Pricing in ATM Network with Feedback

Hyoun Jong Kim
Telecommunications Management Section, ETRI
Kajong-dong 161, Yusong-gu, Taejon

Abstract:

In most of the recent research literature, network performance is expressed in terms of network engineering measures such as delay or loss. These performance measures are important to network owners and operators, but it is believed that user preferences should be the primary consideration which drives the resource allocation scheme.

A network is only as valuable as its users perceive it to be. Therefore, it is advocated that the users themselves determine relative traffic priorities. This paper describes the role of feedback in network resource allocation, which could be part of a user-oriented framework for network operation and control. Feedback mechanism can also be used to improve the two types of efficiency in the network; network efficiency and economic efficiency.

1. Introduction

Admission control and bandwidth allocation are important issues in telecommunications networks, especially when there are random fluctuating demands for service and variations in the service rates. In the emerging broadband communications environment these services are likely to be offered via an Asynchronous Transfer Mode (ATM) network. In order to make ATM future safe, methods for controlling the network should not be based on the characteristics of present services.

There are many problems with traditional approaches to the connection admission control area. Users may not know enough about the statistics of their connections to provide accurate traffic descriptors to the network.

This paper describes the role of feedback in network resource allocation, which could be part of a user-oriented framework for network operation and control. Feedback mechanism can also be used to improve the two types of efficiency in the network; network efficiency and economic efficiency.

A communications network is as good, or as bad, as its users perceive it to be. Network performance should therefore be measured in terms of overall user satisfaction, which in commercial networks is usually determined by the amount of money the users are willing to pay for network services. In non-commercial networks it is also possible to associate some measure of satisfaction with a user's service level, although it can be difficult to quantify if the users have varying preferences.

Most researchers divide users into classes according to their application requirements and traffic characteristics; for example, real-time video, real-time audio, stored (off-line) video, and off-line data transfer. Each class is then regarded as having a generic user for analytical purposes. Users are usually assumed to be capable of negotiating a contract at connection setup, which specifies their traffic parameters and service level. The network then polices the offered traffic and enforces the contract if users misbehave.

Several literatures, [3][4][7][9], which studied similar problem in this paper, are considered to usually ignore two obvious user characteristics.

- Even within a class, user preferences can be expected to be heterogeneous. If this internal variation within a class is substantial, then any assumptions about generic user behavior will be inaccurate for many users in this class.
- Some users may be capable of responding to control signals from the network during their connections, not just at connection setup. Users are sometimes thought as a part of the network and can regulate how much and when traffic is input, given appropriate indicators of the network status.

2. Resource Allocation of ATM Networks

There are many resources in ATM networks and the ones that are of interest here are the buffer spaces and bandwidths that are available in the network. The resources issue at the connection level is how to achieve reasonable network efficiency for source types, where traditional resource allocation techniques may not apply. Reactive control and feedback schemes which are under investigation may be useful in certain network environments or on longer time scales, e. g. connection level rather than burst or cell level. It is fair to say that how bandwidth will be allocated in ATM networks is still an open question. There is also some work on the cell level resource allocation in terms of the overhead and efficiency of the throughput of the cells.

2. 1. ATM and Resource Allocation

ATM has been adopted as the transfer mode for the Broadband Integrated Service Digital Network (BISDN), a service-independent network capable of supporting all existing and future communication services. ATM-based networks are therefore expected to accommodate a wide range of users, including some whose applications require guarantees on cell loss and/or delay. The ATM Forum has recognized 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.

ATM is also designed to enjoy some of the efficiency benefits of statistical multiplexing, while providing the resources necessary to support an integrated services network. The performance of the network depends on the extent to which resources are effectively shared among the applications running on the network. At the same time, increased sharing risks congestion, which degrades service quality. One of the great challenges for ATM is to design mechanisms that balance the competing objectives of resource sharing and congestion avoidance. It is proposed that the efficiency of the network can be improved by more closely involving the users in session control.

