Feedback and Pricing in ATM networks

Liam Murphy[*] John Murphy[†]

Abstract

Admission control and congestion control can provide traffic guarantees in ATM networks. However some customers may not be able to describe their traffic accurately enough for the network to provide these guarantees. By sending a dynamic feedback signal about the current utilisation of network resources, the network could provide loss guarantees to adaptive users who respond appropriately. One possible feedback signal is a price based on the level of network load : when the load is high, the price is high, and vice–versa. We outline our proposed distributed iterative pricing algorithm, and then explore some arguments often raised against usage–sensitive pricing along with some counter–arguments. We also show through simulations that it is possible to simultaneously gain both network and economic efficiency by using pricing.

1 Introduction

Asynchronous Transfer Mode (ATM) has been adopted as the transfer mode for the Broadband Integrated Services Digital Network (BISDN) [1], a service–independent network capable of supporting all the communication services that users now require or may require in the future. ATM is also emerging as a local area networking technology, since it provides flexible bandwidth–on–demand and internetworking capabilities for conventional data communications. ATM networks are therefore expected to accommodate a wide range of users, including some whose applications require **guarantees** on cell loss and/or delay. These guarantees will vary widely in magnitude and firmness across the spectrum of application requirements and user preferences.

Admission control and congestion control can provide performance guarantees and are therefore two of the most important ATM network functions. In order to obtain these guarantees from the network, users have to describe their traffic inputs by specifying values for network–defined **traffic descriptors** such as peak cell rate (PCR) or sustainable cell rate (SCR). However some users may not be able to describe their traffic accurately, because their applications cannot be sufficiently well characterised by the given traffic descriptors or because their actual traffic inputs depend on factors outside user control (such as the number of active applications competing for access to a server). A common assumption in many proposed admission control schemes is that traffic which is not well–described cannot get specific guarantees beyond the level of service currently being provided to best–effort traffic.

The ATM Forum has recognised the problem of providing guarantees to users whose traffic cannot be well–described, and in response has developed a specification for Available Bit Rate (ABR) service [2]. Users who choose ABR service receive feedback from the network about

---

[*] Department of Computer Science and Engineering, Auburn University, AL 36849, USA; tel: +1 334 844-6326; fax: +1 334 844-6329; **lmurphy@eng.auburn.edu**

[†] School of Electronic Engineering, Dublin City University, Glasnevin, Dublin 9, Ireland; tel: +353 1 704-5444; fax: +353 1 704 5508; **murphyj@eeng.dcu.ie**

the current level of network resource utilisation, and can get cell loss guarantees[1] if they respond appropriately (by reducing their input rates in times of congestion, for example). Therefore ABR service is suitable for **adaptive** users, whose applications are flexible with respect to delay but not necessarily to loss. This flexible behaviour represents a tool that network operators can use to increase network utilisation while continuing to serve guaranteed traffic such as CBR and VBR applications. In addition, this type of network feedback would modify an adaptive user's traffic at the source rather than after it has been injected into the network. This would help to localise the effects of feedback to the edges of the network and allow simpler internal network operation.

Most suggestions for supporting ABR service assume that well–described traffic which requires performance guarantees gets priority in the use of network resources such as bandwidth or buffer space, and that the remaining resources are fairly shared among the ABR users. Two issues which are not explicitly addressed are

- why more "demanding" traffic should get priority over ABR traffic;
- what constitutes "fair" sharing. Should the available bandwidth be shared equally among all ABR users, for instance ? Or should it be shared according to the various application requirements (some applications are more flexible than others ...), or in some other way ?

It is important to note that, just because such issues are not addressed explicitly, does not mean that these proposals are neutral on what are often regarded as *policy issues*. On the contrary : sharing the available bandwidth equally among all ABR users implicitly values all such traffic equally, although the users themselves may put widely differing values on their service; giving CBR and VBR users priority over ABR users ignores the possibility that ABR users may be willing to pay more for network access than users with well–described traffic. We are not saying that these assumptions are wrong or undesirable, but instead we advocate making them explicit so that both users and network operators can make decisions based on as much information as possible.

Admission control and congestion control in ATM are difficult problems which so far have not been satisfactorily solved. Perhaps because of the technical challenges involved, most researchers have ignored two key issues (or relegated them to the "policy issues" layer) which are nevertheless crucial to the success of ABR implementations :

- **how should congestion be defined and measured ?** This is a difficult question because individual user requirements vary considerably, so that one user may think the network is congested while another does not; and because in internetworks the responsibility for detecting congestion may be distributed among several network operators, each of which applies a different test at their bottleneck points.
- **how should limited resources be allocated under congestion ?** Currently, randomisation with First-In/First-Out queueing is used, but some proposals call for users to indicate the relative priority of their traffic – leading to the problem of providing incentives so that all users will not choose the highest priority.

