DD

KEK Preprint 95-189
January 1996
H/D

Stability of Coherent Data Traffic in a Switching Network

Y. NAGASAKA, M. NOMACHI, O. SASAKI and A. MANABE

959615

*Submitted to the Real Time 95 Conference, East Lansing, MI, U.S.A., May 23 - 26, 1995,
and will be published in the IEEE Transactions on Nuclear Science.*

National Laboratory for High Energy Physics, 1996

KEK Reports are available from:

Technical Information & Library
National Laboratory for High Energy Physics
1-1 Oho, Tsukuba-shi
Ibaraki-ken, 305
JAPAN

Phone: 0298-64-5136
Telex: 3652-534 (Domestic)
(0)3652-534 (International)
Fax: 0298-64-4604
Cable: KEK OHO
E-mail: Library@kekvax.kek.jp (Internet Address)

Stability of coherent data traffic in a switching network

Yasushi NAGASAKA, Masaharu NOMACHI, Osamu SASAKI, and Atsushi MANABE

National Laboratory for High Energy Physics
Oho 1-1, Tsukuba, Ibaraki 305, Japan

Abstract

Event data flow to switch-type event builder has a coherent traffic shape. Random data traffic has been intensively studied for telecommunication data. However, coherent traffic has not been studied yet. We have analyzed the stability of coherent data traffic. The study is indispensable to use a switching network for an event builder.

I. INTRODUCTION

A switching network with local data flow control is proposed to be used as a switch-type event builder. Such a switching network is designed for random traffic. Random traffic means the arrival of packets is a Markov process and the destinations of the packets are randomly labeled. Random traffic does not have a global structure. Therefore, the system behavior can be analyzed locally. Congestion in a switching network happens statistically on heavy traffic condition.

Figure 1 The number of event fragments in input queues. The results are plotted for only two input queues, crosses and circles.

However, an event data flow from front-end electronics to an event builder has a coherent traffic shape. Event fragments arrive coherently. The destinations of those event fragments are the same. In this case, local data flow control doesn't result in the most efficient data flow control. Congestion in a switching network or at the destination nodes cannot be avoided by local data flow control even for a small volume of

traffic. In the case of local data flow control, the congestion generates a back pressure to reduce the traffic to the congested area.

Figure 2 Latency profile of building an event as a function of an event rate.

An ATM (Asynchronous Transfer Mode) [1] switch with back pressure has been studied by RD-31 [2,3,4,5]. The ATM switching network system they have simulated has 1024 inputs and 1024 outputs, which is constituted by 5 stages of 4x4 ATM switches, and a local data flow control mechanism. The simulation result shows two unexpected behaviors. One is a profile of the number of event fragments for the simulated time (fig. 1). The number of event fragments doesn't increase gradually but it grows up to a very large value on the way to stabilize. The other is a latency profile of building an event as a function of the event rates (fig. 2). It has stable states but also large jumps. These behaviors happen only in coherent traffic. Therefore, the stability of coherent data traffic in a switching network must be studied to use it as an event builder system. We have studied the dynamics of coherent data streams in a switching element.

We would like to present the simulation results of coherent data traffic. The conditions that make the system stable will be described.

II. SIMULATION OF COHERENT TRAFFIC

A. Simulation model

The switching network system we assumed in our model has 512 inputs and 512 outputs. It is constituted by 9 stages of 2x2 ATM switches. The network configuration is a

baseline network [6] which is the same network topology as the simulation by RD31. The link configuration of a baseline network in the case of a 512x512 switching network by 2x2 ATM switches is illustrated in figure 3. If we look at one ATM switch at the first stage of the 512x512 switching network, the first 256 events from each input, i.e., from two detector sub-systems, try to go to the same output. Therefore, congestion has already occurred in the first stage. At first, we aim to examine only one ATM switch in the first stage and to analyze it in this paper.

Figure 3 The link configuration of baseline network in case of 512x512 switching network.

Coherent data traffic has been simulated to understand the behavior found in RD-31 (fig. 1 and 2). We simulated a 2x2 ATM switch and assumed it is in the first stage of a 512x512 switching network, because we considered the unexpected behaviors by RD31 are caused by the congestion in the first stage. The simulation was performed by MODSIM II [7].

