Performance Comparison with Layer 3 Switches in cases of Flow- and Topology-driven Connection Setup

Shogo Nakazawa, Kenji Kawahara, Suguru Yamaguchi, Yuji Oie

S. Nakazawa is a Ph.D candidate in Graduate School of Information Science at Nara Institute of Science and Technology (NAIST), Ikoma, JAPAN. E-mail: shougo-n@is.aist-nara.ac.jp .

K. Kawahara is with Dept. of Computer Science and Electronics, Kyushu Institute of Technology (KIT).

S. Yamaguchi is with Graduate School of Information Science at NAIST.

Y. Oie is with Dept. of Computer Science and Electronics, KIT

Abstract

The layer 3 switch enables us fast to transmit IP datagrams using the cut-through technique. There are mainly two schemes about connection setup; one is the flow-driven case and another is the topology-driven one. In this paper, we analyze the cut-through rate, the datagram waiting time and the mis-order rate as performance measure of both cases and compare these performances. In the analysis, by using interrupted Bernoulli process (IBP), we model the arrival process of IP flow and IP datagram from each source. Furthermore, we investigate impacts of the arrival rate and the average IP flow length on performance.

Keywords

Layer 3 switch, Cut-through, Flow-driven, Topology-driven, IBP, Performance analysis

I. Introduction

On the Internet, the number of hosts have been growing exponentially every year. Within this overall growth rate, there is also increasing demand for multimedia applications in real-time, which have quite stringent quality of service (QoS) delivery in terms of packet delay, loss rate and minimum bandwidth. Furthermore, as the World Wide Web is increasingly used for business, there are most users for whom delay-bounded access of information is very important. Thus, the Internet backbone must increase the bandwidth available to end users as well as offer QoS provisioning. However, conventional IP router may be a bottleneck in terms of high-speed data transfer, since they carry out forwarding decisions in software based on address resolution with hop-by-hop transmission of IP datagrams of variable length [1].

On the other hand, a layer 2 switch based on ATM achieves high bit rates through hardware switching of fixed-length cells and commonly support QoS by connection-oriented cell transfer. Therefore, ATM can be exploited in an IP internetwork. This combination has been prompted by some proposals including classical IP over ATM [2] by Internet Engineering Task Force (IETF) and Multi-Protocol over ATM (MPOA) [3] by ATM Forum, in which some servers mapping next hop's IP address to its ATM one are required. However, this address resolution mechanism is not scalable. In recent years, so-called a **layer 3 switching** mechanism is employed to combine layer 3 routing with layer 2 high-speed switching efficiently and enables us to transmit IP datagrams fast. It has been implemented in Ipsilon's (now Nokia's) "IP switch[4]", Toshiba's "CSR(Cell Switch Router)[5]", Cisco's "Tag Switching[6]"and IBM's "ARIS(Aggregated Route-Based IP Switching)[7]" so far.

The layer 3 switch consists of the layer 3 datagram routing kernel and the layer 2 switching fabric. From now on, we will identify the switching fabric with ATM switch. The specific source-destination IP address pair is regarded as *flow*. When a datagram arrives at the layer 3 switch, the routing kernel distinguishes its source and destination IP addresses from the IP header, creates the specific *flow*, and maps it to some virtual connection (VC) on the ATM switch. The subsequent datagrams of the *flow* are directly transmitted on ATM switch without address resolution on the layer 3 routing kernel. This will be called a cut-through mechanism. There are mainly two schemes of VC association in this mechanism. One is the flow-driven case in which VC setup is done by soft-state and cut-through VCs can be dynamically allocated/released according to the IP flow generation/termination. Another is the topology-driven case in which VC setup is done by hard-state and cut-through VCs are pre-determined according to provided routing information. Therefore, the number of VC for cut-through on the ATM switch is one of the most important parameters in both cases of layer 3 switch. If there is not enough VCs for transmitting datagrams via the layer 2 switch, the rate of cut-through datagrams becomes small.