Our aim in this paper is to propose a feedback control scheme which explicitly addresses these issues.

2 Different types of efficiency

A network is as good, or as bad, as its users perceive it to be. This fundamental assertion leads to the conclusion that network performance should be measured in terms of overall user satisfaction with

---

[1] No specific delay guarantees can be provided, hence ABR users must be prepared to absorb delays at the traffic source before being allowed to input traffic into the network.

the service they receive. Network engineering measures (such as average packet delay or loss rate) are inadequate reflections of user satisfaction when user requirements are not roughly similar. Due to the difficulty in accounting for individual user's requirements, however, such aggregate network–oriented performance measures are usually used in design and operations problems. User objectives are averaged across all users and over time, and these averaged objectives are used to drive the network control process as shown in Figure 1. Therefore the design loop is not closed all the way to the users when making network operational decisions. We propose to bring the users *back into the loop* and thereby ensure that performance measures are user–oriented. Network performance measures will continue to be important to network owners and operators, but we believe that user preferences should be the primary consideration driving resource allocation and congestion control schemes.

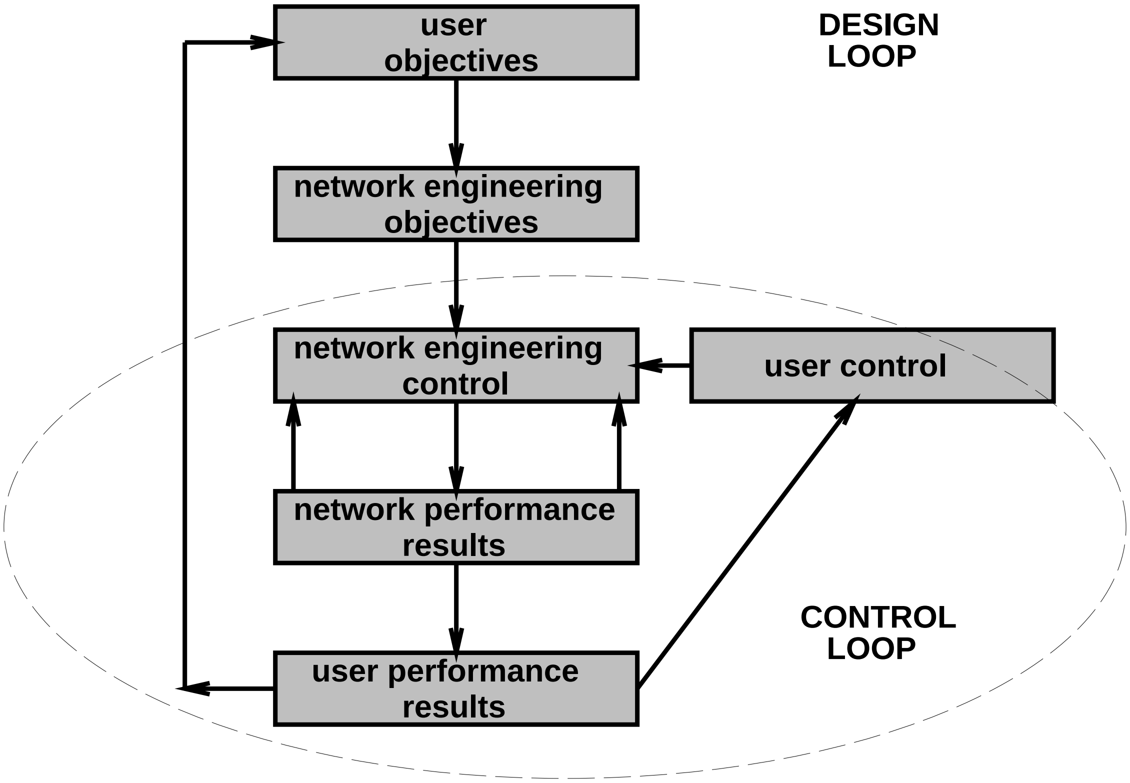

Figure 1: Network design and control loops

In focusing on user preferences, we need to distinguish two very different notions of efficiency :

- **Network efficiency** refers to the utilisation of network resources such as bandwidth and buffer space.
- **Economic efficiency** refers to the relative valuations the users attach to their network service.

If a network can maintain an acceptable level of service while minimising the resources necessary to provide this service, we say that its operation is (network) efficient. If users who value network service more are served ahead of users who value it less, then we say that operation is (economically) efficient.

