25

Simple integrated media access - a comprehensive service for future internet

J. Ruutu [a] and K. Kilkki [b]
Nokia Research Center

[a] P.O.Box 422, FIN-00045 NOKIA GROUP, Finland
tel:+358 9 4376 6188, fax: +358 9 4376 6851
e-mail: jussi.ruutu@research.nokia.com
[b] 3 Burlington Woods Drive , Suite 260, Burlington, MA 01803,
USA, tel: + 1 781 564 9609, fax: +1 781 564 9696
e-mail: kalevi.kilkki@research.nokia.com

Abstract
The basic objectives of future Internet are to increase the network capacity, to offer a practical real-time service, and to develop a feasible charging scheme. These objectives introduce very strict requirements for the traffic control system. This paper presents a new simple approach for traffic management: Simple Integrated Media Access (SIMA). The strength of SIMA lies in its wide area of applications. There is no need to build complex systems with several service classes each appropriate to only certain applications.

Keywords
SIMA, Internet, Charging, Quality of Service, ATM

1   INTRODUCTION

The Internet is at a phase of great changes. There are several stringent new requirements because of two reasons: the invasion of new users, and the rapid development of new applications. These requirements mean that network capacity must rapidly be increased, real-time service has to be fundamentally improved, and a feasible charging scheme must be introduced. There is not too much time for the development as the penetration is growing fast (Snyder, 1997).

The current Internet approach for meeting these requirements consists of several service specifications, Resource Reservation Protocol, QoS routing, etc. We can make an interpretation that the basic philosophy of Internet development is to define different services for different basic communication needs. The supposed advantage of this approach is that by dividing the service specification task into several smaller parts the specification process is easier than if the all the service types were included in a single specification. However, this advantage is somewhat questionable because the whole service concept (with all the different service types) is what the network operator should manage and sell to customers and what the customer should buy and use.

The charging problem must also be solved before there is a real possibility to offer Internet with real quality and reasonable bandwidth. As stated in (Cukier, 1997) the developing of billing has not been a high priority for Internet engineers who prefer flat-rate charging for its simplicity. A lot of carriers would prefer to throw bandwidth at the problem of managing QoS as J. McQuillan has expressed the situation (McQuillan, 1996). This type of approach, adopted for instance by Telecom Finland (Greenfield, 1997), means that even a pure best-effort type of service is capable to meet most of the current service demand.

However, there are different kind of views as well, for instance, usage-sensitive pricing, priority-based pricing, additional value-added services and Intranet tolls are possible charging schemes (Firdmand, 1997). Especially, the priority-based pricing is a promising approach as it is based on premise that customers who want better service should pay more. The best approach is the one which can flexible combine different properties but at the same time allows a very straightforward network management and charging. We believe that the Simple Integrated Media Access (SIMA) concept is able to reach this ambitious target (SIMA, 1998).

2   SERVICE SPESIFICATION

The SIMA specification covers the whole Internet service including charging, QoS and performance aspects, and traffic control functions in the network. The primary idea of the SIMA service is to maximize the exploitation of network resources with a simple control scheme while keeping the ratios of QoS levels offered to different flows unchanged under changeable traffic conditions. The maximization is based on three key features: all flows with different QoS requirements share the total

capacity of every link, the network attempts to avoid any unnecessary packet discarding, and flow level blocking can be totally avoided. The approximate constancy of QoS ratios and simplicity are achieved by using eight drop preference levels which make possible a fair packet discarding scheme inside the network without keeping track on every flow (see Chapter 4).

2.1 Nominal Bit Rate

When the network operator offers the SIMA service, a customer first pays for some Nominal Bit Rate (NBR, kbit/s) and then he/she can trade the speed for QoS. Let us assume that a user pays X $/month. This charge is translated to a NBR using an arbitrary function. The function NBR = F(X) could be linear, but there is no reason to specify the relationship between NBR and charging.

