A new pricing mechanism for a high-priority DiffServ-based service

Christos Bouras[1,2], Afrodite Sevasti[2,3]
[1]*RA Computer Technology Institute-RACTI, Kolokotroni 3, 26221 Patras, Greece*
[2]*Department of Computer Engineering and Informatics, University of Patras, 26500 Rion, Patras, Greece*
[3]*Greek Research and Technology Network-GRNET, 56 Mesogion Av., 11574, Athens, Greece*
bouras@cti.gr, sevasti@grnet.gr

***Abstract* — QoS provisioning according to the DiffServ framework has raised the requirements for pricing mechanisms that preserve the potential and flexibility of DiffServ. At the same time, such mechanisms should reflect resource usage, allocate resources efficiently, reimburse costs or maximize service provision profits and lead customers to requesting services that will maximize their revenue. This work proposes a policy for pricing based on resource allocation by a particular category of DiffServ-based services for aggregated traffic in the case of transport networks. Our research takes into account the particularities that apply to the case of DiffServ services' provision over transport networks while imposing minimal overload and a-priori estimation of costs. The detailed pricing methodology is presented and experimentally evaluated.**

***Keywords* — Internet Service Pricing, SLAs and business process management, Quality of Service Management, Service design, provisioning and quality assurance**

1. Introduction

The evolution of Internet in our days is undoubtedly towards a direction of providing more advanced services than the traditional best-effort model to its users. The flat pricing models that have been in effect in the traditional Internet so far do not provide users with the incentives to make reasonable use of resources. Introducing service differentiation while preserving flat pricing will therefore not prevent users from requesting the best quality possible and thus congestion in the high-priority, high-quality services will be inevitable. Hence, an important issue in designing pricing policies for today's networks, is to balance the trade-off between engineering and economic efficiency. Another factor in designing pricing policies that have good chances of being adopted, is to keep the costs' calculation simple and the monetary amounts that the customers will be asked to pay predictable.

Our work focuses mainly on pricing schemes for services provided in accordance with the DiffServ framework, which seems to gain significant importance in transport networks worldwide. The evolution that has emerged from the introduction of service differentiation and QoS provision by specifications such as that of DiffServ has affected traditional network pricing and shifted the interest from fixed access and connection fees to usage-based fees. Usage-based fees have been considered appropriate to account for congestion costs, differentiated services, QoS provision and other relevant costs for pricing today's connectionless IP networks ([1]).

With respect to the research work performed in this area, the establishment of long-term contracts between the customer and the service provider, instead of detailed accounting, was proposed in [3]. The traffic profiles comprising these contracts are in turn a good approximation of the 'expected capacity' that the customer purchases from the network services' provider. 'The effective bandwidth of a flow is considered a quantity that represents the 'expected capacity' that a customer buys when signing an SLA (Service Level Agreement), specifying a certain traffic profile and resource allocation, with a provider. In [4], [5] two compatible approaches for charging flows that obey to traffic contracts (or SLAs) according to their effective bandwidth are presented.

The 'smart market' approach that was introduced in [2], requires customers to declare their willingness to pay by bidding for network resources for each packet sent. This way, each customer is charged for the marginal cost imposed by the transmission of an additional packet during congestion. The clearing price, determined from the bids supplied, is then used together with per-packet accounting to charge the best effort service. Apart from the 'smart market' approach, [6] proposes another auction mechanism called "Progressive Second Price" (PSP) which does not assume any specific mapping of resource allocation to QoS. Despite the difficulties in their implementation, such bidding approaches are generally efficient for internalizing negative externalities to costs and therefore their concept will be exploited by our approach.

In [11] it is proposed that best-effort packets are blocked from entering the network in the event of congestion and only packets for which users are willing to pay a marginal congestion cost are allowed to enter. In the attempt to identify this marginal cost it is shown that its dominant component is the delay imposed by high-priority traffic to the best effort traffic. However, it occurs that a major requirement for a pricing scheme, that of predictable charges is thus difficult to achieve through a per-packet marginal cost approach.

