Charging IP Network Services in a Multi-Segment Infrastructure Supporting QoS and Mobility

N. Blefari-Melazzi, D. Di Sorte, M. Femminella, G. Reali
D.I.E.I. - University of Perugia – Italy
{blefari, disorte, femminella, reali}@diei.unipg.it

Abstract

Internet is becoming an integrated services network, and enhancing the classic IP Best Effort Service is a pressing need. At the same time, flat-rate pricing model, commonly applied today by Internet Service Providers to charge subscribers for the connectivity, will become inadequate soon. This paper presents usage-based tariff models to charge added value services, and the relevant accounting framework to support charging by means of a per-usage approach. The network architecture we consider is the one of the IST SUITED project, that aims at providing different degrees of QoS guarantees in a broadband terrestrial/satellite mobile Internet infrastructure. Therefore, the paper faces also the issue of pricing management in a mobile scenario.

Keywords: IP, Accounting, Mobility, QoS

1. Introduction

This work has been carried out in the framework of the SUITED project, sponsored by the EU. The network scenario of SUITED is constituted by two portions: a wireless one and a fixed one. The former is made up of four different access segments: UMTS, GPRS, a satellite network (the EuroSkyWay one [1]), and a Wireless LAN based on the 802.11 standard. The fixed part is constituted by the Internet network, interfaced with the wireless segments.

Nowadays, Internet does not provide an admission control function. The consequence is that the Quality of Service (QoS) perceived by users is not guaranteed.

The goal of the SUITED project is allowing mobile and portable broadband connectivity to end users, while providing different degrees of QoS performance to support adequately future Internet applications, such as voice and video over IP. Clearly, network has to charge users differently, according to the performance levels requested.

Presently, the pricing model mainly used in Internet is the flat-rate one, according to which the charge is not dependent on the volume of traffic involved in transmission sessions.

This approach has been one of the major reasons of the wide Internet development, but it will become soon inadequate, particularly for charging applications with different QoS requirements. As underlined in [2], “The introduction of QoS services creates a strong impetus to move to usage-based tariffs, where the tariff is based on the level of use of the network’s resources. This, in turn, generates a requirement to meter resource use...”. Therefore, it is necessary to realise an accounting architecture for supporting usage-based pricing model. As defined in [3], accounting is concerned with the collection of resource consumption data to support procedures of capacity and trend analysis, cost allocation, auditing, and billing.

Beyond the classic flat-rate approach, different usage-based pricing models have been proposed in the literature for packet networks (see [4] for a survey).

It is worthy to note that the main goals of charging are recovering costs for Internet Service Providers (ISPs), stimulating and controlling the use of shared resources. Moreover, usage-based tariffs must depend on measurement of simple and clear quantity to limit the degree of complexity of relative meter devices, and to be intelligible to users, who have the right to understand the relation between the cost and the use of services. The quantities that well fit to be measured and to be involved in the definition of a tariff are the volume of the traffic exchanged in the communication session, and the length of time of the connection.

In this paper, we present tariff models to charge the classic Best Effort Service (BES), and added value qualitative and quantitative services in the SUITED network architecture, economical procedures for setting prices are beyond the scope of this paper. We assume that pricing procedure are limited to the access ISP and are not realised in a distributed fashion along the path. We face the issue of the need of an accounting model architecture for supporting charging by means of a per-usage approach. Moreover, since the SUITED network scenario provides users with wireless
access, we consider the problem of the pricing management in a mobility environment.

The paper is organised as follows. In Section 2 we describe the SUITED project network architecture. In Section 3 we present the general tariff models to charge the network service. In Section 4 the accounting model and the accounting management architecture are proposed. In Section 5 we face the issue related to the handoff management. Finally, in Section 6 we report some conclusive remarks.