   If NBR is permanent, it is probably related to an interface. The next level of NBR is the NBR assigned to a user (or IP-address). The bottom level is the NBR of a flow (determined, for example, by a pair of IP address and port number). Both interface based and user based approaches have the drawback that they do not separate different applications properly: a high-speed file transfer may disturb other flows, although the user may consider the file transfer a background process which uses only the capacity left by other more demanding applications. Therefore, as regards the performance and QoS of the SIMA service the most useful approach is the one where every flow has its own NBR.

2.2 Real-time vs. Non-real-time

The other part of the SIMA service concept is the possibility to request a real-time service. The user is entitled to him/herself determine whether the flow is a real-time (rt) or non-real-time (nrt) one. In practice, this decision can be made usually at the application level: a real-time service is requested only for interactive audio or video applications. If a real-time service is requested, the SIMA network attempts to offer as short delay and small delay variation as possible by using small buffers reserved only for real-time connections. The measurement for the drop preference determination shall be more sensitive for traffic variations in case of real-time service than with non-real-time service.

   If the user changes a VBR connection from nrt-service to rt-service without changing NBR or traffic process, the delay will decrease, but the cell loss ratio may increase because real-time measurement gives worse drop preferences during peak rates. If the user wants to obtain the same quality, this impairment of loss ratio should be compensated by increasing NBR. Real-time-service could, in this respect, be more expensive than non-real-time service although there is no difference in the actual tariffs. In consequence, if the application is a real-time one, it is advantageous for the user to select the real-time class, because it is the only way to attain small delay and delay variation.

2.3   Quality of Service expectations

The total SIMA service requested by a user consists of a nominal bit rate and of a possible real-time service request. This half of the service is as clear and reasonable as possible. The other half of the service is the expected QoS of the flow, or actually, the expected QoS of the application that the customer uses over the SIMA network. An essential issue for the success of the SIMA service is how reasonable and acceptable this part of the service concept will be.

In a circuit switched network a busy period means that the call blocking probability increases. In packet networks the packet loss ratio increases during busy periods, and effectively, the available capacity for a flow decreases if a TCP/IP type of protocol is used. In a SIMA environment, when a user buys a NBR for a flow and then sends traffic into a SIMA network, there is usually no flow level blocking. The quality of the flow depends on two issues: the NBR to actual bit rate ratio, and total load in the network. Therefore, a potential difficulty is that the customer cannot precisely know what the QoS of a flow will be because rapid traffic variations may bring about unexpected changes of QoS (however, even in the case of services using resource reservation the actual quality of flows using certain quality class may vary significantly).

Because the quality of existing flows is not in the same way predictable as with services using complicated resource reservation mechanism, the SIMA network shall be implemented in a way that the users can rely on the fairness of the service. The fairness of the SIMA service is based on the fact that all flows with the same actual bit rate to NBR ratio perceives similar QoS. Thus, a home user with 10 kbit/s NBR receives the same QoS as a large company with NBR of 100 Mbit/s provided that both are transmitting at their own NBR.

Another aspect of fairness is the possibility to obtain more quality with higher price or lower price with less quality by changing the actual bit rate or NBR. This means that each customer is entitled to change the NBR to actual bit rate ratio and by that means to optimize his/her quality to charge ratio. If the ratio increases, the quality of the flow is enhanced. If the user sends traffic by using a constant bit rate, the SIMA service offers 7 different quality. Although the absolute quality of each drop preference depends on the network dimensioning and on actual traffic process, the quality levels can be described approximately as follows:

    7 = reserved for non-SIMA services with resource reservation
    6 = excellent quality: negligible packet loss ratio
    5 = high quality: packet losses only during exceptional traffic peaks
    4 = good quality: small packet loss ratio even during busy hour
    3 = moderate quality: usually small packet loss ratio except during busy hours
    2 = satisfactory quality: from time to time very high packet loss ratio
    1 = suitable for best-effort traffic during busy hour
    0 = suitable for best-effort traffic during non-busy hours

The charge of drop preference j will be $X * 2^{j-4}$, if the charge of level 4 is X, and if the charging is proportional to NBR. However, quality level 0 can be in practice

obtained free of charge, and the operator may use different charging scheme with level 7. The network operator may try to dimension the network in a way that the traffic of the three lowest levels is able to fill the network during busy hour. Now, as the charge of level 6 service is 16 times higher than that of level 2, we can assume that there will be much more traffic offered to the lowest drop preferences. It is reasonable to assume that the most intense traffic variations occur at the lowest quality levels, whereas the charging may dampen the variations at the highest quality levels. Note that even very high load of the low quality levels has no significant effect on the packet loss ratio of the highest levels.