Until today, many proposals for pricing of DiffServ-based services have followed the 'usage-based per service class' model. We claim that, for DiffServ services, a flat per packet or per transmitted-volume-unit price within a service class is not efficient from an economical and engineering point of view. There has to be some kind of differentiation in charging within the packets belonging to the same traffic class as well, especially for those packets of a customer that impose negative externalities to the rest of the service users. We propose a pricing scheme that demonstrates engineering and economic efficiency, preserves simplicity in calculation of customers' charges and effectively reveals the details of service differentiation and QoS provision. It allows for renegotiation of tariffs based on actual usage and customers' valuation of the service provided. Our approach is innovative because it anticipates for externalities hidden in the costs involved and caused by the nature of such DiffServ services and also because it goes all the way up to the determination of actual prices.

After this introductory section, section 2 outlines the architectural framework for the proposed pricing methodology. Section 3 describes the proposed methodology for provisioning and pricing bandwidth and buffer space. Section 4 outlines the proposed pricing mechanism and section 5 presents an experimental evaluation of our proposed methodology. In section 6 our proposed future work is outlined and the conclusions of our work are summarized.

2. The architectural framework

The case that will be further investigated in this work is that of pricing a high-priority, low latency QoS service for the customers of a transport network. Such a service is built according to the Expedited Forwarding Per-Hop-Behavior (EF PHB) of the DiffServ framework. For a detailed specification and analysis of the EF-based service, the reader can refer to [9]. Reliable transmission of data with the least possible end-to-end delay, almost zero packet loss and the minimum possible variation between the end-to-end delay experienced by different packets are the most crucial factors from the customer's point of view.

In an EF-based service, the provisioning of transmission resources is taken for granted and the focus shifts to the transmission quality obtained. The provision of such a service by a transport network provider has an analogy to the best-effort service provision: instead of bandwidth, the resource under contention is buffer space. The negative externalities imposed by congestion in best-effort service provision have their analogy to the negative externalities imposed by delay due to buffer occupancy and packets' waiting time in an EF-based service. In economic theory, externalities are referred to as costs (for negative externalities) or benefits (for positive ones) that do not accrue to the consumer of the good ([7]).

A pricing scheme for the EF-based service must lead each customer to select the amount of buffer space that he will buy from the provider in such a way that the negative externalities imposed by that amount of space are compensated (included in the price for this buffer space) and the customer does not have to shape his traffic, in an effort to reduce that amount, more than he can endure. Contrary to the existent approaches, what we are proposing is a distinction between the costs imposed to customers for the rate of their token bucket traffic profiles and the costs imposed to customers for the depth of their token bucket traffic profiles. This approach provides to the customers the incentives to provide EF traffic aggregates as well shaped as possible to the network provider. It also provides them with the incentive to provide the most accurate description of their detailed traffic profile (average rate and burrstones), rather than just an accurate description of their expected mean rate as proposed in [7].

For the purposes of our detailed analysis, the downstream domains or customers are modeled as sources of EF traffic. Between each of the customers and the Transport Domain (TD) there exists an SLA that specifies the characteristics (traffic envelope) of the marked as EF traffic injected by each customer into the TD and the specific bounded end-to-end delay guarantee (D) provided by the TD itself. EF aggregates are considered legitimate after being policed each one by its own token bucket (r,b) policer that imposes conformance to an average rate (r) and a maximum burst (b) to the corresponding aggregate.

3. Pricing the SLAs

Over-provisioning and careful dimensioning can be intuitively assumed to guarantee the required transmission rate and low end-to-end delay for the EF traffic aggregates traversing a transport domain. In such a situation, the utility function of customers is no longer dependent upon the amount of traffic being transmitted and the congestion experienced. It depends upon the equivalent capacity that each aggregate perceives and the quality metrics guaranteed (end-to-end delay, jitter and packet loss). We assume that over-provisioning ensures that no EF packets are dropped due to overflow in high-priority queues along the TD and that EF aggregates obtain a throughput, which is at least equal to their token bucket profiles' rate (r). Thus, the utility function of customers depends upon the rate (r) and burrstones allowance (b) purchased from the provider through the SLA S_i as well as the end-to-end delay (D) that the packets of each aggregate experience. We make the simplifying assumption that the utility perceived by the jitter guarantee is included in the delay factor. If we depict by $p(S_i)$ the costs that a customer has to pay for purchasing an SLA with the $S_i = (r_i, b_i)$ token bucket profile, then the objective of a pricing mechanism should be (apart from reimbursement of the provider's expenses for providing the EF-based service) that of maximizing

