ROUTING IN ATM NETWORKS

Matti Palo

HELSINKI UNIVERSITY OF TECHNOLOGY
Laboratory of Telecommunications Technology

Abstract

Various routing methodologies are implemented both in connection-oriented and connectionless networks. From the perspective of ATM networks especially those that address issues of QoS and dynamic networks are important ones. ATM was originally developed as a switching standard for support of broadband integrated services in public, switched telephone network. This makes it natural to consider routing schemes used in telephone networks also in ATM networks. However, the proper choice of the routing algorithm is difficult, because the performance of the algorithm depends on the implementation of the nonstandardized ATM technology, like size of the buffers and CAC.

Contents

Abbreviations

ATM	Asynchronous Transfer Mode
B-ISDN	Broadband ISDN
CAC	Connection Admission Control
ISDN	Integrated Services Digital Network
PNNI	Private Network-to-Network Interface
QoS	Quality of Service
UNI	User-to-Network Interface

1. Introduction

Asynchronous Transfer Mode (ATM) is the transmission and switching technique, which has been chosen to be the transfer mode for implementing Broadband Integrated Services Digital Network (B-ISDN) [1]. Because the B-ISDN will support a wide range of audio, video and data applications in the same network, ATM has to be able to handle multiple classes of traffic with different traffic characteristics and Quality of Service (QoS) requirements.

ATM is a connection oriented networking technology. It means that a virtual circuit is set up across the ATM network prior any data is transferred. Because the path selected for the new connection will remain in use for a potentially long period of time, the consequences of inefficient routing decision will affect a connection for as long as that connection remains active. Thus it is critical that the path is selected carefully.

ATM call establishment consists of two operations [2]: the selection of a path and the setup of the connection at each node along the selected path.

1.1. Connection setup

A source ATM end-system that desires to set up a new connection sends a connection request into the ATM network across its user-to-network (UNI) interface. The request includes the destination ATM address, traffic parameters, QoS requirements and other essential information for the network to be able to find a path from source to destination. A request is propagated through the network, setting up the connection as it goes, until it reaches the destination end-system.

Despite the place where the route decisions are made, the processing of the call setup at each node along the path confirms that the requested resources are available [2].

1.2. Path selection

There are two basic routing techniques that can be used to determine the path: source routing and hop-by-hop routing. In source routing the originating system select the path to the destination and other systems on the path obey the source's routing instructions. In hop-by-hop routing, each system independently select the next hop for that path, which results in progress towards the destination provided that the decisions made at each hop are sufficiently consistent.[2]

In connection-oriented networks, such as ATM, both techniques can be applied. However, hop-by-hop routing does have some disadvantages in this type of network. The first is the creation of routing loops. The possible causes for the loops are inconsistency in routing decisions when different switches use different routing algorithms or when routing databases are inconsistent.

Inconsistency in routing decisions leads to a fundamental constrains of hop-by-hop routing, which is that the path selection must be fully specified and all systems must implement it exactly as specified. It is also possible that, even there is no loops, the individual hop by hop decisions made are far from optimal decisions.[2]

In source routing, the source is responsible for selecting the path to the destination. The decision is based on the local knowledge of the topology and since only one database is involved, loops are not possible. Furthermore the algorithms used need not to be the same in every system.

The routing technique used in telephone networks uses hop-by-hop routing. Also the B-ISDN routing, as an extension to Integrated Services Digital Network (ISDN) is based on the same method [3]. On the other hand, PNNI-routing has adopted the source routing technology [2].

2. ATM Network Model

2.1. Virtual Path networks

ATM standards specify two types of connections - Virtual Path (VP) connections and Virtual Channel (VC) connections. While VCs are used as the virtual circuits on which data is transferred, VPs are used to bundle several VCs together, thus decreasing the amount of entities to be managed. That's also why VPs play an important role in traffic control and resource management in ATM networks.

A VP is defined as a logical direct link between two nodes (Figure 1) in the network that are connected via two or more sequential physical links [4]. A VP is identified with the Virtual Path Identifier (VPI) and each VP has its own bandwidth, limiting the number of VCs that it can accommodate. The number of VPs in each link is not limited. The only limit is that the sum of bandwidth of the VPs, does not exceed the capacity of the link.