An obvious question is, why will either type of efficiency continue to be important ? Some observers have suggested that the widespread deployment of fibre optic lines, and continuing exponential decreases in processor and memory costs, will result in these network resources becoming essentially "free" so that efficiency in their use will not be important in the future, and all users can always be accommodated. We do not believe these arguments apply in the short or medium terms, if indeed they will ever apply. User demands are increasing exponentially, so that it is not clear when – if ever – network resources will be "free"; experience suggests that application developers will have no difficulty in designing new services that use up all available resources, perhaps after an initial adjustment period; and market economics dictates that commercial network operators should be aware of the differing valuations that users attach to the same level of network performance. The same considerations apply even to privately owned or operated networks : the ultimate goal will continue to be to maximise some suitable measure of the value of using the network.

2.1 Improving efficiency with feedback

Adaptive users can help to increase network efficiency if they are given appropriate feedback signals. When the network load is high, the feedback should discourage adaptive users from inputting traffic; when the load is low, the feedback should encourage these users to send any traffic they have ready to transmit. In this way many of the congestion problems that can occur if the offered load is regarded as fixed can be avoided. One possible feedback signal is a price based on the level of network load : when the load is high, the price is high, and vice–versa. Similarly, by associating a cost measure with network loading, all users can be signalled with the prices necessary to recover the cost of the current network load. Price–sensitive users – those willing and able to respond to dynamic prices – increase economic efficiency by choosing whether or not to input traffic according to their individual willingness to pay the current price. Users who value network service more will choose to transmit, while those who value it less will wait for a lower price. When the network is lightly loaded then the price will be close to zero, and all users can input traffic.

One important point needs to be clarified before going on :

- when most people think of prices, they think in monetary terms, e.g. dollars and cents. However, **there is nothing inherently monetary in applying pricing principles to communication networks.** As long as the appropriate cost and valuation functions can be defined, a pricing mechanism can be applied even if money is not directly involved. For example, in a private network where one organisation controls all the users, the "prices" would be control signals which summarise the state of network resources. In this case the users (or their applications) are co-operative and can be programmed to obtain a desirable traffic mix.

We envisage that the charge to a user in an ATM network might have many components, such as a connection fee, a charge per unit time or per unit of bandwidth, premium charges for certain services, and so on. We suggest that there should also be a usage–sensitive component, to increase both network and economic efficiency. However we recognise that many people are concerned about the use of pricing in network operations. Concerns range from questions about the feasibility and overhead of usage–sensitive pricing, to more philosophical issues such as profit opportunities and fairness. While some of these concerns may or may not be borne out by future developments, others are based on misconceptions of what is being proposed or on other non–technical grounds. We do not expect that decisions on implementing usage–sensitive pricing would be made solely on technical grounds, but we do believe that a clear understanding of the nature of what is being proposed is necessary on all sides. Therefore we first outline our proposed dynamic pricing scheme, and then address some of the objections often raised in discussions of dynamic network pricing.

2.2 Distributed iterative pricing algorithm

It is important to note that our proposed pricing algorithm would only be applied to users who are able and willing to respond to dynamic prices. All other users would be charged according to another pricing scheme. How to co-ordinate the various pricing schemes to achieve some overall objective (such as fairness) is a complex issue and we do not address it in this paper.

The network and its users are considered to form an economy or economic system. The system has various resources such as link bandwidths and buffer spaces that can be used to meet user demands for service. Network constraints such as buffer sizes or link capacities are translated into cost functions on the demands for resources.

Each adaptive user is viewed as placing a benefit, or willingness–to–pay, on the resources they are allocated. Given a price per unit of bandwidth or buffer space, a user's benefit function completely determines that user's traffic input. A benefit function could follow the usual economic assumption of diminishing incremental benefit as more of the resource is consumed (Figure 2(a)). Or it could be a simple threshold rule, or series of threshold rules, for deciding how much of the resource to request based on the current price (Figure 2(b),(c)). Users are allowed to change their benefit functions every feedback interval so the examples in Figure 2 are for a particular interval.

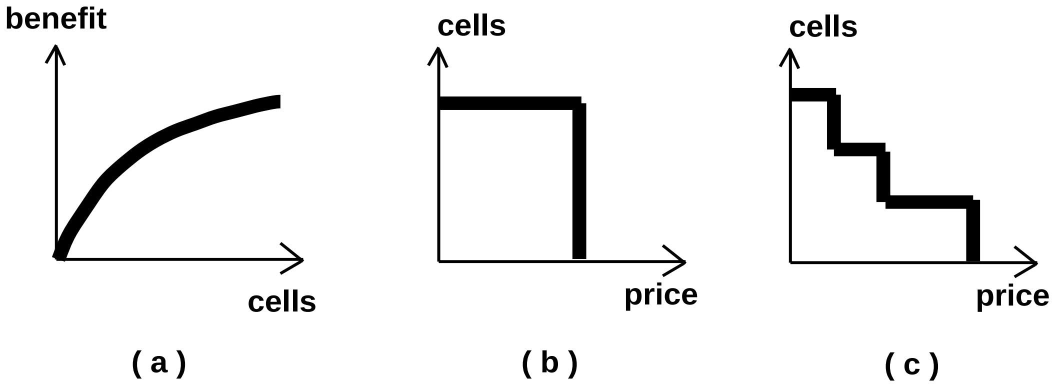

Figure 2: Possible user benefit functions

The network operator sets the prices so that the marginal benefit the users place on their resource allocation is equal to the marginal cost of handling the resulting traffic in the network[2]. The basic property of the cost function is that price should go to infinity as usage of the resource approaches capacity. The network operator dynamically adjusts the prices based on current network conditions. It turns out that it is not necessary for the network operator to know the user benefit functions; therefore this pricing scheme is suitable for both public and private networks.