The simulation model of a 2x2 ATM switch system is illustrated in figure 4. The input queue is a queue just before the switching network. The output queue is a queue in each destination switch at the next stage. However, the larger queues than that in the switch emulate all queues in the following switches.

Figure 4 Simulation model of a 2x2 ATM switch system that is in the first stage of 512x512 switching network. The bottom comments are the size of each queue.

The conditions of our simulation are: (1) the assumed switching network size is 512 inputs and 512 outputs; (2) the link bandwidth is 100 MB/sec; (3) the event fragment size has an exponential distribution that is an average value of 1 KB;

and (4) output data are transferred to output queues that emulate the following stages. The output queue has finite length and output bandwidth from the output queues is 50 MB/sec.

B. Simulation result

The first figure of our simulation results is the profile of the number of event fragments in two input queues (fig. 5). The simulation was performed for various traffic intensities. The traffic intensity (ρ) is the ratio of an actual event rate to the maximum one. In this figure, only the case of the traffic intensity 0.4 is shown. As the link bandwidth is 100 MB/sec in our simulation condition, the actual event rate is 40 MB/sec when the traffic intensity is 0.4. It shows a similar tendency to the result of the full 1024x1024 switching network by RD-31, although we simulate only one ATM switch in the first stage. The numbers of event fragments in two input queues are shown as crosses and circles. The profiles for both queues have a similar tendency. The starting points of decreasing and stabilizing, however, are different from each other.

Figure 5 Profiles of the number of event fragments in two input queues in case of traffic intensity 0.4.

III ANALYSIS OF COHERENT TRAFFIC

A. Phase of data stream

In the case of a 512x512 switching network, it is considered that the input data stream is packed by a unit of 256 event fragments. The first 256 event fragments from both detector sub-systems are going to the same output at the same time. Therefore, in this in-phase case of two data streams, the collision of event fragments occurs and the congestion in the switching element exists as illustrated in the top figure in figure 6. When congestion exists, the actual bandwidth is reduced and then the number of event fragments in each input

queue increases. But as the event fragment sizes from each detector sub-system are not exactly the same, the phase difference between them exists. The difference becomes larger and larger, until it becomes out-phase. That is, the data are going to the different destination and no congestion exists as illustrated in the bottom figure in figure 6. In this case, the actual bandwidth almost becomes the maximum one and each number of event fragments decreases toward a stable value.

Figure 6 The coherent data flow in ATM switch;
top: congestion exists in case of in-phase,
bottom: no congestion exists in case of out-phase.

To make sure, we have scatter plots of summation and difference of the number of event fragments in two input queues (fig. 7). The difference means the phase difference of two packed data streams which try to go to the same output. The unit of the phase difference is the number of event fragments. The plotted data proceed clockwise for time increasing. The first part which rises vertically indicates the case of congestion, and the other part means the case of no congestion. Finally the data stabilize. In this case the stable region is between 50 and 75 event fragments in phase difference.

Figure 7 Scatter plots of sum and difference of the number of event fragments in two input queues.

B. Speed of the number of event fragments

We have further analyze the figure 5 to understand the unexpected behavior. First we look at the profile before the number of event fragments is stable.

The number of event fragments changes according to the difference of actual input and output bandwidths. One output link is shared by two data streams in the in-phase case. Therefore, the actual average bandwidth will be about a half of the output bandwidth. On the other hand, in the out-phase case, the event fragments will be able to be transmitted with about maximum output bandwidth because of no congestion.

The amount of changes of the number of event fragments that depends on time. Namely a speed of the number of event fragments is defined. In figure 5, the speeds are calculated by slopes of each part of the profile. The first part is an increasing part and the next is a decreasing part. The last part is in the stable region and the speed of the part is zero. The speeds of various traffic intensities are plotted in figure 8. Positive speed means increase of the number of event fragments and negative speed means decrease. Circles are the plots in the case of increase and crosses are that in the case of decrease. Lines are also drawn in the figure. They are calculated by the following method with input and output bandwidths.

Figure 8 Speed of the number of event fragments in input queues for traffic intensities. Lines are calculations.

