Distributed Traffic Control Method for Tbit/s Multi-stage ATM Switching Systems

Kohei Nakai, Eiji Oki and Naoaki Yamanaka
NTT Network Service Systems Laboratories
3-9-11 Midori-cho, Musashino-shi, Tokyo 180-8585 Japan
tel: +81-422-59-3608, fax:+81-422-59-4819
E-mail: nakai@nttssl.nslab.ntt.co.jp

Abstract

This paper proposes a distributed traffic control method for large multi-stage ATM switching systems. In the proposed switching system, each port of the basic switches has its own traffic monitor, and each line unit (LU) periodically obtains congestion information about available paths from OAM cells in order to route a new virtual circuit independently. The performance of the proposed system depends on the interval between OAM cells. We show how an appropriate interval can be determined in order to maximize the number of user cells that each LU can send.

1 Introduction

Asynchronous transfer mode (ATM) is believed to yield the best high-speed multimedia infrastructure. ATM networks can handle various services, such as high-speed data communications, real-time video conferences, and video-on-demand (VOD). With demand for such services expanding, ATM networks will require switching systems that offer Tbit/s throughput[1, 2].

We previously proposed a scalable three-stage ATM switching architecture that uses optical WDM (wavelength division multiplexing) grouped links and dynamic bandwidth sharing[3, 4]. The former reduces the number of cables required and the latter prevents the statistical multiplexing gain from fading as the switching system expands.

That switching system required explicit path selection. When a new virtual circuit is to be added from some input line unit (LU) to some output LU, the system must find an appropriate path through the switching network with sufficient unused bandwidth on each of its links to accommodate the new virtual circuit.

A conventional system has a central control processor which manages the bandwidth on all of links in the switching network. The processor handles all virtual circuits in the system. And it finds a potential path having enough unused capacity. However, it is difficult for a central control processor to handle all the virtual circuits when a switching system becomes large. A large switching system whose throughput is Tbit/s class, requires a distributed traffic control method.

We propose a new distributed traffic control method for large multi-stage switching systems. In this switching system, each port of the basic switches has its own traffic monitor. And each LU periodically obtains congestion information about available paths from OAM cells in order to route a new virtual circuit independently.

The performance of this system depends on the interval between OAM cells. If we set it too long, the congestion information in LUs is delayed, but if we set it too short, the number of OAM cells in the system increases. We show how an appropriate interval between OAM cells can be determined in order to maximize the number of user cells that each LU can send.
The remainder of this paper is as follows. Section 2 briefly reviews a scalable 3-stage ATM switch architecture with optical WDM grouped links. Section 3 presents a distributed traffic control method and shows how an appropriate interval between OAM cells can be determined. Finally, section 4 summarizes the key points.

2 Scalable three-stage ATM switching architecture with WDM grouped links

Figure 1 shows a scalable three-stage ATM switch architecture with optical WDM grouped links [3, 4]. The expansion factor is $M$ times the basic network, so the total throughput of this system is $MN$ times that of the basic switch. For example, a system capacity of 5.2 Tbit/s can be achieved using 8 x 8 80-Gbit/s basic switches ($N = 8, M = 8$).

Each stage consists of $N$ switch groups and each switch group consists of $M$ basic switches. We call the basic switch a member to more clearly explain our new switch architecture. The $(n, m)$ basic switch is the $m$th member of the $n$th group.

Each stage consists of $N$ input ports and $N$ output ports. Each port multiplexes $M$ links corresponding to the $M$ wavelengths that are multiplexed into one optical fiber. The set of $M$ wavelengths is called a WDM group.

Each output port of the basic switch is connected to its intended switch group in the next stage. Each wavelength belonging to a WDM group of an output port is connected to its intended member in its intended switch group of the next stage by wavelength switching.

The bandwidth of the switch port is shared among each wavelength in the WDM group. The total bandwidth of wavelengths in the WDM group is limited to the speed of the switch port ($C$ bit/s). The bandwidth of each wavelength changes dynamically. The statistical multiplexing gain is not reduced even if the scale of the switching system is increased. When the basic network expands by $M$ times, the maximum bandwidth of each link ($C$ bit/s) is not divided into $C/M$ bit/s.

An example of cell routing is shown in Fig. 1. Let us consider the routing of an ATM connection from the first stage $(1, 1)$ to the third stage $(1, M)$. In the first stage, cells are routed
3 Distributed traffic control method for large multi-stage switching systems

Figure 2 shows our proposed traffic control method applied to a three-stage ATM switching systems using WDM grouped links. In this switching system, each port of the basic switches has its own traffic monitor. And each LU periodically obtains congestion information about available paths from OAM cells in order to route a new virtual circuit independently.

In each port of the basic switches, the residual bandwidth is estimated by counting arriving cells in a certain period. When the residual bandwidth in a port is below $B_r$ bit/s, the congestion information bit of arriving OAM cells is set to 1.

Each input LU sends OAM cells forward to output LUs every $T_o$ seconds. The output LUs return OAM cells to the input LUs. Each input LU routes virtual circuits independently using the congestion information in OAM cells. The routing lookup in each input LU inserts a path specification into the cell headers of arriving cells.

If we set $T_o$ too long, the congestion information in LUs is delayed. LUs route many virtual circuits to congested ports in the system. The total bandwidth assigned in a port exceeds the speed of the switch port ($C$ bit/s). We call this overbooking.

Figure 3 shows the relationship between $B_r$ and the overbooking probability. We must set $B_r$ to a wide bandwidth so as not to exceed the specified overbooking probability. Figure 4 shows $B_r$ for assuring that the overbooking probability $< 10^{-9}$. If we set $T_o$ to 1 ms, then $B_r$ must be set to 0.9 Gbit/s. ¹

On the other hand, if we set $T_o$ too short, the number of OAM cells in the system increases. The rate of the OAM cells sent by each LU ($B_o$ bit/s) is inversely proportional to $T_o$, as shown in Fig. 5. If $B_o$ increases, the number of user cells in the system decreases.

¹The rate of connections is 100 Mbit/s.
Figure 3: The overbooking probability when $B_r$ increases

Figure 4: $B_r$ for assuring the overbooking probability $< 10^{-9}$
The rate of user cells sent by each LU ($B_u$ bit/s) is given by

$$B_u = C - B_r - B_o$$  \hspace{1cm} (1)

Figure 6 shows the relationship between $T_o$ and $B_u$. $B_u$ has a maximum at a certain value of $T_o$. When the rate of connections is 100 Mbit/s, each LU can send 8.6 Gbit/s of user cells if we set $T_o$ to 1 ms.

4 Conclusions

This paper proposed a distributed traffic control method for large multi-stage ATM switching systems. When a new virtual circuit is to be added from some input LU to some output LU, a system must find an appropriate path through the switch network with sufficient unused bandwidth on each of its links to accommodate the new virtual circuit. In the proposed switching system, each port of the basic switches has its own traffic monitor, and each LU periodically obtains congestion information about available paths from OAM cells in order to route a new virtual circuit independently. The performance of the proposed system depends on the interval between OAM cells. If we set it too long, the congestion information in LUs is delayed, but if we set it too short, the number of OAM cells in the system increases. We showed how an appropriate interval can be determined in order to maximize the number of user cells that each LU can send. This traffic control method will suit future Tbit/s ATM switching systems.

References

Figure 6: The relationship between $T_o$ and $B_u$