2. Network Architecture

The Integrated Services (IntServ)/RSVP [5][6] and Differentiated Services (DiffServ) [7] can be seen as complementary technologies in the pursuit of end-to-end QoS. The goal is to exploit both the possibility for the hosts to request quantifiable resources along end-to-end data paths, provided by the IntServ approach, and the scalability properties provided by the DiffServ approach. The SUITED system adopts a hybrid IntServ-DiffServ architecture, shown in Fig. 1. We stress that we enhance the standardised DiffServ model with admission control function, as suggested in literature ([2][8]).

In the core DiffServ portion of the Federated ISPs network, in order to improve the user-perceived QoS performance, an innovative resource reservation solution, called Gauge&Gate Reservation with Independent Probing (GRIP), is implemented [9][10].

Routing are classified as access, edge or core routers, according to their position in the network.

Given that one of the most challenging objectives of the SUITED project is to provide IP end-to-end QoS guarantees, appropriate and specific QoS and accounting procedures have to be executed both in the wireless and fixed segments/domains. The resource and accounting management co-ordination is required not only at connection set-up, but also while the connection is in progress.

A QoS contract includes the following fundamental information:
1) the definition of the so-called compliant traffic;
2) the definition of the QoS performance of the compliant traffic;
3) the definition of a rule for handling the non-compliant traffic;
4) the definition of the price of the service.

We assume that pricing procedures to charge users must be realised locally, at the access ISP according to the edge pricing model [11]. Each local ISP applies to its subscribers the tariffs that fit better to the requirements of its own administrative domain, particularly according to the agreements with neighbouring and/or hierarchically higher ISPs. Some Authors have presented a different solution based on the RNAP protocol [12], which provide a price negotiation distributed along the path from the source to the destination. We limit the procedures to the access ISP according to the well-known scalability requirements of the Internet architecture.

3. Service-Based Tariff Models

We consider three classes of IP service: the classic BES, the Better than Best Effort Services (BBESs), and the High Quality Services (HQSs).

In Internet, the most common method of payment is based on the flat-rate model. According to it the charge does not depend on the amount of traffic exchanged. It is necessary to meet only the expenses for accessing the service (access charge). Under this model, it is possible to consider the following payment options: free access, connection charge, periodical subscription fee, and periodical subscription fee plus connection charge. The prices are usually based on the speed of the access link. It is fair to classify connection charge as metered, depending on the time and unmetered, as for the case of access through telephone networks.

Flat-rate model presents some advantages. First of all it is very simple and does not need an added accounting architecture. The second important feature is that users and network management have an accurate idea of costs and
revenues, respectively. Finally, flat-rate charge incites users to a wide use of the service, so that it aids the development of Internet businesses.

At the same time, flat-rate model has some marked limitations. Due to the fact that charge does not depend on use, this model penalizes light users with respect to heavy ones. Moreover it is not economically efficient; in fact it does not provide the possibility to pay on the basis of the received QoS. This feature is not attractive, especially if we consider that Internet is developing as a multi service network. Finally, flat-rate charge does not favor network efficiency. It encourages resource waste, and therefore it does not guarantee high resource utilizations, while maintaining a good level of performance.

In our framework, we assume the flat-rate model for the BES only.

As attested by some studies [13], users are willing to pay, beyond the flat-rate access charge, an added usage charge for improved receiving performance (quality of service charge) on demand, for particular applications, or in situation of marked network congestion; this is very attractive for ISPs as well. This way, users can pay these per-call additional tariffs periodically, for example by means of a flat-rate fee. Otherwise, billing and payment procedures may take place on line, e.g. using rechargeable credit card bought by users according to their needs, and updated on line by the network.

We stress that when resources are particularly expensive, e.g. Internet access by means of wireless sections, it could be advisable to provide, beyond the flat-rate access charge, a per-usage charge for BES as well; in particular we assume a per-volume charge $c_v V$, where $V$ (i.e. bytes) is the traffic volume exchanged during the communication session, and where the constant $c_v$ (per-byte charge) is settled by the local ISP and depends on the speed of the access link and on the time of the day. In fact it is useful to apply tariffs which are a function of the different hours of the day (time of day pricing), so as to contribute to the congestion control. Usually, day is divided in off-peak time and peak time. A more pretentious idea is making tariffs depending dynamically on the network congestion degree, monitored on line. In this case, the charge would consist of a fixed component plus a dynamic one (congestion charge).