$$U_j(S_i) - c_j(D) - p(S_i) \qquad (1)$$

for each customer K_j, where $U_j(S_i)$ the utility perceived by customer K_j serviced by the TD according to SLA S_i, $c_j(D)$ is the cost of end-to-end delay D for customer K_j and $p(S_i)$ is the price to be paid by each customer signed with the SLA S_i and receiving EF treatment. Expressing $c_j(D)$ separately from $U_j(S_i)$ in (1) is one of the novelties of our approach. It is proposed to reflect that, unlike the S_i traffic profile, D is a factor involving externalities that concern the whole group of customers and this will be later addressed by our proposed pricing scheme. For ensuring reimbursement of costs for provisioning of the EF-based class (c_{EF}), the provider should charge the SLAs provided so that:

$$\sum_{i \in \{set\, of\, SLAs\, offerred\}} p(S_i) \geq c_{EF} \qquad (2)$$

The pricing mechanism proposed should aim at restricting the customer's demands in such a way that, at the equilibrium, each customer's revenue calculated by (1) is maximized, without equation (2) being violated. The principles of the proposed pricing mechanism are analytically presented in [10], while this work introduces the notion of re-negotiation phases of both the SLAs and their pricing between the customers and the EF service provider over long-term intervals. A brief outline of the pricing mechanism is repeated here for clarity purposes.

As thoroughly explained in [10], the TD provider can guarantee a worst-case end-to-end delay bound to all its customers, provided that the ratio a of the aggregated EF traffic injected to the TD links over the TD links' capacity reserved for EF traffic is bounded. The upper bound to a can be determined as a function of the capacity reserved for EF traffic on each link (C_l), assumed constant $\forall l, l \in TD$ and equal to C, the maximum rate with which the EF traffic aggregate is injected at each TD node and the maximum number of hops within the TD that a customer's EF traffic can traverse. Also, the prerequisite of over-provisioning, upon which an EF class is based, provides a lower bound for a. It requires that, if N is the set of customer aggregates routed through a node, for every node n of the TD then:

$$\sum_{i \in N} r_i \leq aC \Rightarrow a \geq \frac{\sum_{i \in N} r_i}{C} \qquad (3)$$

Assuming that each customer will ask for the highest r_i possible, the network administrator has to turn up with a set of acceptable r_i values and corresponding prices for the customers so that one or more values for a can exist without violating the aforementioned bounds. The range $\{a_{\min}......a_{\max}\}$ within which a can vary is quite limited. In fact $a_{\max}$ is constantly bounded by the upper bound, which is constant for a certain topology and traffic engineering and $a_{\min}$ is provided by (3). The TD provider can vary the selection of a value for a below $a_{\max}$ according to the total EF capacity he wishes to sell to his customers. For the rest of this section, we will assume that the TD provider selects a value for a so as to isolate the charging for EF traffic methodology from its side effects on the rest of the traffic. After the selection of a, the TD provider has to distribute a total of

$$r_{tot} = \sum_{i \in N} r_i = aC \qquad (4)$$

EF capacity among his customers. As explained in [10], the TD provider is suggested to distribute r_{tot} to his customers during the pricing mechanism's initialisation phase in a fair way according to

$$r_i = \frac{C^i_{access}}{\sum_i C^i_{access}} r_{tot} \qquad (5)$$

In this way, each customer K_i receives a share of the EF capacity available according to the capacity (C^i_{access}) of his access link to TD. In later, re-negotiation phases the TD provider might update the distribution of r_{tot} to each customer according to a ratio ρ_j that might differ from their access link ratios so that $r_j = \rho_j \times r_{tot}$, while (4) is always respected.