2.4 SIMA service chain

The total service chain of SIMA is outlined in Figure 1.

User	Charging	Input to network	Traffic control	Performance

C ──▶ X Ecu
          +
      tariff
      function ──▶ NBR for the flow

I ◀········· /   +

T ─────────────▶   actual offered
                      traffic of
                      the flow
                          +

P ─────────────▶   rt/nrt ─────▶ SIMA packets
                                    +
                                    SIMA traffic control
                                    +
                                    aggregate traffic
                                    process
                                    +
                                    network capacity ──▶ network
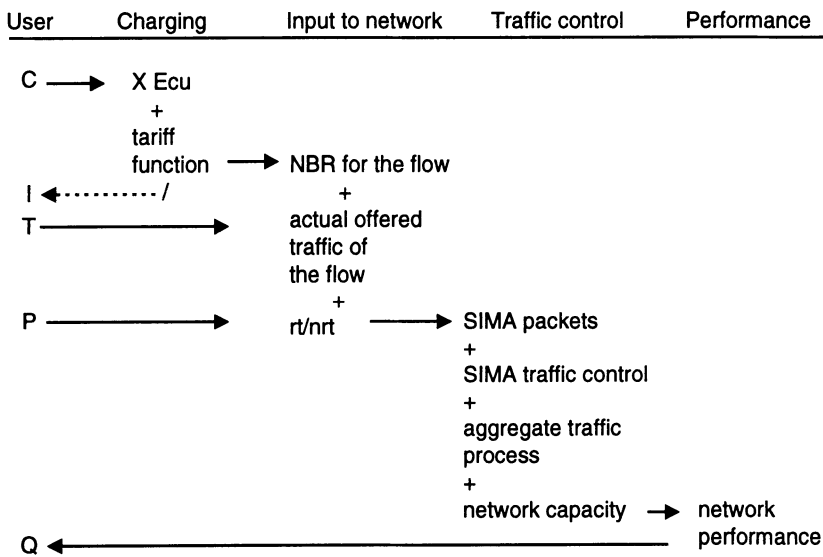Q ◀─── performance

Figure 1 Service chain of SIMA. C is the user's readiness to pay, I is information given by the network, T is actual traffic sent into the network, P is the parameters needed to control the flow, and Q is the quality experienced by the user.

The user input to the SIMA network consists of charge (C → X $), actual traffic sent into the network (T), and rt/nrt selection (the only pre-defined traffic or quality parameter used with SIMA). The network may inform the user of the offered service tariffs by announcing the NBR or as the user may know the charging function, he can select directly a proper NBR. The main output of the network is the actual QoS of the flow, which depends directly on the network performance. Note that there is a (one-way) connection from the charge of the flow to the actual QoS of the flow via NBR, the SIMA control and network performance. Therefore,

although there is no pre-defined exact relation between charging and QoS, the user may optimize the charging of the flow by trying firstly a low charge, and then doubling the charge until the quality level is sufficient. Another important feature is that the traffic control information is conveyed purely by the SIMA packets or cells, which means that there is no need to have any traffic control information transported between different network nodes. Finally, as there is no packet or cell discarding based on a separate traffic flow, packets or cells are discarded only if the total load exceeds the link capacity.

3   COMPARISON OF SIMA AND OTHER SERVICE APPROACHES

3.1   Current service schemes

The starting point of the development of Internet services is the current best-effort service model. The well-known problems of best-effort service are that there is no relation between quality and charging, and that there is no way to offer high quality (small packet loss or small delay) for those flows that need these features. The prevalent approach to solve this problem is to design guaranteed service classes, each of which has certain quality features. A simplified service chain of this approach is presented in Figure 2.