A distributed iterative pricing algorithm has been developed [3] (see Figure 3). The distributed nature of the pricing algorithm suggests that it may be possible to meet the real–time feedback requirement. In addition, the computation required per iteration at each user and ATM access switch is simple, which suggests that inexpensive processing elements may be sufficient in executing the algorithm.

In [4] the pricing algorithm of Figure 3 was extended to include adjustment of VP capacities using a similar iterative scheme. The VP adjustments are not purely local, since contention between users is not confined to the network access points as before. On the other hand the VP adjustments have a longer time period in which to carry out their computations.

---

[2] These prices only address the variable costs corresponding to network constraints.

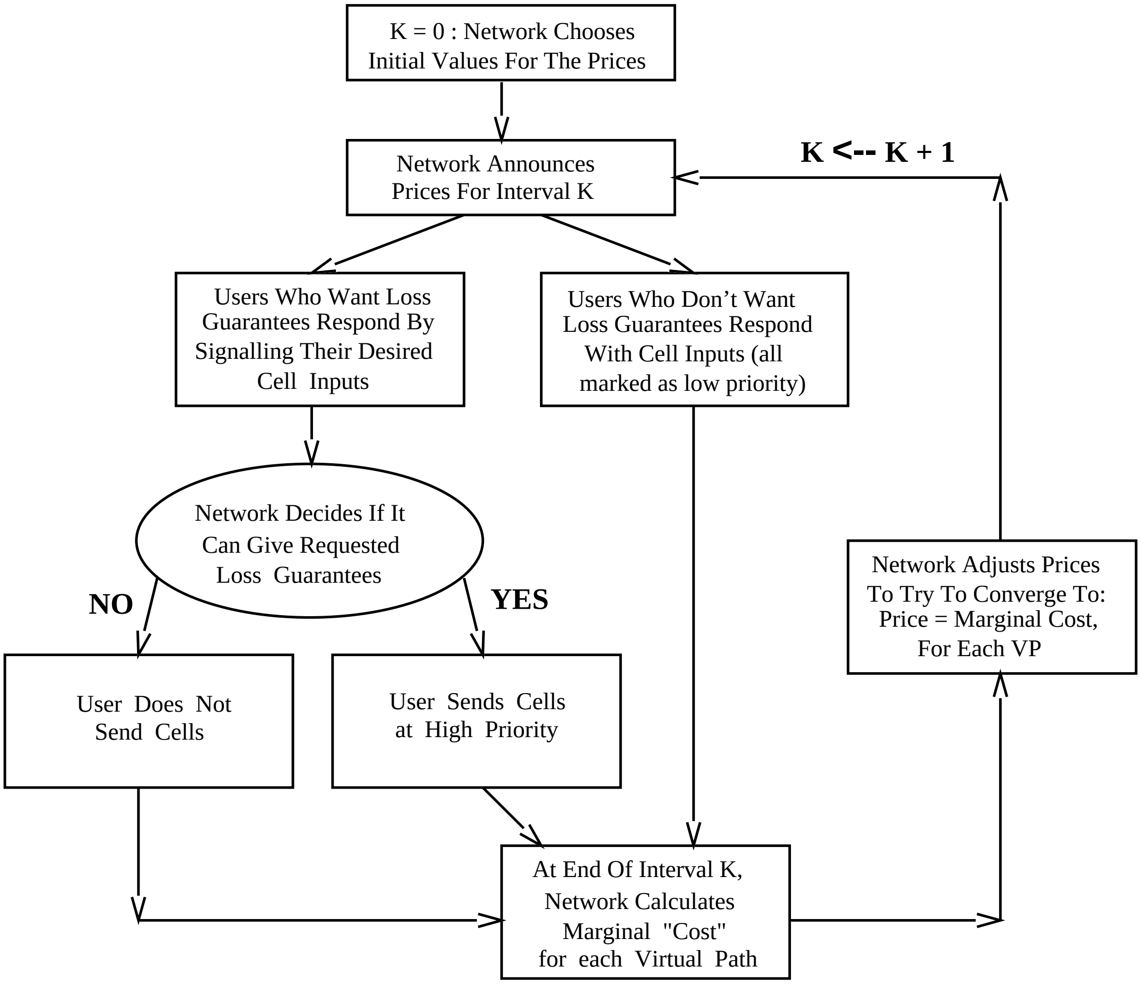

Figure 3: Distributed iterative pricing algorithm

2.3 Concerns about usage–sensitive pricing in networks

We explore some common arguments against usage–sensitive pricing in network operations in this Section, and provide some counter–arguments. Some previous work along these lines is contained in [5].