BBEs are qualitative services; they do not give absolute guarantees on performance, but assure that the quality will not completely degrade during network congestion periods, as regards delay and/or informative integrity. Exigent users and some Internet applications needing a minimum of guaranteed performance, like multimedia services, would benefit from this type of service. The Controlled Load Service [14] in the IntServ paradigm and the Assured Service [15] in the DiffServ model are consistent with BBEs. The traffic is assumed to be regulated by a Leaky Bucket device [16][17]. The compliant traffic is described by Leaky Bucket parameters: $(r,b)$, where $r$ is the sustainable rate, and $b$ is the token bucket size, representing the sustainable rate tolerance. The out of profile packets are classified as Best Effort ones.

We assume that this service is charged on a per-call basis, dependent both on the volume $V$ (i.e. bytes) involved and on the time duration $T$ (i.e. seconds) of the connection. This way, call charge $= c_r T + c_v V$, where the constants $c_r$ (per-second charge) and $c_v$ (per-byte charge) are a function of the sustainable rate, the type of service and the time of the day. It is our opinion that for BBEs a per-call charge is necessary, since a per-call resource reservation is involved. Note that reservation generally precludes other customer from using the reserved resources. We assume that such charge is a moderate one, because, in this case, the reservation is soft.

We call this component allocation charge; it characterises the potential use of resources and it is a protection against unfair behaviours of users. Anyway, we assume that the heaviest component of the above tariff is per-volume (effective usage charge); it depends on the actual use of resources.

HQSS aim at giving quantitative end-to-end guarantees on performance, in terms of delay and losses (real time guarantees). Future Internet applications, such as voice and video over IP, would benefit from this type of service. The Guaranteed Service [18] in the IntServ paradigm and the Premium (or Expedited) Service in the DiffServ model [19], enhanced by admission control function, are consistent with HQSS. This traffic is assumed to be regulated by a Dual Leaky Bucket device [16][17]. The compliant traffic is described by Dual Leaky Bucket parameters: $(P,r,b)$, where $P$ is the peak rate. The out of profile packets are dropped.

The quantity that is adequate to be reserved and charged is the equivalent bandwidth; it is a scalar quantity dependent on traffic characteristics and requested performance of a flow, and on the availability of resources in a node (see e.g. [20][21]); it represents the resource consumption of the flow at a node. However charging this type of service is complex, since the amount of allocated resources is conspicuous. We can consider the matter from various angles. First of all, once defined the traffic and performance descriptors, the amount of allocated resources is generally conservative, and it is often based on worst case assumptions. This is the reason why network administrators would like to charge traffic in proportion to this quantity. At the same time, it is difficult for users to predict their traffic rate process exactly, so that they would consider convenient to be charged according to the actual used resources instead of the estimated (and eventually over-estimated) ones. Some Authors [22][23] have proposed an approach that lies between the previous ones; they charge an amount of equivalent bandwidth that depends on both static (i.e. negotiated) and measured parameters. We propose a simple tariff model, that consists of both an allocation and an effective usage charge component: thus we have call charge $= c_r T + c_v V$, where $c_r$ and $c_v$ depend on the allocated equivalent bandwidth and on the time of the day. In this case, we assume that the allocation charge component is the
heaviest one, because the amount of reserved resources is substantial. It is our opinion that tariffs must also depend on the difference between the terminals. For instance, it could be advisable to classify calls into national, international and intercontinental ones. Moreover, since the concept of equivalent bandwidth is relative to a single node, it could be useful to define a standard node, to compute the equivalent bandwidth for pricing purpose univocally.

We observe that in a multi-service network it could be useful to provide each user with a user agent [24], that we call Host Intelligent Agent (H-IA). It is a software entity whose tasks are to choose the type of service depending on the users willingness to pay, the requested performance and the network status. In particular, once fixed an acceptable level of a given service, it has to maximise the customer surplus function, that is the difference between the amount of available money (utility function) and the applied charge. The H-IA has also to exchange data with its ISP counterpart (ISP Intelligent Agent, ISP-IA) about service consumption, price and resource negotiation, and billing (see Fig. 2).