Hence, in this paper, we analyze three important performance measures of layer 3 switch, called the cut-through rate, the datagram waiting time and the mis-order rate. So far, we have only analyzed the flow-driven model by using 3-state Markov modulated Bernoulli process (MMBP) in [11]. In this paper, by using interrupted Bernoulli process (IBP) as the arrival process of IP flow and IP datagram, we can deal with the greater number of sources than in [11]. In addition, by extending the analysis for the topology-driven case, we can compare performance in both cases of layer 3 switch. Furthermore, we will investigate impacts of the arrival rate and the average IP flow length on performance through some numerical results.

The rest of this paper is structured as follows. Section 2 explains the mechanism of the layer 3 switch and interrupted Bernoulli process (IBP) as a source model. In section 3 and 4, we analyze the steady state probability of the layer 3 switch and derive performance measures for the cases of the flow-driven and the topology-driven, respectively. In section 5, we provide numerical results and examine the effects of certain parameters on performance. Finally, section 6 gives a brief conclusion and future works.

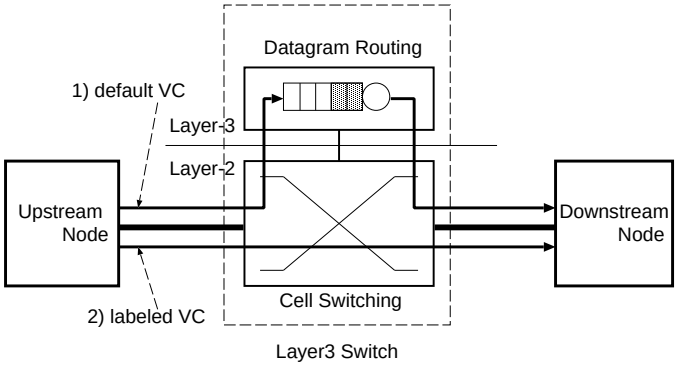


Fig. 2.1. Concept of layer 3 switching

II. Traffic Model

A. Layer 3 switch: Cut-through mechanism

Figure 2.1 illustrates the concept of the layer 3 switching mechanism. The layer 3 switch consists of the layer 3 datagram routing kernel and the layer 2 switching fabric or ATM switch. When a datagram from the upstream node via *default VC* arrives at the layer 3 switch, the datagram is firstly raised to the routing kernel to check its source and destination IP addresses, then it is forwarded to the downstream node which is determined according to its destination IP address. Simultaneously, the routing kernel regards this source-destination IP address pair as a *flow* and maps it to some VC, which is called *labeled VC*. Subsequent datagrams of the *flow* are directly transmitted on ATM switch via the labeled VC without address resolution on the routing kernel. This mechanism is called *cut-through* and it allows datagrams to be quickly transmitted. Here, VC association is flow-driven for cut-through VCs. On the other hand, it is called topology-driven when cut-through VCs are pre-determined according to the routing entries managed by the routing protocol. The former type is IP switch[4] and CSR[5], and the latter is Tag Switching[6] and ARIS[7].

In this paper, we will assume that the system is in the ideal state in forms of cut-through VC of the layer 2. That is, in case of the flow-driven type, a cut-though VC is assigned during just one datagram transmission, and in case of the topology-driven type, a VC is allocated while one IP flow is established. Under these assumption, we will analyse both type of layer 3 switches.

B. Source model description

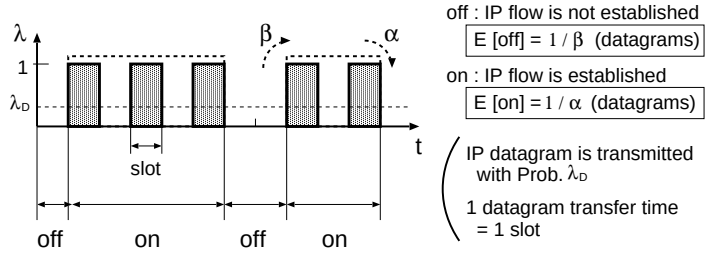


Fig. 2.2. Source model : IBP

The IP flow model is shown in Fig. 2.2. In this paper, we assume that an IP datagram length is fixed and consider a discrete-time queueing system with time slot being equal to a datagram transmission time. Each source generates datagrams according to an interrupted Bernoulli process (IBP), which is characterized by a parameter set of $(\lambda_D, \alpha, \beta)$. IBP has two states: on-state and off-state. In each slot, the process changes its state from on to off with probability α, and from off to on with probability β. In on-state, one datagram arrives in each slot with probability λ_D, while no datagram arrives during off-state. We suppose that the period of on-state correspond to the IP flow established by the source. Then, the steady state probability of each state is easily given by:

$$\pi_{\text{on}} = \frac{\beta}{\alpha + \beta} \quad , \quad \pi_{\text{off}} = \frac{\alpha}{\alpha + \beta} \quad . \tag{2.1}$$

Thus, we have the average arrival rate λ of each source as:

$$\lambda = \pi_{\text{on}} \lambda_D . \tag{2.2}$$

III. Analysis of flow-driven model

First, we will consider the flow-driven case in this section and deal with the topology-driven one in the next section.

A. Analytical model assumption

Figure 3.1 illustrates our queueing model in case of a flow-driven connection setup. We suppose that the layer 2 switching fabric transmits at most N datagrams in a slot while the routing kernel forwards only one datagram and its buffer size is infinite. N sources are multiplexed at the layer 3 switch and they are homogeneous in terms of traffic characteristics, that is, each source has a parameter set of IBP, $(\lambda_D, \alpha, \beta)$. When the IP flow is established by some sources, the

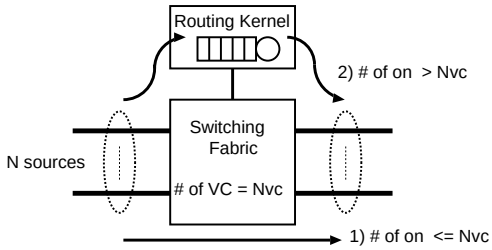


Fig. 3.1. Analytical model : flow-driven model

cut-through VC has been assigned during just transmitting one IP datagram. Since we do not suppose the buffering on the ATM switch, the waiting time of datagrams is equal to 0. Here, let N_{VC} denotes the number of assigned VC for cut-through. Note that if the number of sources generating datagrams exceeds N_{VC}, then N_{VC} datagrams are cut through while the rest of them are lifted up to the routing kernel.

B. Random variables of the system

We define following random variables in the t-th slot:

- $D(t)$: the queue length in datagrams of the buffer at the routing kernel;
- $S_{\text{on}}(t)$: the number of sources being in the on-state.

Since sources are homogeneous in terms of traffic characteristics, we can completely describe the state of our system by above variables. Furthermore, each source generates datagrams according to the IBP described in the previous section, so that the system state $\{D(t), S_{\text{on}}(t); t = 0, 1, 2, \cdots\}$ can form a discrete time Markov chain. Note that when $S_{\text{on}}(t) = i$, the number of sources in the off-state is given by $N - i$, where N denotes the number of sources in our system.

In the rest of this section, we first give the transition probability of the number of sources and the datagram arrival probability. Thereby we derive the steady state probability, finally we obtain some performance measures.

C. Transition probability of the number of sources being in on-state

We first define the transition probability of the number of sources in the on-state.

$$P_{i,j} \stackrel{\triangle}{=} \Pr\{S_{\text{on}}(t) = j | S_{\text{on}}(t-1) = i\}. \tag{3.1}$$

By assuming that q sources change their states from on to off, we can completely give the

Table 3.1. Transition of the number of sources in the on-state

	$S_{\text{on}}(t)$	$S_{\text{off}}(t)$	
$S_{\text{on}}(t-1)$	***i-q***	***q***	***i***
$S_{\text{off}}(t-1)$	***j-i+q***	***N-j-q***	***N-i***
	j	***N-j***	***N***

transition of the number of sources as shown in Table 3.1. From this table, we thus have $P_{i,j}$ as

$$P_{i,j} = \sum_{q=\max(0,i-j)}^{\min(i,N-j)} \binom{i}{q} \alpha^q (1-\alpha)^{i-q} \binom{N-i}{j-i+q} \beta^{j-i+q} (1-\beta)^{N-j-q}. \tag{3.2}$$