The user input to a network with a guaranteed service consists of requested traffic and quality parameters (P), and actual traffic sent into the network (T). Actually, this is quite a big difference between SIMA and guaranteed service. With SIMA a customer mainly informs how much he/she is willing to pay. With guaranteed service the customer must first predict the parameters of his/her flow, something that is not easy even for an expert. The network informs users of the charge of the flow by using a complicated tariff table including all possible combinations of traffic and quality parameters.

The guaranteed service approach means that the network attempts to give a statistical prediction of the actual quality of the flow: certain service class will generate certain average quality (it is assumed that each user is willing to understand the meaning of the parameters). However, the connection between requested quality and actual quality is more complicated than can be concluded directly from service specifications due to variations in aggregate traffic process.

The output of the network is the actual QoS of the flow, but in this case the quality (or rather, quality impairments) consists of two parts: first one (Q1 in Figure 2) contains the possible packet loss ratio due to the control of each flow (UPC in ATM networks), and the other one (Q2) contains the effects of control functions directed to the aggregate traffic load. In order to improve quality, these effects have to be discerned since either the traffic parameters or service class should be changed. This optimization needs quite profound understanding of the properties of services, network and traffic.
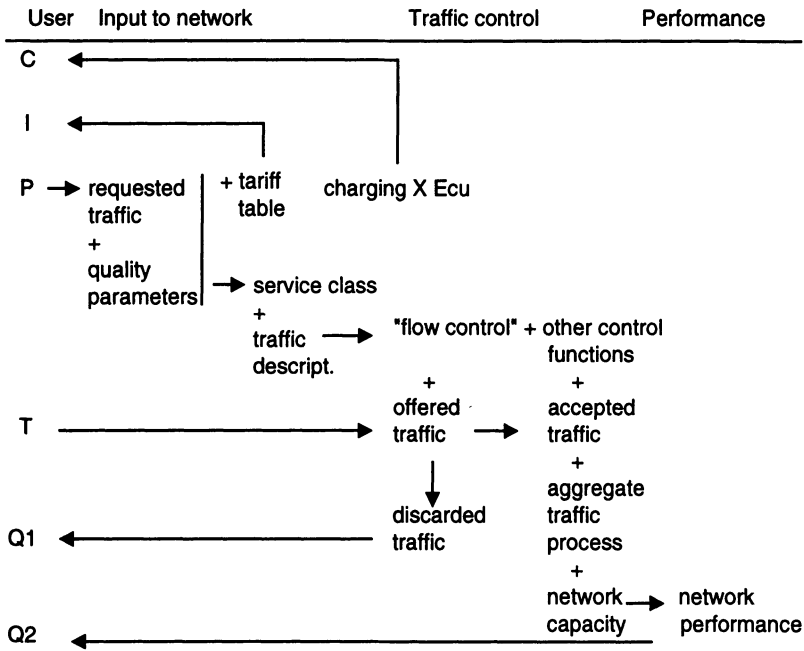
User   Input to network   Traffic control   Performance

C

I

P → requested   + tariff   charging X Ecu
     traffic   table
     +
     quality
     parameters → service class
                  +
                  traffic → "flow control" + other control
                  descript.   functions
                            +   +
                            offered   accepted
T   traffic   →   traffic
                                           +
                              ↓   aggregate
                            discarded   traffic
Q1 ◄─────────────────────── traffic   process
                                           +
                                           network → network
Q2 ◄──────────────────────────────────   capacity   performance

Figure 2 Service chain of a guaranteed service.

3.2 Service comparison

As C. Heckart has expressed it in (Heckart, 1995) gone are the days when carriers could be content to provide one small component of larger picture; carriers must now offer comprehensive service packages that increase their value to users and make it difficult for customers to find another provider to supply the same end-to-end solutions. This real need may induce a serious problem with services using guaranteed quality: how to build a reasonable service package offered to ordinary customers not familiar with technical details. The principal question is whether the SIMA service can in this respect be better than the other approaches for integrated services. It is important to note that it is not reasonable to only compare individual services realized by SIMA, or some other service models, but the whole service package offered to customers.