- once a network is installed, any load–dependent costs of transferring data are minimal – the fixed costs of network management and maintenance dominate. These fixed costs can be efficiently recovered through connection fees and capacity prices (proportional to the size of the access link). Why implement an elaborate pricing mechanism to recover the relatively small variable costs ?
  - *Counterpoint* : this ignores the congestion cost which one user's traffic imposes on other users sharing the resources. Bandwidth or buffer space occupied by one user's traffic is not available to other users. When this reduces other users' quality of service (through increased delays, loss rates, blocking probabilities, and so on), they suffer congestion costs which may translate into significant actual costs of service degradation. One mechanism to capture these costs is a price which is sensitive to some indicator of congestion, such as load.
- even if we want to recover congestion costs, how can the network determine what actual costs the current load is imposing on users who probably have widely varying service requirements ? Getting users to reveal these costs is likely to be extremely complicated, if not impossible.

  - *Counterpoint* : it is true that providing users with the right incentives to reveal their actual costs of service degradation is complicated. However, by giving users the opportunity to formulate their willingness–to–pay as a function of their usage, the network allows users the choice of trading off their usage costs against their service requirements. Users who choose not to take load–sensitive prices into account are in fact asserting that their usage is at least as valuable to them as the current price, and they should be required to verify that when other users are experiencing congestion.

- why won't some non–pricing scheme be enough ? Administrative controls can be used to impose some appropriate notion of fairness, for example; or users can choose a traffic priority level which matches their requirements.

  - *Counterpoint* : who decides what is fair ? The network operator can; but according to a user–oriented objective, fairness should be determined collectively by the users. We might all agree that telesurgery is more important than email, but what about interactive video games versus email ? Also, every time a new application is developed it has to be slotted into the priority order, an increasingly complex process. Suppose the network simply supports priority levels and allows each user to choose their own level. Why wouldn't they all choose the highest priority ? To guard against such abuses, there would have to be some penalty for "inappropriate" declarations, implying the need to define "appropriate" priority levels or to assign increasing charges to higher priorities. A user's choice of priority level would then be based on economic considerations : balancing the benefits of higher priority against the costs and/or the penalties for inflating their application's perceived priority level. Pricing represents the limiting case of a continuous spectrum of priorities, eliminating the quantisation errors that discrete levels introduce.

- most users will want to know their charges in advance, and will not want to deal with prices that change during the lifetime of a typical connection.

  - *Counterpoint* : we are not advocating that all users must face usage–sensitive prices. Any user can choose not to face dynamic prices, even if their application is adaptive. They would then be charged according to some other pricing scheme, which presumably should be co-ordinated with the dynamic pricing mechanism (although we do not address this issue here). Or a user faced with dynamic prices can always choose to ignore those prices by transmitting at their application's natural information rate, and paying the resulting charges. Finally, in our scheme – and in any realistic pricing scheme – it would be possible for a user to set the maximum charge they are willing to pay, which is what is usually required for budgetary purposes.

- bits/bytes/cells are not the correct units to charge for – it's information that users care about. Any scheme which proposes to look inside every packet or data unit to determine how it relates to other packets is likely to be too complex to be justified. Also, lower–layer mechanisms (such as Ethernet collisions) or cell losses requiring packet retransmissions make it difficult to predict how much "raw" data has to be transferred to transmit a given amount of information. Should users be charged for retransmissions that they have no control over, or cells which are dropped by the network ?

  - *Counterpoint* : our proposal involves pricing for transport, not for content. The "importance" of a particular cell or packet, and its relation to other cells or packets, is a higher–layer issue determined by the application (or ultimately by the users). We are not proposing that the network be aware of these issues; on the contrary, the network view in our scheme is that cells are cells and it's up to the users to decide how cells are used to transfer information. It's true that it is in general impossible to determine beforehand exactly how many cells are required to transmit a block of information, but again this is a higher–layer issue. The basic question is whether the users or the network should bear the uncertainty. If the network is expected to offer a "file transfer" service, the file transfer charge per megabyte could be computed by averaging over many such transfers. If the user is expected to pay for all transmitted cells, they could define a maximum number of cells they are willing to transmit per megabyte of information, and invoke an application–layer process if this threshold is exceeded.