After the mechanism's initialization phase, we propose re-negotiation phases of all the contracted traffic profiles simultaneously over long-term intervals. During re-negotiations, each customer is able to base his new traffic profile's r value selection for the next period on statistical data for the utilization of the rate value allocated to him in the elapsed period. This data can directly be retrieved by the statistics of the token bucket policer of the customer's aggregate in the ingress of the TD, so that no per-packet accounting is required and overhead is avoided. Long-term re-negotiation phases will allow customers to evaluate their needs for resource provisioning based on solid, single-dimensional measurements and request the corresponding resources from the provider. This model will be shown to demonstrate fluctuations in the beginning, leading to more stable distribution of resources after a number of re-negotiations. Fluctuations are also possible when a new customer requires EF services from the provider.

In terms of charging the provided EF rates for each phase, the TD provider is proposed to fairly spread the cost of over-provisioning that EF traffic requires among the EF class customers. Thus, instead of charging each customer just for the EF contracted capacity r_i provided to him, the provider has to calculate EF capacity unit price according to

$$p_j^{unit} = \rho_j \times \{cost\ of\ capacity\ C\ in\ the\ TD\} \qquad (6)$$

so that the unit price occurs as if the customer is using $\rho_j \times C$ instead of the actual $\rho_j \times a \times C$ capacity for his EF traffic. The total cost for providing an EF average rate of r_i to a customer is then

$$\begin{aligned} P_j &= p_j^{unit} \times r_j = \\ &\rho_j \times \{cost\ of\ capacity\ C\ in\ the\ TD\} \times r_j \end{aligned} \qquad (7)$$

After the selection of a, the provisioning of resources for servicing EF traffic throughout TD is possible, by configuring all nodes' PQ schedulers to provide a service rate of C to the EF traffic on all TD links. It can be then shown ([8]) that the end-to-end delay D is bounded by a function of the same factors as in the upper bound of a, thus topology and capacity configuration related factors, as well as the total buffering space b_{tot} reserved at each TD router for EF traffic and the over-provisioning factor a itself.

The TD provider can thus calculate his available b_{tot} for a certain D guaranteed to its customers. It is apparent that according to the current TD's topology and capacity there is a limited amount of total buffer space at each router that can be distributed to its customers. The customers must thus be prompted by the bucket depth charging policy of the TD provider to restrain themselves from selecting large values for b_i by the fact that this will penalize them and others in terms of the delay perceived by their packets. Also the TD provider has to distribute b_{tot} among his customers so that if N is the set of all customers, it holds that

$$\sum_{i \in N} b_i \leq b_{tot} \tag{8}$$

In [10] it is explained how the 'smart market' approach already presented can be adopted so that buffering resources can be provisioned to those customers who value them most, while distribution has a direct impact on all customers (the end-to-end delay guaranteed by TD). The clearing price for a buffer position (P_b) is set at the point where the sum of demands for buffer space, starting to add from the higher-bids' demands, reaches the amount of available buffer space b_{tot}. So each customer will be notified of the cost he will have to pay for buffer space when signing a token bucket (r_i, b_i) SLA as equal to

$$P_{b_i} = b_i * P_b \tag{9}$$

where P_b is the price to be paid for a unit of b_i and should be set by the provider so that (2) holds.

In a real-life scenario, it is envisaged that the TD provider will distribute the available buffer space b_{tot} during the initialization phase according to intuitive bids placed by customers, since no real-use data will be available. At the moment of re-negotiations, instead of speculating for the future, the customers are able to place bids on the available buffer space based on the statistics of the token bucket policer (r_i, b_i) of their aggregates for the elapsed period. Again, fluctuations will be observed in the first phases or when a new customer will require EF services from the TD provider. However, since the 'smart market' and bidding are proven to successfully integrate externalities in goods provision costs, it is envisaged that in equilibrium, the buffer space will be distributed to those who value it most and are willing to compensate for the delay their bursts might cause to others.

4. Proposed pricing mechanism

Based on the theoretical analysis already made, it is proposed that the following algorithm is used for the provision and pricing of an EF-based service over a transport domain:

Step 1: Each customer agrees that his EF aggregate will be policed by a (r_i, b_i) token bucket policer as the traffic enters the TD.

Step 2: Based on his local policy for EF provisioning, the provider determines a (the provisioning factor) for EF traffic on the TD topology, so that it obeys to the lower and upper limits mentioned in section 3. For a TD topology composed of links over which 2.5 Gbps capacity is provisioned for EF traffic, a maximum fan-factor equal to 3 and a diameter h (maximum number of hops for a packet) as shown in the first column, the provisioning factor a_b for providing an end-to-end delay bound to 20 customers attached with 155Mbps links is provided in Table 1.