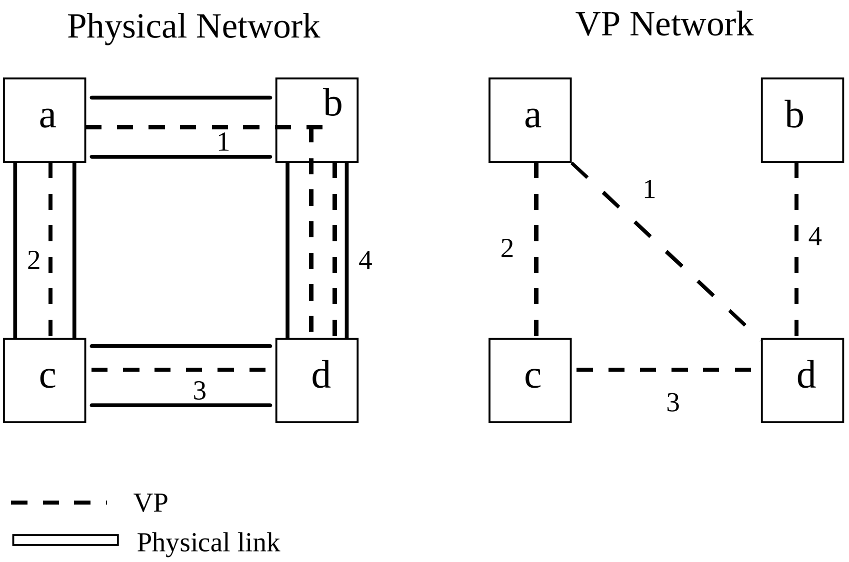

Figure 1: ATM Network Model

From the point of view routing, ATM network may be viewed as a collection of VP routes and each VC connection may be viewed as a route composed of a concatenation of VPs. When establishing a VC between two users, a route consisting of a set of VPs is selected. If a route consists of one VP, it is a direct route, otherwise it is an alternative route. The number of VPs used by a VC connection (VP hop count) directly affect the efficiency of the setup. Thus it is important to keep the number of VP hops on every VC as low as possible [6].

The collection of VPs defines a VP network with the switches as nodes and an edge between two nodes if and only if at least on VP lies between the corresponding switches. A VP networks can be used to segragate different traffic types, which require different QoS, to different logical networks.

2.2. Network classification

From the point of view routing algorithms, ATM networks can be classified according to whether ATM network supports homogenous or heterogeneous sources and whether statistical multiplexing is performed in the interior of the network or not [4].

If at least one VP supports VCs that have different traffic characteristics and QoS requirement, then the sources are said to be heterogeneous. On the other hand if the various traffic types are segrageted into classes so that VCs in a class have the same traffic characteristics and QoS requirements, then the sources employing a VP network are homogeneous. [4]

An ATM networks uses the statistical multiplexing if it is allowed that a connection exceeds the bandwidth allocated for the connection and benefits other connections' free capacity. On the other hand, if bandwidth sharing between connections is not permitted the network is said to operate in nonstatistical multiplexing mode. Statistical multiplexing is one of the main features in ATM networks.

This combination of the two properties presented leads to four classes of ATM networks [4].

1) Homogenous traffic with nonstatistical multiplexing
2) Homogenous traffic with statistical multiplexing
3) Heterogenous traffic with statistical multiplexing
4) Heterogenous traffic with nonstatistical multiplexing

The class 1 is not considered in different studies because the class is equivalent to single-rate circuit switching and therefore the routing algorithm currently in use in the telephone networks can be employed. Also the class 3 is usually omitted, because it is found that multiplexing of sources with very different traffic characteristics is not an efficient method.

3. Dynamic Routing

3.1. General

Dynamic routing is an efficient method of traffic control in which call routing is frequently altered according to the status of the network and/or anticipated demand shifts, so that the network can respond quickly and properly to the changes in traffic and/or facility conditions.

Compared to static routing dynamic routing gives a better change of success to an individual call by increasing the number of ways the call can traverse the network. When a new call requests a connection between a pair of switches, it is possible that the call is established along the direct route or along an alternative allowable route. Usually routing schemes attempt first the direct route and if it is unavailable then the alternative routes are considered. The routing algorithms differ mainly in how they choose the one route from the set of allowable routes.

Usually two-link alternative routes are considered. Removing the restriction of two VPs, allows a wider choice of alternative routes and thus tends to reduce the blocking probability. On the other hand, it also tends to reduce the effective capacities of the physical links [7]. In general the use of multiple VPs for a single call means inefficient use of network resources, because the same resources could be used to complete the several separate calls [7]. Also it is shown that providing even as few as 2 or 3 multiple paths per source-destination pair will result in a marked decrease in call blocking probability [8].

An ATM network, as noted earlier, is connection-oriented network and in this respect it resembles the existing telephone networks. Since early 1980s, the trend in the telephone industry has been to implement dynamic alternative routing within nonhierarchical networks. Several routing schemes, which have been shown to offer excellent performance throughput and robustness, has been proposed and implemented in telephone network. This makes it natural to consider dynamic routing of VCs also in ATM networks.