- dynamic pricing schemes are unworkable in practice due to the overheads involved in accounting and billing for usage on such a detailed level. In addition, a significant portion of the revenue raised is needed to defray the cost of doing dynamic pricing in the first place.
  - *Counterpoint* : the costs of dynamic pricing may outweigh the benefits for a particular implementation but we do not believe this is necessarily true for all dynamic pricing schemes. In particular, online pricing mechanisms may reduce the actual cost to an acceptable level; there is no reason to think that current billing and accounting costs in other industries, such as telephone or electricity networks, will necessarily apply to dynamic pricing in ATM networks. This concern can only be answered for each scheme individually, and should obviously be part of the overall decision on what usage–sensitive pricing scheme to implement, if any.
- usage–sensitive pricing is just another way for network operators to make more money. Users will lose out as network operators maximise their profits.
  - *Counterpoint* : it's true that there is the potential for profiteering anytime prices are charged, especially when the conditions under which prices are set are not immediately accessible to ordinary users. But in a competitive environment, network operators have market incentives to keep their margins of revenue over actual cost as low as possible. This incentive is missing in the case of a monopoly provider or a cartel of price–fixing providers. But whether abuse is possible in this case depends on policy and regulatory decisions rather than on the specific pricing scheme. And usage–based prices eliminate the cross–subsidisation of heavy users by lighter users inherent in flat–fee or connection prices.
- charging for cells transmitted fails to capture cases where the benefit of a transfer is with the receiver. If senders are charged for receiver–initiated transfers, we could see a drastic reduction in the number of open–access servers with a corresponding decrease in the value of using the network.
  - *Counterpoint* : the problem of allocating the benefits of a particular information transfer is a higher–layer issue. We do not believe that associating the charge for a transmission with the sender constrains the actual flow of money in any way. It is easy to imagine multiparty connection protocols which initially negotiate each party's responsibility for the total charge, or "reverse–charges" servers which only transmit data once the receiver has indicated willingness to pay the resulting transmission costs.
- dynamic pricing is impractical because users cannot respond to prices which are updated many times per second. If the update interval is increased to the minimum period in which users can respond, congestion can arise and disperse in between price updates, so that prices no longer influence user behaviour.
  - *Counterpoint* : our scheme assumes an intelligent network interface at price–sensitive user sites, so the processing necessary to respond to dynamic prices would be done automatically based on pre-programmed user preferences. Current ATM connection admission control schemes already assume enough user intelligence to be able to negotiate quality of service parameters, so our scheme adds a little more complexity rather than a new

requirement. This software would play a similar role to current TCP implementations, which respond to network feedback by adjusting their traffic inputs, except that the feedback in our case is the current price.

- economics is important in network planning but has nothing to do with the technical operation of a network, whether public or private.
  - *Counterpoint*: economics has a lot to do with network operation. Packet–switching was developed for computer communications because around 1970 it became more economical to use switching and routing to statistically multiplex several connections into one transmission link, rather than dedicating one circuit to each connection as in circuit–switching. Economics plays a role in formulating and solving decision problems in all types of network; current price schemes differ from ours only in the frequency with which prices are updated. Our proposal simply moves this updating into "real–time" for those users who are able and willing to respond on that time scale.

3 Simulation models & results

The simulation set up is a high speed ATM 155 Mbps link with two types of users connected, inelastic and elastic [6] (see below). Video sources are modelled as inelastic users; data sources are modelled as elastic users. The link model is shown in Figure 4.