The value of $\sum_i r_i$ that can be supported by the TD is shown in the third column of Table 1. It is important to stress out at this point that the indicative values of a_b provided are upper bounds for the provision of end-to-end delay guarantees. Usually a TD provider will determine a value for a that also allows for the use of the majority of resources by non-priority, best-effort traffic. Provided that the TD provider's policy requires rate provisioning for priority traffic that does not exceed a_u for all backbone links (customary values for a_u are in the interval $\{0.05 \ldots 0.2\}$), then the provider has to select a so that

$$a = \min\{a_b, a_u\} \tag{10}$$

Table 1. Provisioning factor and allowed total of EF capacity for a series of h values

h	a_b	$\sum_i r_i$
3	0.6	1.5 Gbps
4	0.43	1.075 Gbps
5	0.33	825 Mbps
6	0.27	675 Mbps
7	0.23	575 Mbps

Step 3: Initialization phase for EF rate provisioning. The TD provider calculates

$$r_{tot} = \min_l a \times C_l \tag{11}$$

over all the links l of the TD topology and then distributes SLA token bucket rates to all customers according to (5). From (7), the cost for providing an EF average rate of r_i to each customer is calculated and the customers are then informed in advance about one part of the cost they will be asked to pay for the upcoming operation phase.

Step 4: Initialization phase for EF burrstones provisioning. According to the end-to-end delay demands of the applications supported and the advertised quality that the TD provider wishes to sell to all EF customers, the TD provider determines the end-to-end delay guarantee provided (D) and then calculates the buffer space b_{tot} that can be distributed among all EF customers as explained in section 3. In the case of a TD with $a = 0.05$, maximum number of hops equal to

8, a topology fan factor of 4, $MTU = 4700bytes$ and $C = 622Mbps$ the bound on D provided to all customers for different b_{tot} values is provided in Table 2.

Table 2. Bounds on end-to-end guaranteed delay in a transport domain with a maximum EF space of b_{tot} for any node

b_{tot} (pkts)	10	20	30	50	100	150	200
D (ms)	7.9	14.88	21.86	35.82	70.72	105.62	140.52

After b_{tot} is determined, SLA token bucket depths to all customers can be distributed according to

$$b_i = \left\lfloor \frac{b_{tot}}{k} \right\rfloor \tag{12}$$

where k is the total number of EF customers. As a result in the initialisation phase, each customer is asked to pay for allowed burrstones an amount of

$$P_{b_i} = \frac{b_i}{b_{tot}} \times P_b \tag{13}$$

Step 5: Operation phase. The service is initialised and provided for a number of days n_d. During the operation phase, at the interface of the edge router where each customer's EF traffic aggregate is policed according to the token bucket (r_i, b_i), the following statistics are maintained at regular intervals Δt:

$$r^i_{average} = r_i + \frac{\#of\ packets\ dropped\ by\ the\,(r_i, b_i)\,token\ bucket}{\Delta t} \tag{14}$$

$$b^i_{current} = current\ burst\ size \tag{15}$$

It is important to note at this point that these statistics can be collected without computational complexity, since only dropped packets are counted in the case of (14) and the value of a counter is recorded in the case of (15).

Step 6: SLA re-negotiation and prices' adjustment phase. After an operation phase is terminated, the statistics collected must be evaluated and the SLAs preserved or adjusted. Each customer is presented with the vectors $\{r_{average}\}, \{b_{current}\}$ for the previous operating period and, ideally, a graphic representation of the values of the collected statistics.