D. Definition of datagram arrival probability at the routing kernel

By letting $a(t)$ denote a random variable that represents the number of datagrams arriving at the routing kernel in a slot, we define the datagram arrival probability $a_{i,j}(x)$ and obtain it as follows.

$$\begin{aligned} a_{i,j}(x) &= \Pr\{a(t) = x, S_{\text{on}}(t) = j \,|\, S_{\text{on}}(t-1) = i\} \\ &= \begin{cases} P_{i,j} & \text{if } x = 0, i \le N_{VC} \\ 0 & \text{if } x > 0, i \le N_{VC} \\ \sum_{k=0}^{N_{VC}} P_i(k) P_{i,j} & \text{if } x = 0, i > N_{VC} \\ P_i(N_{VC} + x) P_{i,j} & \text{if } x > 0, i > N_{VC} \end{cases} \end{aligned} \tag{3.3}$$

where $P_i(x)$ denotes the probability that x datagrams are stored in the queue of the routing kernel from i sources. $P_i(x)$ is given by the binomial distribution as follows:

$$P_i(x) = \binom{i}{x} \lambda_D^x (1-\lambda_D)^{i-x}, \quad \text{for } i \ge x. \tag{3.4}$$

E. Steady state probability

We define the system state probability at the t-th slot as

$$P_t(b, i) \stackrel{\triangle}{=} \Pr\{D(t) = b, S_{\text{on}}(t) = i\}, \tag{3.5}$$

and the steady state probability $x(b, i)$ and its vector representation as

$$\begin{aligned} x(b,i) &\stackrel{\triangle}{=} \lim_{t\to\infty} P_t(b,i) \\ \boldsymbol{x}_b &= [x(b,0), \cdots, x(b,i), \cdots, x(b,N)] \end{aligned} \tag{3.6}$$

We next show the transition probability matrix $\boldsymbol{A}(x)$ representing that x datagrams arrive at the routing kernel with changing state of sources, as follows:

$$\boldsymbol{A}(x) = \begin{pmatrix} a_{0,0}(x) & \cdots & a_{0,j}(x) & \cdots & a_{0,N}(x) \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{i,0}(x) & \cdots & a_{i,j}(x) & \cdots & a_{i,N}(x) \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{N,0}(x) & \cdots & a_{N,j}(x) & \cdots & a_{N,N}(x) \end{pmatrix}. \tag{3.7}$$

Since the system state forms Markov chain of M/G/1 type queueing system, the steady state probability vector $\boldsymbol{x}_b$ can be related to the transition probability matrix $\boldsymbol{A}(x)$ as follows [9]:

$$\begin{aligned} \boldsymbol{x}_b &= \boldsymbol{x}_0 \boldsymbol{A}(b) + \sum_{i=1}^{b+1} \boldsymbol{x}_i \boldsymbol{A}(b+1-i), \quad \text{for } b \geq 0 \\ \sum_{b=0}^{\infty} \boldsymbol{x}_b \boldsymbol{e} &= 1, \end{aligned} \tag{3.8}$$

where $\boldsymbol{e}$ is a column vector of ones and Eq. (3.8) can be stably solved by using the algorithm proposed in [10].

F. Derivation of performance measures

By using the steady state probability obtained in the previous subsection, we can get following performance.