The prevalent approaches are service specifications developed at IETF's Integrated Service working group (IntServ), and ATM specifications. The two main services needed in future Internet are a real-time service with high quality, and a file transfer service with loose requirements for packet loss ratio and delay. In addition, some customers may benefit from a service which guarantees a small packet loss ratio but does not provide small delay. If the network operator attempts to satisfy all these requirements by using the current specifications, he must

implement several dissimilar services. Possible combinations are: guaranteed service, controlled load service and best-effort service if IETF's specifications are used (IETF, 1994), and CBR+rt-VBR, ABR and UBR if ATM is used (ATM Forum, 1996) (ITU-T, 1996). Table 1 provides a brief summary.

Table 1 Comparison of network services.

	IntServ			ATM			SIMA
	Guaran. serv.	Contr. Load	Best effort	CBR+ rt-VBR	ABR	UBR	SIMA
Charging	?	-	flat rate	based on traffic, QoS parameters	?	flat rate (+usage)	based on NBR
Traffic parameters	Many	Many	-	PCR, SCR, BT	-	-	(NBR)
Pre-defined QoS par.	yes	yes	no	yes	yes	no	no
Small delay	yes	no	no	yes	no	no	yes or no
Loss ratio	small	small	high	small	small	high	small-high
Controlled service	no	yes	no	no	yes	no	possible

One of the main problems with the current approaches is different charging principles of different services. The charging of any service relying on resource reservations is likely to be based on the traffic and QoS parameters used at the reservation phase. Best-effort service uses, instead, flat rate or usage based charging schemes. The charging of controlled load services may combine these two schemes (and be quite complicated). In total, if we take into account the need of different charging levels for busy and idle hours, the charging structure tends to be very complicated due to the large amount of parameters. On the contrary, the charging of the SIMA service can be based purely on one parameter, NBR.

When a customer requests a service she/he shall inform the network what kind of service is needed. This information consists usually of some traffic parameters and quality parameters and perhaps service class. In order to successfully use a service, the customer shall understand the meaning of these parameters (if they cannot be totally hidden from end-users), and even to make proper guess for their values. Taking into account the reluctance of many Internet users to learn technical details, the current service concepts seem to be unsatisfactory in this respect. With SIMA there is only NBR and the selection between real-time and non-real-time service. The latter selection can be usually left for the application.

The next question is whether a SIMA network can offer all the necessary service types. SIMA can provide efficient real-time service, different packet loss ratios from negligible to high, and a free combination of these two categories (delay, packet loss ratio). The most unclear service class is the controlled load service with small packet loss ratio. However, it should be stressed that if there is no strict delay requirement, a small packet loss ratio can be always attained by using efficient

upper layer protocols. Therefore, when using TCP/IP or a similar protocol, there is no urgent need for a controlled load service as an end-user service, rather the objective of the controlled load service is to optimize the use of network resources. A SIMA network offers good possibilities for an application using TCP/IP or similar protocols, because the packet loss ratio decreases rapidly when the transmission rate goes down enough (say, to a level of 3*NBR). In this respect SIMA service is essentially better than a pure best-effort service.

Because traffic and network management is one of major costs of telecommunication network, it is very important to keep management functions as simple and efficient as possible. The realization of a guaranteed service requires traffic parameters for every flow, controlling of these parameters, resource reservation at every network node, complicated signaling for the transfer of parameters, dimensioning of complicated buffer and switching structures, etc. It will be very difficult to implement and manage this type of network. In contrast, the SIMA service may work without such ordinary management functions as Traffic Descriptor, QoS parameters, Connection Admission Control (CAC), or Usage Parameter Control (UPC), etc. All these functions are replaced by the measuring unit at access nodes and the scheduling and buffering unit (SBU).