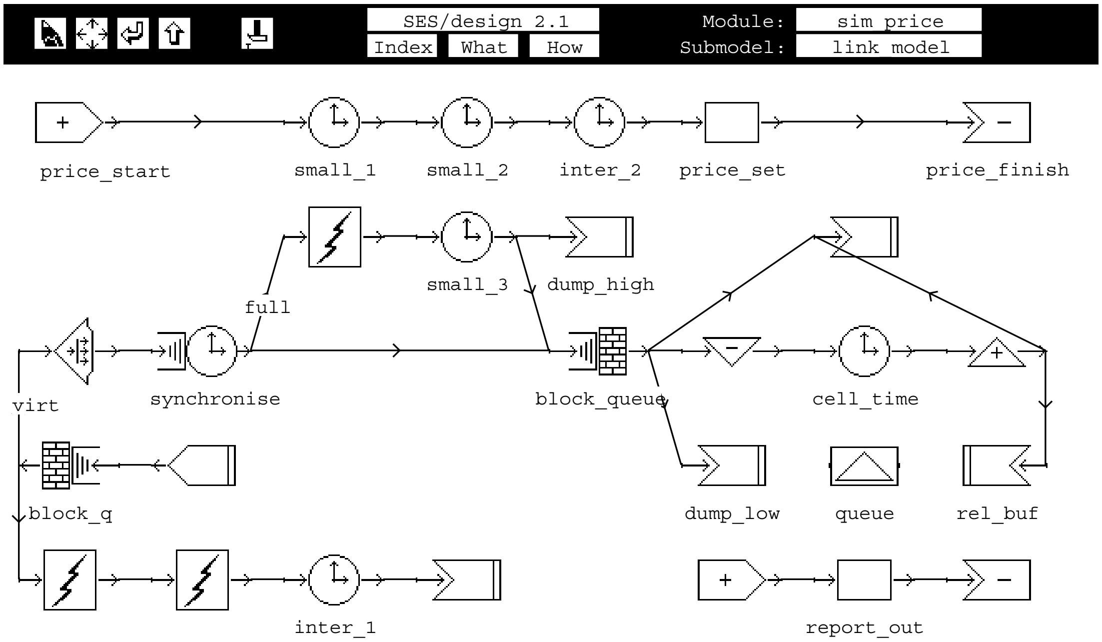

Figure 4: Simulation model for economic efficiency

In our proposed scheme a price is generated by the network based on the present state of the network buffer, and the sources adapt their demands based on this. What we propose and simulate adheres to the UNI 3.0 specification from the ATM Forum [7]. A leaky bucket is also implemented on top of our scheme so that if there is cell loss we can discard the marked ones first.

The model takes in cells over a pricing interval and gives a price to all the sources sharing the link. The price reflects the congestion (if any) in the buffer and hence on the virtual path. The pricing interval is short compared to the video frame time : a value of about 0.05 of a frame time was chosen. To achieve feasible run times we neglect cell scale effects. This neglecting of cell scale effects is critical to the speed up of the simulations. The total utilisation of the link is then high, at a value of around 0.85. We can compare the performance of the pricing scheme to other access control schemes by looking at the loss and at the value of the traffic carried.

3.1 Inelastic users model

This user has a delay bound on the traffic, but can tolerate only sending a fraction of the cells that are ready to go in the interval in question. We assume that if they are not sent in the interval then there are discarded and useless to the user. An example would be the second layer cells of a video codec, which only enhance the picture, but are useless if they arrive after the frame has been shown. We assume that the user has a concave benefit function, $benefit(X)$, on the number of cells sent, X. This is fixed for each interval but could vary from one interval to another. If the user sends in interval t the number of cells X_t then the benefit to the user is $benefit_t(X_t)$. How the user decides on the number of cells to send in interval t depends on the price given to the user P.

The video source model that we use here is a standard one for video conferencing [8]. The codec has a compressed bit rate of 2.3 Mbps, which adheres to the H.261 standard for video [9]. The video sources each have a mean of 2.3 Mbps and a peak of 5 Mbps. These are input at a rate of 30 frames per second, all synchronised together, so that the resulting inelastic users are (more or less) stationary on the millisecond scale, that is with respect to the pricing interval.

3.2 Elastic users model

There is one essential element to preferences for elastic traffic which is that the traffic has value even if it experiences delay. However the value should be declining the longer the delay is (if it isn't, then the decision rule to send is simple : never send unless the price is zero – or the lower bound, if not zero). Another characteristic that is not necessarily intrinsic to all elastic traffic, but seems to be characteristic of most types (e.g., files, messages), is that a transfer has zero value unless the entire message (file) is delivered.

The elastic users can be thought of as one user with a lot of files to transfer independently, many users each with one file, or some combination of these types. The negotiations for file transfer or connection set up will only occur when a new video frame is to be sent, i.e. every 1/30 second, because this simplifies the simulation and makes it possible to speed up the run time. This makes the modelling for high speed easier and is not restrictive. Therefore the network re-negotiates the PCR every 1/30 second with the elastic users. A data source is one of the most difficult sources to model as the source type depends on what applications are being run and on what systems. We model a file transfer application. This captures the bursty nature of data communications as well as its looser delay requirements relative to voice and video. A model was built based on transferring files from one computer to another. An empirical distribution for file size ranges was obtained from actual files stored on one of our computers. In the simulations a range was chosen according to this empirical distribution, and then a file size was chosen from a uniform distribution within this range. The amount of data transfer can be varied depending on how many files are to be transferred. The peak–to–mean ratio of this source can be high with values up around 1000. There are on average 20 data sources in use and these are taken from a uniform distribution between 1 and 39 sources. Each file to be sent is also taken from a uniform distribution between 20 and 660 cells. Therefore the average bit rate of a single data source is about 4.3 Mbps.

3.3 Simulation details

The network and source models were simulated using SES/*workbench* [10], a discrete-event simulator that allows hardware and software simulation. The models were mainly created by use of its graphical user interface. SES/*workbench* compiles the graphical code to C and creates an executable. The simulation execution platform was a cluster of Sparc-10 workstations.

