B_{top,i}}\mu_i B_i}, \dfrac{\lambda_i}{\mu_i} H_i\left(1-\xi_i\right) \leq B_i \leq \dfrac{\lambda_i}{\mu_i} H_i,$ (35)

$$p_i\left(\lambda_i, B_i\right) = P_{\max,i}\cdot\left(1 - \frac{\lambda_i}{D_{\max,i}}\frac{B_{top,i}}{H_i}\right)$$

when $\lambda_i \leq \dfrac{D_{\max,i}}{B_{top,i}} H_i, B_i > \dfrac{\lambda_i}{\mu_i} H_i,$

$$p_i\left(\lambda_i, B_i\right) = 0 \qquad \text{otherwise.}$$