The SIMA service is able to meet the simplicity requirement essentially better than a network with several service classes, and it can satisfy the basic service needs of most customers. The remaining questions are related to the performance and QoS of SIMA networks. As to the throughput, the main advantage of SIMA is that there is no need to fragment the network capacity, instead, all services and all flows divide the whole capacity of every link. In this respect SIMA is very efficient. In a SIMA network, sufficient quality can be obtained by proper network dimensioning. The operator may offer satisfactory QoS to nominal connections (i.e., to those connections in which actual bit rate is equal to NBR). In practice, this may mean that the operator measures the average packet loss ratio of packets with drop preference 4. This ratio should remain on a reasonable level. If this packet loss ratio is exceeded continuously, the operator shall firstly identify the bottlenecks in the network and then increase the network capacity in those points. The network operator simply throws bandwidth, and the SIMA service manages QoS. This network dimensioning scheme provided by SIMA is a natural extension of the prevalent way of managing Internet.

4   IMPLEMENTATION OF SIMA SERVICE

The implementation of the SIMA service consists of two main parts: access nodes and core network nodes presented in Figure 3. There is a fundamental difference between these node types: the traffic measurement of every flow is performed at access nodes whereas at the core network nodes the traffic control functions do not need to know anything about the properties of separate flows.
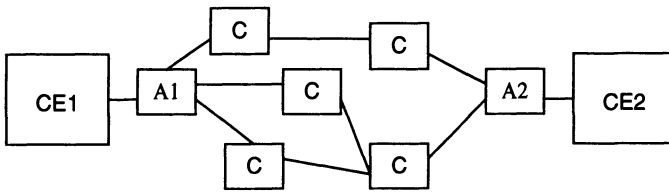
Figure 3 Customer equipment (CE1) connected to an other customer equipment (CE2) through a SIMA network with access nodes (A) and core nodes (C).

4.1. Access node functions

Let us suppose that there is an IP flow (i) at an access node. A nominal bit rate, NBR(i), is associated to the flow and the user is transmitting IP packets (which may be converted into ATM cells) into the network according to an arbitrary traffic process. At the user/network interface there is a measuring device which measures the momentary bit rate of the flow at the arrival of the j:th packet (or cell). This rate is denoted by MBR(i,j). The device gives every packet a drop preference, DP(i,j), based on the MBR(i,j) to NBR(i) ratio:

$$x = 4.5 - \ln\left(\frac{MBR(i,j)}{NBR(i)}\right) / \ln(2)$$

$$DP(i,j) = \begin{cases} 6 & \text{if } x \geq 6 \\ \text{Int}(x) & \text{if } 0 < x < 6 \\ 0 & \text{if } x \leq 0 \end{cases} \qquad (1)$$

where Int(x) is the integer part of x. Consequently, if MBR(i,j) = NBR(i) the packet (or cell) gets drop preference 4, if MBR(i,j) > 5.66 NBR(i) the packet gets the lowest drop preference (0), and if MBR(i,j) < 0.17 NBR(i) the packet gets the highest preference (6). Drop preference 7 is reserved for those connections that use a network service with guaranteed bandwidth and quality. The accepting and discarding of packets inside SIMA network is only based on the drop preferences.

The proper value for time constant of actual bit rate measurement depends on the buffer capacity reserved for the service class used by the connection. With real-time services the buffer should be small, and thus the measurement period must be short. When using a non-real-time service the user may want to send bursts of packets without high packet loss ratio. As a consequence the averaging period should be much longer.

4.2. Scheduling and buffering unit

The key issue in the implementation of the SIMA service in a high capacity core network is the packet discarding system before the actual buffering shown in Figure 4. At any instant there is an accepted level of drop preference (DP_a): if an incoming packet has the same or higher drop preference, it is accepted, otherwise it is discarded. The calculation of DP_a is based on the buffer occupancy levels of the real-time buffer (M_{rt}) and non-real-time buffer (M_{nrt}).
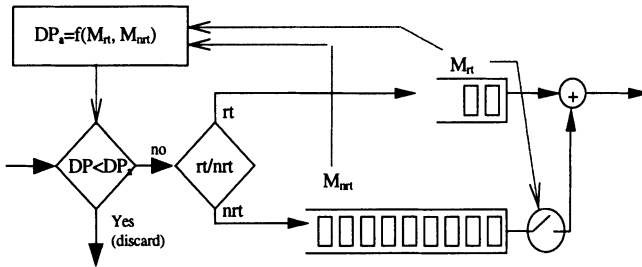

Figure 4 A packet scheduling and buffering unit (SBU).