Figure 2 - Interactions between a user and its local ISP.

4. Accounting Model

Accounting concerns the collection of resource consumption data to support usage-based pricing. It is necessary a general model supporting both the realisation and the management of Internet accounting procedures.

As regards the former feature, a model of recursive accounting has been proposed in [25]. To describe it, we refer to the general network architecture model shown in Fig. 3. Local (or access) ISPs offer connectivity to users, and make use of the network service of hierarchically higher ISPs (backbone providers). The whole network is divided in different administrative domains. Each domain applies tariffs that fit better to its own requirements, according to the models of the previous Section. We suppose that the administrator of domain D is interested in charging on a per-usage basis the traffic passing through its network. It is possible to identify two traffic classes. The internal traffic, that has source and destination within D, and the external one, that has the source and/or the destination outside D. Correspondingly we can speak of intra-domain and inter-domain accounting.

Figure 3 - Network architecture model.

If we consider a usage-based pricing model, the first step is determining the subject responsible for payments.

At first, we assume the sender-pays approach, according to which such subject is the sender. As regards the intra-domain accounting, it is sufficient implementing measurement procedures at the access routers, and then processing and storing accounting data for successive billing. This approach may be extended to inter-domain accounting, in which measurements are realised at edge routers as well. For example, consider a user belonging to domain A, who sends traffic to a user belonging to domain D. The relevant traffic is measured at the edge router between A and D and, according to these measurements and the adopted tariffs, D charges A. Then, if A wants to charge its customer by means of a per-usage approach, it measures his traffic at the relative access router. Note that, to identify the subject responsible for payments, it is necessary to realise a packets classification based on a voucher [26] in the IPng packets, updated along the path. It indicates, time by time, the subject who sent the packet: an host at access routers or an administrative domain at edge routers.

The sender-pays model is not always adequate. The receiver, and not the sender, is often the adequate subject responsible for payments. It is sufficient thinking to a better than best effort data download from WWW servers, or to an high quality IP telephone call, in which both corresponding users send data. Different solutions can be applied. In the first case, users can pay by electronic money the service to the server, which is in turn responsible for the transmission charge. Therefore, the receiver is only indirectly responsible for payments to the network (dual approach charging [27]). This approach protects users against unfinished or incorrect services due to server inefficiencies. Note that users can also request to the server a determined level of service. In this way, the receiver is in charge of selecting the level of QoS. It is up to the server negotiating such level with the network.

We call this model dual approach QoS negotiation. Another solution presented in [28], suited for instance for telephone applications, is tagging each packet at the network access...
point by an appropriate field. This field, that we call charge bit, indicates if the responsible for payments of the per-volume charge is the receiver or the sender. For example, if the charge bit is equal to 1, then the responsible for payments is the receiver, determined at nodes by routing algorithms; otherwise, if it is 0, the responsible is the sender, and the measurement devices read the voucher to determine it. In the former case, money and data flows go in the same direction, while in the latter one they follow opposite directions, as shown in Fig. 4.

![Figure 4 – Money and data flows.](image)

The subject responsible for the per-time charge is clearly settled during the set-up phase.

Note that it is also possible thinking to a division of charge between the interested terminals, for instance depending on the zones crossed by the flow (zone based cost sharing [28]).

Another important issue for accounting purposes is the realisation of an accounting management architecture within each administrative domain. In [3] three main entities are identified: network devices, accounting servers, and billing servers.

Network devices collect measurement data, and send them by means of an accounting protocol (RADIUS, TACACS+, DIAMETER, COPs or SNMP for instance) to accounting servers, which generate the session records in which are stored the accounting information [29]. As mentioned above, network devices are implemented at access routers and at edge routers; network devices and accounting servers can be only logically distinct. Session records are then sent, with HTTP, SMTP, or FTP, to the billing server, which generate invoices to its customers and to the neighboring domains. The rules for generating, transporting and storing accounting data are known as accounting policies [30].