The current resource allocation methods used for services in ATM rely on many techniques. At present the constant bit rate (CBR) traffic and circuit emulation traffic is decoupled from the remaining services. This is achieved by using different virtual paths (VPs) for each of them. This allows each sub-problem to be tackled independently. Effectively the bandwidth for the CBR traffic is split from the other traffic's bandwidth, by reserving the amount needed in advance. The other traffic can then have no effect on the CBR traffic. However not all the advantages of statistical multiplexing can be got by doing this. For variable bit rate (VBR) services like compressed voice there are a number of models using effective bandwidths that look to be useful and provide good bounds for the connection admission control (CAC). The CBR and VBR services will be policed by either a single leaky bucket or possibly a dual leaky bucket.

ABR traffic is more difficult to control due to the unknown nature of the sources and there have been a number of proposed schemes to control it. The factors that have influenced the decisions on the choice of schemes are numerous and vary from what types of sources are expected to what might be possible in real networks. The final decision might not even have been arrived at, and this area is still under intense research.

An early scheme proposed, is to just allocate the resources for the peak rate and resign the network to inefficiency. However this is probably only a short term solution as competitive forces will force this option out. There is also the possibility of offering different levels of priority that the user can allocate to different services that they might have multiplexed together. While this might work when the user is in control of the multiplexing, in bigger networks this might not be possible.

2. 2. Importance of Resource Allocation

The system made up of the users together with the network has various resources that can be used to meet service demands. However in all realistic systems these resources are limited and some method of allocating them is needed when total demand is greater than the resource limit. The resources are the capacities of the ATM connections. The bandwidth allocated to a user is considered to be a commodity which is sold by the network to the user. The view here is that the users place a benefit, or willingness-to-pay, on the bandwidth they are allocated. Given a price per unit of bandwidth, a user's benefit function completely determines that user's traffic input. Users are assumed to act in their own best interests and to be capable of responding to changes in the price for bandwidth.

Will the network have plentiful resources in the future? Some commentators have suggested that the widespread deployment of fiber 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. However three points should be kept in mind.

- Demands are continuing to increase exponentially, so that it is not clear when - if ever - network resources will be free.
- Past experience suggests that application developers will have no difficulty in designing new services that use up all the available resources, perhaps after an initial adjustment period.
- When a significant number of users become involved in defining their service characteristics, efficient network operation will be critical in a competitive network provider environment. Put simply, if network operation is not efficient, the users will be efficient by multiplexing their traffic before submitting it to the network, for example. Legal barriers or tariff disincentives to this kind of user behaviour may not be feasible, so network inefficiency could lead to a financial penalty for the network operator.

3. Pricing for ATM Network Efficiency

3. 1. Pricing and Feedback

There are two distinct scenarios that need to be considered when deciding on an appropriate feedback signal: private and public networks. In a private network the users are applications owned and controlled by one organization. Therefore the users are cooperative, since their responses to feedback can be programmed to obtain a desirable traffic mix. In a public network the users must be considered as separate entities, with their own private rules for deciding their traffic inputs. The network cannot assume the users will be cooperative without being given the right incentives.

These incentives would correspond to actual money in a public network as this would be a good method of getting the users to adapt their behaviour. However in a private network the prices could correspond either to actual money or to control giving incentives to the users signals. These control signals would summarise the state of network resources such as bandwidth or buffer space.

The price that is being discussed can be thought of as a component of the total charge that the user faces from the connection. The total connection charge may be made up of many parts, like a connection charge, a time charge, a cell charge and then maybe a congestion charge.

Two fundamental points in this study need to be spelled out.

- Price is one possible feedback signal which has some attractive properties (compactness, quantifiable, etc.). Economists have developed a large body of theory of pricing mechanisms, and there is a lot of experience with the use of prices in real-world markets. However we do not rule out the possibility that there are other feedback mechanisms that, for one reason or another, may be preferable in communication networks.
- When most people think of prices, they think in monetary terms. 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, as in a private network where one organization controls all the users. In the private network the users (or their applications) are cooperative and can be programmed to obtain a desirable traffic mix. Of course some incentives are needed to ensure that the users respond as if the prices were going to be applied to their usage, but these incentives could be based on aggregated usage or some other indirect measure.