All the packets which have been accepted in the scheduling unit are situated either in the real-time or non-real-time buffer. Both buffers may apply the First In First Out (FIFO) principle. In order to obtain a small delay and delay variation, the real-time buffer should be relatively small (e.g., 10 kbyte). All packets (or cells) in the real-time buffer shall be transmitted before any packet in the non-real-time buffer. It should be emphasized that the delay priority of real-time flows has no effect on the packet loss ratios. The non-real-time buffer should be much larger (e.g., 1 Mbytes) because of the packet scale fluctuations in typical non-real-time traffic processes. Moreover, large buffers make it possible to offer reasonable service for those flows that are capable of adjusting their bit rate.

It should be emphasized that the function of each scheduling and buffering unit (SBU) is independent of all other SBU's; all the tasks of SBU are performed based on the information of incoming packets, and moreover, all the necessary functions for the implementation are described in Figure 4. Thus, the management of the SIMA network is very straightforward.

5 CONCLUSIONS

The current specifications fail to adequately address the need of simple management and feasible charging for future Internet and other networks with high capacity and quality requirements. Accordingly, there is a need in the communications industry for a network management architecture that is simple in concept and in its implementation, yet adequately addresses the quality of service

requirements to support a variety of network services, including real-time and non-real-time services. There exists a further need for a system and methodology that provides for the implementation of a simple and effective charging capability that accounts for the use of network services. The present SIMA service introduced in this document is capable to fulfill these and other needs which remain unaddressed by current traffic management approaches.

The SIMA service is technically based on three key ideas: the use of nominal bit rate concept, the use of eight drop preference levels for every packet, and separation of real-time and non-real-time connections at the buffer level. If a user needs a connection over an IP or ATM network, he should select a nominal bit rate which could be even a constant proportional to a monthly fee. The user shall select also either a real-time or a non-real-time service class. In addition to these two parameters the user does not need to give any information about the properties of the connection like required bit rate or quality of service. After the connection establishment the capacity division among different connections is based on a drop preference which is determined using a ratio of the measured bit rate and the nominal bit rate. Drop preference in addition to the real-time/non-real-time separation is sufficient for every network node to properly manage the traffic.

Because there is no need for various traffic classes, traffic parameters and network services, the SIMA service makes possible a simple and efficient implementation of network nodes, a simple and fair charging scheme, and very simple traffic management in the core network. The SIMA concept is a very promising scheme for solving the most acute traffic control problems in Internet.

6   REFERENCES

ATM Forum (1996) Traffic Management Specifications v.4.0, af-tm-0056.000.

Cukier, K. (1997) ...Net tariffing lags behind, *Communications International*, 7 April, p. 4.

Firdmand, E. (1997) Rx for the Internet: Usage-Based Pricing, *Data Communications*, Jan., pp. 27-28.

Greenfield, D. (1997) Euro-ATM Get Ready for the Rollouts, *Data Communications*, Feb., pp. 48A-48G.

Heckart, C. (1995) ATM Services, The Truth, *Telephony*, Sept. 18.

IETF (1994) RFC1633 Integrated services in the Internet architecture: an overview.

ITU-T (1996) Recommendation I.371, Traffic control and congestion control in B-ISDN. International Telecommunication Union, 90 p.

McQuillan, J. (1996) Internet poised for great change, *Broadband Networking News*, Nov. 26, p. 7.

SIMA (1998), SIMA information is available at http://www-nrc.nokia.com/sima/.

Snyder, B. (1997) California crunch, Pacific Telesis documents woes of Internet traffic, *Telephony*, March 31, 1997, p. 6.