In our simulation condition, the link bandwidth is 100 MB/sec and the output bandwidth from output queue is 50 MB/sec. The input data rate is defined by $100 \cdot p$ MB/sec, where p is the traffic intensity. The output data rate is $50 \cdot 0.5$ MB/sec in the case of a complete collision. The solid line is calculated by $100 \cdot (p - 0.25)$ MB/sec from the above input and output data rates. On the other hand, the dotted line is calculated by $100 \cdot (p - 0.50)$ MB/sec in the case of decrease of them. Because it's considered without collision, that is, the

output data rate is nearly equal to the maximum bandwidth. The plots are little smaller than these calculation lines. However, the tendency of slope for the traffic intensity is almost the same.

C. Stable point of the phase difference

We have also analyzed the stable region. The profile in figure 7 is caused by the actual bandwidth changing because of changing phase difference. The stable point of the phase difference is equal to the number of event fragments in the stable region in figure 5. The value, therefore, is given by the unit of the number of event fragments. It is considered that an actual input bandwidth is almost the same as the output in the stable region. Therefore, the state can be calculated by the relation that the actual input bandwidth is equal to the output one. In our model, the input actual bandwidth in the stable region is $100 \cdot p$ MB/sec and the output is $50 \cdot (128/(256-x))$ MB/sec, where x is a phase difference of two packed data streams. $128/(256-x)$ means a ratio of the actual bandwidth to the maximum bandwidth including the consideration of the next stage queues.

From the above equations, the stable points are calculated as a function of a traffic intensity, $256-64/p$. This calculated line and the simulated values are plotted in figure 9. As shown in figure 5, in the stable region, one queue has a stable value of zero and the other has the value other than zero. These values plotted circles and crosses, respectively, in the figure.

Figure 9 Stable point of phase difference as a function of traffic intensity.

D. Latency

The number of event fragment is product of the latency and the event rate;

$$(\# \text{ of event fragments}) = (\text{event rate}) \cdot (\text{latency}).$$

This is the Little's Formula in the queuing theory [8];

$$L = \lambda W,$$

where L is an average queue length, that is, the number of event fragments, λ is the number of arrival in an input queue, that is, the event rate, and W is an average waiting time, that is, the latency. The event rate is the product of the maximum event rate and the traffic latency. According to the previous consideration, when the phase difference has the difference value of 1024 from the stable point; $(\text{stable point}) + n \cdot 1024$, where n is integer, in the case of a 1024x1024 switching network.

Figure 10 Latency profile with three calculated lines for $n=0, 1$, and 2 .

Namely, the stable point of the phase difference exists at every 1024. Therefore, the latency is calculated by

$$(\text{latency}) = (\text{constant}) + 1024 \cdot n/p.$$

We show the profile of latency (fig. 2) again with the calculated line of n equal to 0, 1, and 2 (fig. 10).

IV. CONCLUSION

We have simulated an ATM switch and analyzed the number of event fragments and latency of building an event. We have analyzed the switching network system with an ATM switch and found that they are stabilized when the data streams are out-phase. And some tendencies have been mostly understood and been able to be calculated analytically. However, to apply this type of switching network for an event builder, we need more studies.

V. ACKNOWLEDGMENTS

We would like to thank the RD-31 collaboration, especially I.Mandjavidze, for showing us some simulation results and for discussions with us.

VI. REFERENCES

- [1] For example, J. L. Boudec, "The Asynchronous Transfer Mode: a tutorial", *Comp. Net. and ISDN Systems*, 24 (1992) 279-309.

- [2] I.Mandjavidze, "Review of ATM, Fiber Channel and Conical Network Simulations" presentation at International Data Acquisition Conference, Fermilab, October 1994, submitted.
- [3] J. Christiansen et al., NEBULAS - A high performance data-driven event building architecture based on an asynchronous self-routing packet-switching network, CERN / DRDC / 93-55.
- [4] M. Costa et al., NEBULAS - High performance data-driven event building architectures based on asynchronous self-routing packet-switching networks, CERN / DRDC / 95-14.
- [5] Masaharu Nomachi, "Other research projects (Fermilab /LBL /KEK)" presentation at International Data Acquisition Conference, Fermilab, October 1994, submitted.
- [6] For example, Joseph Y. Hui, "Switching and Traffic Theory for Integrated broadband network", Kluwer Academic Publishers, Norwell, Massachusetts, 1990.
- [7] MODSIM II - The Language for Object-Oriented Programming, CACI Products Company, La Jolla, California, January 1993.
- [8] For example, E. Page, "Queueing Theory in OR", RUSSAK, New York, 1972.