A number of proposals have been made concerning the use of feedback in ATM networks. A crucial issue ignored in most feedback proposals is the basis on which decisions should be made, both by the network and by users. The scheme that is proposed addresses this issue by developing an economic framework in which incentives are provided to enable rational decisions and resource allocations.

3. 2. Feedback and Efficiency

In focusing on user preferences, it is needed to distinguish two different types of efficiency.

- *Network efficiency* refers to the utilization 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 over time while minimizing 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.

Feedback signals from the network can help to increase both types of efficiency. These increases in efficiency (of either type) require flexible or price-sensitive users (from now on, called adaptive users) to be more closely involved in controlling their connections than is currently the case. Not all users are capable of, or want to be involved in, dynamically adjusting their traffic inputs.

A user is adaptive if they are able and willing to respond to feedback signals from the network during their connection lifetime by changing their offered traffic. The addition of willingness-to-respond reflects the fact that some users may run applications which are capable of adaption, but do not want to be concerned with network feedback during the connection. A typical adaptive scheme is where there is a limited set of traffic parameters given to the network and the network gives a limited set of QOS to the user. However these parameters would be instructed in advance by the user on how to adapt to the network feedback signals. Not all users will be adaptive and therefore the network will contain a mixture of user types.

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 injecting cells; when the load is low, the feedback should encourage these users to send any cells they have ready to transmit.

4. Concluding Remarks

It is shown that it is possible to maximize, using price as a feedback signals, both network and economic efficiency at the same time. However it is felt that it is still possible to use economics in networks even if it is not part of a pricing scheme in other area of network management. The feasibility of using this pricing approach here as a dynamic adaptive priority scheme in this way is one avenue for further research.

Whether the prices used in the scheme here could be used as part of an ATM network tariff is not clear. However it is believed here that the mechnism suggested is an interesting one and may be helpful in solving these difficult network management problems.

There are many unresolved questions about applying pricing principles to networks as follows.

- What network resources can be priced?
- Can the pricing scheme be made understandable to the users?
- How much overhead is associated with a particular pricing scheme ?
- Is there a possibility of unfairness and if so how can this be resolved ?
- How will dynamic pricing affect the development of new applications ?

It is believed that pricing - and feedback in general - may be useful in resource allocation and congestion control for integrated-services networks. At the very least, such mechanisms deserve to be investigated : as an alternative to more traditional allocation and control mechanisms, or in tandem with them.

References

1. ATM Forum, `ATM User Network Interface Specification', Prentice Hall, 1993.

2. C. Parris, S. Keshav and D. Ferrari, `A Framework for the Study of Pricing in Integrated Networks', Technical Report TR-92-016, International Computer Science Institute, Berkeley, CA, March 1992.

3. F. Bonomi and K. Fendick, `The Rate-Based Flow Control Framework for the ABR ATM Service', IEEE Network, Vol. 9, No. 2, March 1995.

4. H. Kanakia, P. P. Mishra and A. Reibman, `An Adaptive Congestion Control Scheme for Real-Time Packet Video Transport', Computer Communication Review, Vol. 23, No. 4, pp 20-31, Oct. 1993.

5. ITU-T I.311, `B-ISDN General Network Aspects', ITU 1993.

6. ITU-T I.371, `Traffic Control and Congestion Control in B-ISDN', ITU 1993.

7. K. K. Ramakrishnan and P. Newman, `Integration of Rate and Credit Schemes for ATM Flow Control', IEEE Network, Vol. 9, No. 2, March 1995.

8. R. Händel, M. N. Huber, and Stefan Schröder, ATM Networks: Concepts, Protocols, Applications, 2nd Ed., Addison-Wesley, 1994.

9. S. Yazid and H. T. Mouftah, `Congestion Control Methods for BISDN', IEEE Commun. Mag., Vol. 30, No. 7, pp 42-47, July 1992.