Interactions among the three entities are shown in Fig. 5.

![Figure 5 – Accounting management architecture.](image)

5. Pricing Management in a Mobile Environment

Since the SUITED scenario provides also wireless and mobile multi-segment access, we need to face the issue of the pricing management when the active mobile user moves from an access segment of his administrative domain to a different access segment belonging to the same or to a different administrative domain. In fact, once the type of service is fixed, it may be that different domains apply different tariffs, and it is clear that, even in the same domain, different access segments can charge different prices (e.g., the satellite segment is probably more expensive than the other segments).

To provide mobility, a user entity has to be introduced for the handling of a seamless handover, intending that an application session can go on without noticing the change of access segment. This entity is called InterWorking Unit (IWU), and it is located in the mobile terminal [31]. The IWU is a router with interfaces connected to several radio modems and with a call set-up layer for each segment.

The availability of all segments is known to the IWU by means of receive power feedbacks. Some reasons for starting a handoff procedure could be for instance:

1) a user leaves the coverage of a segment (or a domain);
2) a cheaper (in terms of cost) segment (or domain) becomes available.

We focus our attention on pricing issues, when the handoff implies a tariff change.

When a user faces a handoff procedure, we assume that an amount of resources equal to the previously allocated ones will be available in the new segment/domain as well. Under this hypothesis, we distinguish between two types of handoff: the forced handoff and the voluntary handoff.

In the former case, the handoff procedure is needed to maintain the connection active, as in the first of the two events listed above. Such procedure needs a pricing negotiation between the user and the administrative domain, depending
on the requested type of service and on the new tariff relative to the interested segment. Each access domain applies its own tariffs according to the models previously described. This is the reason why a user has to choose, depending on his willingness to pay as well, the new QoS level to be requested. In particular, when the new segment/domain charges higher prices, the user decides if accepting a worst service while maintaining the same tariff (or a similar one), or if accepting to pay an higher tariff, in order to maintain the same QoS of his connection.

In the case of voluntary handover the change of segment/domain is performed because there could be the possibility to receive the same level of service with a lower cost, or a better service with the same cost, (see the second of the events listed above). In this case, the IWU reveals radio signal power of different segments/domains higher than a minimum pre-defined threshold, deemed necessary to provide connectivity. Therefore it is necessary to evaluate the best trade-off between the price and the required level of service. The previous considerations, can be formalised in the definition of a decision function in the user terminal. The algorithm can be a priori settled on the basis of the user preferences in terms of desired QoS and willingness to pay; this way, the handoff results transparent from the pricing management point of view as well. In the definition of such decision function, we have to satisfy two basic constraints: the user perceived performance has to be better than a minimum reference level, and the price has to be lower than a maximum value. Once these constraints are fulfilled, the decision function has to calculate the best trade-off. At the end of the price negotiation, the IWU starts the call set-up phase in order to require the QoS level previously chosen.

6. Conclusions

The main objective of this paper has been to present an accounting framework supporting different pricing policies in the hybrid IntServ-DiffServ environment proposed for the SUITED project. A classic flat-rate model for the classic BES, and a usage-based model for BBESs and HQSs have been proposed. We have presented tariff models to charge added value services, and the accounting model for supporting usage-based charging. We have assumed that pricing procedures are limited to the local ISP, instead of considering them distributed along the path from the source to the destination. Moreover, since the SUITED network architecture provides mobility to users, we have considered the issue of price negotiation in the mobile scenario as well.

References

[10] G. Bianchi, N. Blefari-Melazzi, "A migration path for the Internet: from best-effort to a QoS capable infrastructure by means of localized admission control," Lecture Notes on Computer Science, Springer-Verlag, volume 1989, M. Aymone-Marsan, A. Bianco, Eds. (this paper and also a more detailed technical report can be requested to the authors by writing to blefari@diei.unipg.it).