Based on the data collected from the previous operating period, each customer is applying for a new token bucket policer ($r_i^{'}, b_i^{'}$). The values of $r_i^{'}, b_i^{'}$ can emerge from the $\{r_{average}\}, \{b_{current}\}$ vectors in a number of ways, e.g. the mean or median or upper values of the measured statistics can be used. A negotiation phase is here required and the TD provider can apply different policies in order to reach agreements with all its clients, e.g. first-come-first-serve, or normalizing demand according to available capacity determined in Step 3, providing each customer with a token bucket rate for the upcoming operation phase equal to

$$r_i^{t+1} = \frac{r_i^{'}}{\sum_j r_j^{'}} r_{tot} \tag{16}$$

Customers are also placing bids (bid_i) for the available buffer space b_{tot} in the upcoming operation phase, taking into consideration the sampled data of the previous operating period and the delay guarantee D provided by the TD. Each bid_i is in the form of a vector

$$\hat{v} = (s_i, p_i^s) \tag{17}$$

where s_i is the number of buffer spaces requested at price p_i^s per buffer space. Thus, each customer may request a series of (s, p_s) tuples. The TD provider is evaluating all bids in the order of p_s offers, starting from the highest offer and provides all token bucket positions for which the following holds.

$$\sum_i s_i \le b_{tot} \tag{18}$$

In this way the token bucket b_i^{t+1} values for the next operating period are determined for all customers. The next operation phase can be initiated. Steps 5 & 6 are iterated continuously during the service's operation.

5. Experimental evaluation

For the evaluation of the proposed methodology and algorithm, an experimental set-up investigating the convergence of the iterative procedure of SLAs negotiation and pricing was implemented. The approach followed is rather simplistic, however it demonstrates the effectiveness of the pricing methodology proposed and how it provides to the customers the incentives to better approximate their true traffic profiles and charged in a fair and exact manner.

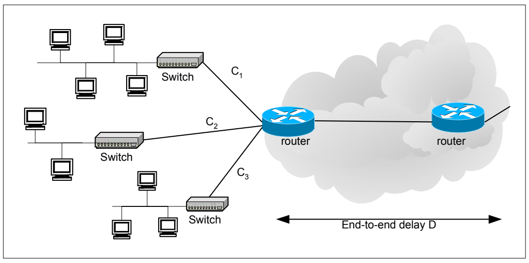

Figure 1. Experimental topology

The simple case of a TD composing of single backbone link was adopted. Three main customers inject aggregated EF traffic to the same PoP of the TD. Each customer's EF aggregate is composed of 4,2 and 3 MPEG video flows for customers C_1,C_2 and C_3 correspondingly. Each video flow is rather bursty with an average rate of 1.3 Mbps, packet size 200 bytes and an average burst size of 1700 bytes. 55Mbps are provisioned for EF traffic on the TD backbone link and an end-to-end delay of 19ms is promised by the TD provider for a value of a equal to 20.5% and $b_{tot} = 30$. Background traffic was also used to load the TD backbone link.

In Table 3 the SLA traffic descriptors that occurred during the re-negotiation phases of the experiment based on the statistical data of (14)-(15) in the form of

$\{r_i\,(Mbps), b_i\,(packets)\}$ are presented. Due to a relatively high P_b value in (9) set by the TD, the customers were led to reduce the burrstones metric b_i in their traffic contracts during the re-negotiation phases.

Table 3. Traffic descriptors for all three customers during the re-negotiation phases.

		Initiali-zation phase	Renegotiation periods						
			1st	2nd	3rd	4th	5th	6th	7th
C_1	r_i	4	4.2	4.3	4.4	4.45	4.48	4.5	4.51
	b_i	30	20	17	14	12	11	9	8
C_2	r_i	4	2.1	2.15	2.17	2.18	2.19	2.2	2.2
	b_i	5	10	9	8	9	8	6	6
C_3	r_i	5	3.3	3.4	3.4	3.45	3.5	3.52	3.53
	b_i	30	20	22	19	13	11	9	8

It is quite important to notice how, with small fluctuations, each customer updated his traffic contract throughout the iterations so as to describe more tightly his EF aggregate and shifted requested resources from the burrstones parameter b_i to the average rate r_i. Of course, the end-to-end delay bound of 19ms was never violated during all phases, since it consisted an upper bound for our case.