4. Routing Algorithms

4.1. Statistical multiplexing and homogenous sources

4.1.1. Least Loaded Routing (LLR)

Various Least Loaded Routing (LLR) algorithms are proposed and developed to reduce the call blocking probability in cases when all the traffic sources have the same traffic characteristics and QoS requirements and the VCs are statistically multiplexed at the inputs to the VPs.

The QoS requirements which should be taken into account when a new call is established are cell loss, cell delay and cell jitter. The direct route is said to be QoS-permissible, if an additional VC does not violate QoS requirements in any directly routed or alternatively routed VCs on the VP in question. On the other hand, the alternative route is said to be QoS-permissible, if an additional VC does not violate QoS requirements in any directly or alternative routed VCs on neither of the links along which the VP is routed.

The LLR algorithms have many common features and they all operate as follows. A new call is always established on direct route, if the route is QoS-permissible. If the direct route is unavailable, then the alternative routes which both are QoS-permissible and Trunk Reservation (TR) -permissible are considered. A alternative route is said to be TR-permissible if in both links along which call will be directed, it is possible to accept the new call only based on required bandwidth and traffic parameters.

In [4] the LLR algorithms are categorised into three different classes according the procedure how they select an alternative route from the set of QoS-permissible routes.

- Unrestricted LLR algorithm chooses, from the set of QoS-permissible alternative routes, the one that is most TR-permissible i.e. has the smallest load
- Restricted LLR chooses from the set of QoS-permissible alternative routes, the one which will minimize the load
- Partially restricted LLR chooses, from the set of TR-permissible alternative routes which have at least one alternatively VC routed on each of its two VPs, the one that minimizes load. If the set is empty then the alternative route which minimizes load is selected.

The simulation made in [4], which include only cell loss requirement, show that the unrestricted LLR gives significantly lower blocking probabilities than restricted LLR or partially restricted LLR when the VC loads are light. When the loads are low or moderate, partially restricted LLR has blocking probability comparable to unrestricted LLR and significantly lower than restricted LLR.

If we take account the implementation efforts of the three algorithms: unrestricted: O(N^2), restricted: O(N), partially restricted: O(N), we see that the performance gains of the unrestricted come with an implementation cost. If also cell delay is included then the partially LLR gives a significantly better performance than restricted LLR. [4]

4.2. No statistical multiplexing and heterogeneous sources

When considering heterogeneous traffic routing we assume that the VC bandwidth is allocated according to peak rate of the connection, which may vary between zero and the bandwidth of the physical line.

Because every VC may have different bandwidth, the routing algorithm has to take account at least the following two issues: how to spread traffic evenly throughout the network and how to pack narrowband connection in order that there is also bandwidth for broadband connections which may require the whole bandwidth of the one VP.

4.2.1. Multirate LLR

In [4] the multirate version of LLR is introduced, which tries to spread traffic evenly in the network. The algorithm attempts first to set up the VC along the direct route, if there is capacity. If the direct route is full, then the alternative route which maximized the available bandwidth after establishing the new connection, is selected.

5. Conclusions

The QoS-sensitive path selection algorithm is still a research issue. The presence of the multiple constraints requires multiple link parameters and that makes it difficult to implement an efficient QoS-sensitive algorithm which also takes into account the traffic characteristics of the new VC as well as the characteristics of the VCs in the progress.

Many dynamic routing algorithms has been implemented in telephone networks. The knowledge and connection-oriented behaviour of the ATM-network makes it natural to consider the same routing schemes also in ATM networks. However, the choice of the right routing algorithm depend on the nonstandardized features of the ATM technology. This means also that the routing algorithms are open to vendor differentiation.

References

[1] ITU-T Recommendation I.121. Broadband Aspects of ISDN. 1990.

[2] ATM Forum. Private Network-to-Netowork Specification version 1.0, March 1996

[3] ITU-T Recommendation E.177. B-ISDN Routing. 1996

[4] Raif O. Onvural, Asynchronous Transfer Mode Networks, Artech House, 1995

[5] Martha Steenstrup, Routing in communications networks, Prentice Hall, 1995

[6] Marzo J.L., Domingo J., Fabregat R., Sole J. Dynamic routing based on a single parameter: Link Congestion Probability. High Performance Networking VI, IFIP Sixth International Conference on High Performance Networking, 1995, p.307-18

[7] Gerstel O., Segall A., Dynamic maintenance of Virtual Paths Layout, IEEE INFOCOM 1995, p.330-7 Vol 1.

[8] Jordan, T.P., Morse, M.J. (1994); Routing algorithms in ATM networks. IEE Eleventh UK teletraffic symposium. Performance engineering in telecommunication networks, 1994, 376 pp