The simulation model is made up of submodules, each of which performs a well defined function. The sources generate cells which are input to a network interface submodule. The network interface takes the source bit stream and forms ATM cells. The cell stream from an interface is then input to the ATM switch buffer submodule. This submodule smooths the arrival of cells to the ATM network and so takes care of cell scale congestion. The switch buffer is the limited network resource in our model. In a typical implementation the buffer is managed by an input control mechanism, such as a leaky bucket scheme.

The leaky bucket is a method for policing the mean rate of a source and still allowing bursts to occur. The mean rate is the rate at which tokens are given to allow cells past the network interface and into the network. If there were no storage of tokens then this would be a pure mean rate policing function. If the source can store tokens, it could burst all those tokens in one go. If a source has no remaining token allowance, then rather than dropping that cell usually it is merely marked. If there is congestion in the network and cells have to be discarded, the marked ones are discarded first. The leaky bucket does not lower the cell loss in a buffer; it merely selects which cells to lose. Our pricing scheme actually lowers the probability of cell loss by smoothing the traffic adaptively based on the buffer occupancy.

The simulation was speeded up by use of time stepping, where the time step was chosen to be the frame time. This was possible as everything was known at the start of the frame time, except the pricing, but if the price was low and the link un–congested then the pricing could be skipped.

3.4 Results

The results that are got are shown in Table 1 and show the difference between using adaptive pricing and no pricing at all. What can be seen is that both the network efficiency and the economic efficiency can increase by using pricing at the same time. Here there was an increase of 15% in the economic benefit to the users as well as a decrease of 70% in the amount of cells lost.

Table 1: Results of both network and economic efficiency

Source Type		% Loss	User Value	User Charge
Unpriced	Inelastic	0	240	
	Elastic	5.5	146	
	Combined	**2**	**386**	**14367**
Priced	Inelastic	1.6	239	
	Elastic	0	204	
	Combined	**0.6**	**443**	**6**

4 Conclusions

We have presented an economic framework for adaptive users in ATM networks. Instead of the typical requirement for traffic descriptors in order to get performance guarantees, these flexible users can get loss guarantees if they adjust their traffic input rates in response to dynamic feedback from the network. This is the basis for recent proposals for ABR service in ATM. Our framework takes these proposals one step further by explicitly defining how that feedback is generated by the network, and what form it takes. In our scheme, the network announces a price which is based on the cost of using network resources, and price–sensitive users adjust their cell inputs based on this price and their own specification of how valuable network service is to them.

What we propose is to use pricing to make users behave in a manner similar to typical users in today's data networks. In these networks there is an understanding that the sources will cut back when instructed by the network. However in a commercial network this may not happen if there is no incentive for the users to do so. We also address the problem of user valuation of the service, and allow for ABR sources to have more demanding traffic than well–described sources. We have proposed a distributed iterative pricing algorithm and shown (by simulation) that it is possible to gain both network efficiency and economic efficiency by using pricing. In other words, the network actually carries more traffic and carry more important traffic from the users point of view.

Acknowledgements

We would like to thank Jeff MacKie–Mason for many useful discussions about this work, and the contributors to the *com–priv* Internet mailing list for their objections to usage–sensitive pricing.

References

[1] M. de Prycker, *Asynchronous Transfer Mode : Solution for Broadband ISDN*, 2nd Ed., Ellis Horwood, 1993.

[2] P. Newman, 'Traffic Management for ATM Local Area Networks', IEEE Communications Magazine, pp 44–50, August 1994.

[3] J. Murphy and L. Murphy, 'Bandwidth Allocation By Pricing In ATM Networks', *Proc. of IFIP Broadband Communications '94*, Paris, France, March 1994.

[4] J. Murphy, L. Murphy and E. C. Posner, 'Distributed Pricing For Embedded ATM Networks', *International Teletraffic Congress ITC–14*, Antibes, France, June 1994.

[5] J. MacKie–Mason and H. Varian, 'Some FAQs about Usage–Based Pricing', available from **URL ftp://gopher.econ.lsa.umich.edu/pub/Papers/useFAQs.html**

[6] L. Murphy and J. Murphy, 'Pricing for ATM Network Efficiency', *Proc. 3rd International Conference on Telecommunication Systems Modelling and Analysis, Nashville, TN, March 1995.*

[7] *ATM Forum*, 'ATM User–Network Interface Specification', Prentice Hall, 1993.

[8] 'Redundancy Reduction Techniques for Coding of Video Signals in Multi–Media Services', COST 211ter, Compendium ATM, Coding Procedures, Simulation Subgroup, pp 1–83, May 1993.

[9] J. Murphy and J. Teahan, 'Video Source Models for ATM Networks', *Eleventh UK IEE Teletraffic Symp.*, Cambridge, UK, 23-25 March 1994.

[10] SES, *SES/workbench Reference Manual*, Release 2.1, Feb. 1992.