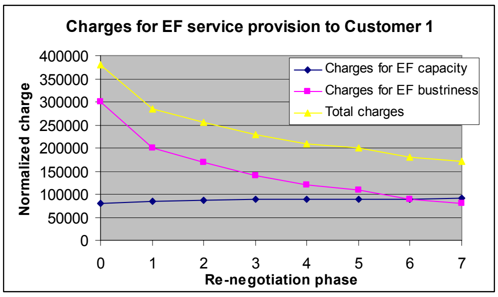

Figure 2. Evolution of the charges paid by Customer 1 during the re-negotiation phases

Finally in Figure 2, the normalized charges imposed to C_1 throughout the consecutive re-negotiation periods are depicted in a graph. One can observe how the statistical data provided to the EF service customer and the incentive-based pricing scheme proposed leads to a tighter traffic descriptor which is also economically beneficial for the customer. From the TD provider's point of view a more efficient allocation of resources is achieved. The decrease in revenue for the TD provider is compensated by new customers that can be accommodated. By providing incentives to existent customers to reveal their true traffic profiles through some iterations, the provider can become aware of the true utilization of resources in his backbone and is then able to accommodate new customers without compromising quality.

6. Future work-Conclusions

Our future work will focus on the case of the provider's profit optimization and on further investigating the customers' utility function in (1), for utility functions that are specified so as to be valid for duration longer than a connection's duration. The length of the re-negotiation is another issue that needs to be further investigated by applying the proposed schema for re-negotiation periods of different granularities and assessing its effectiveness. We also aim at dealing with the case of pricing services based on the Assured Forwarding PHB (AF PHB), as defined within the DiffServ framework.

The pricing mechanism proposed in this work is based on traffic profiles that the customers negotiate with a TD provider and concludes on prices announced to customers prior to the service provision interval. In the case of EF-based services, which is under consideration here, traffic profiles are used by the TD provider in order to dimension the EF-based service and allocate the resources used by it. The proposed pricing mechanism uses the traffic profiles of customers as the intermediate between each customer and the provider. In this way it reflects both the customers' revenue from the EF-based service provided and the costs for the service provisioning that the TD provider undertakes. Moreover, the proposed pricing mechanism takes into consideration the in-elasticity in demand for transmission rate that applies in the case of the customers of a backbone transport domain and efficiently allocates the available buffer space to those customers for which accommodation of their bursts is more valuable. Finally, the proposed mechanism provides indications of the quality that will be provided to customers (in terms of end-to end delay), in order to assist them in the qualitative valuation of the service they will receive and express accurately their needs for resources.

References

[1] L A. DaSilva, 'Pricing for QoS-Enabled Networks: A Survey', IEEE Communications Surveys & Tutorials, Vol. 3, No. 2, 2000

[2] J. MacKie-Mason and H. Varian, 'Pricing the Internet', in 'Public access to the Internet', Brian Kahin and James Keller, editors, Prentice Hall, New Jersey, 1995

[3] D. D. Clark, 'A model for cost allocation and pricing in the Internet', In L. W. McKnight and J. P. Bailey, editors, 'Internet Economics', MIT Press, 1996

[4] F.P. Kelly, 'Charging and accounting for bursty connections', in 'Internet Economics', J. P. Bailey and L. McKnight, editors, MIT Press, Massachusetts, 1996

[5] C. Courcoubetis, F. P. Kelly, and R. Weber, 'Measurement-based charging in communications networks', Technical Report 1997-19, University of Cambridge, 1997

[6] N. Semret, R. R.-F. Liao, A. T. Campbell and A. A. Lazar, 'Market Pricing of Differentiated Internet Services', Technical Report CU/CTR/TR 503-98-37, Columbia University, 1998

[7] T. Henderson, J. Crowcroft and S. Bhatti, 'Congestion Pricing: paying your way in communication networks', IEEE Internet Computing, September-October 2001, pp. 85- 89

[8] A. Charny and J.-Y. Le Boudec 'Delay bounds in a network with aggregate scheduling', in Proc. of QofIS'00, Germany, 2000

[9] C. Bouras, A. Sevasti, "Analytical approach and verification of a DiffServ-based priority service", in Proc. of HSNMC '03, Portugal, July 2003, pp. 11-20

[10] C. Bouras, A. Sevasti , 'Pricing priority services over DiffServ-enabled transport networks', IFIP 6th Interworking 2002 Conference, Australia, 13-16 October 2002, pp. 25-37

[11] L. J.Camp, C.Gideon, 'Certainty in Bandwidth or Price', The 29th Research Conference on Communication, Information and Internet Policy, Washington, D.C. October 2000