F.1 Cut-through rate, R_c

We will first derive the cut-through rate R_c. R_c is the ratio of the average number of datagrams when the number i of sources in on-state and datagrams generated from them become less than or equal to N_{VC}, to the average number of datagrams generated from whole i sources being in

on-state. This is given by

$$R_c = \frac{\sum_{b=0}^{\infty}\sum_{i=1}^{N} x(b,i)\sum_{k=0}^{i}\min(k, N_{VC})P_i(k)}{\sum_{b=0}^{\infty}\sum_{i=1}^{N} x(b,i)\sum_{k=0}^{i} kP_i(k)} \quad . \tag{3.9}$$

F.2 Average datagram waiting time, W_d

We let the z transform of the steady state probability vector $\boldsymbol{x}_b$ be $\boldsymbol{X}(z)$, the first equations of (3.8) lead to

$$\boldsymbol{X}(z)[z\boldsymbol{I}(M_s) - \boldsymbol{A}(z)] = \boldsymbol{x}_0(z-1)\boldsymbol{A}(z), \tag{3.10}$$

where $\boldsymbol{A}(z) = \sum_{i=0}^{\infty} \boldsymbol{A}(i)z^i$ and $\boldsymbol{I}(M_s)$ is the unit matrix of the dimension, M_s. Neuts[9] shows that the n-th moment of the probability density $\boldsymbol{x}_b$ is obtained by the n-times differentiation of $\boldsymbol{X}(z)$ with respect to z. Thus, the average queue length L of the buffer at the routing kernel is given by $L = \left.\frac{d\boldsymbol{X}(z)}{dz}\right|_{z=1}$. Since the total arrival rate at the layer 3 switch becomes $N\lambda$, by using Little's formula, the average datagram waiting time W_d can be obtained as follows:

$$\begin{aligned} W_d &= \frac{\sum_{b=0}^{\infty} b\sum_{i=0}^{N} x(b,i)}{N\lambda} \quad \text{(datagrams)} \\ &= \frac{\sum_{b=0}^{\infty} b\sum_{i=0}^{N} x(b,i)}{N\lambda} \times L_p \quad \text{(cells)}, \end{aligned} \tag{3.11}$$

where L_p is datagram length in cells.

F.3 Mis-order rate, R_m

Finally, we will obtain the mis-order rate R_m. The term "mis-order" means the out-of-order transmission of datagrams in [8] and we also use the term in the paper. The mis-order occurs according to following steps;

1. A first datagram of a flow, it will be called *trigger datagram* is stored in the routing kernel and the cut-through connection is established by its arrival.
2. The subsequent datagrams of the flow are directly transferred on the ATM switch by cut-through before the trigger datagram is served at the routing kernel.

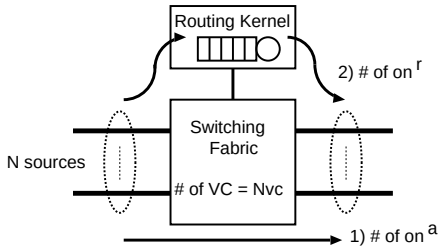


Fig. 4.1. Analytical model : topology-driven model

Therefore, the mis-order rate R_m is obtained by

$$R_m = \frac{\sum_{b=1}^{\infty}\sum_{i=N_{VC}+1}^{N} x(b,i) \sum_{k=N_{VC}+1}^{i} (k-N_{VC})P_i(k) \sum_{m=1}^{b} \{(1-\lambda_D)(1-\alpha)\}^{m-1} \lambda_D(1-\alpha)R_c}{\sum_{b=1}^{\infty}\sum_{i=0}^{N} x(b,i) \sum_{k=0}^{i} kP_i(k)}$$

$$= \frac{R_c(1-\alpha)\lambda_D \sum_{b=1}^{\infty}\sum_{i=N_{VC}+1}^{N} x(b,i) \sum_{k=N_{VC}+1}^{i} (k-N_{VC})P_i(k) \left(1-\{(1-\lambda_D)(1-\alpha)\}^{b}\right)}{(1-(1-\lambda_D)(1-\alpha)) \sum_{b=1}^{\infty}\sum_{i=0}^{N} x(b,i) \sum_{k=0}^{i} kP_i(k)} . \quad (3.12)$$

IV. Analysis of topology-driven model

A. Analytical model assumption

Figure 4.1 illustrates our queueing model in case of a topology-driven connection setup.

This model has the same parameters as the flow-driven model in terms of the switching fabric and the routing kernel. However, in contrast with the case of flow-driven model, all datagrams generated during one on period, that is, IP flow, with the same source-destination IP address pair, can be transmitted via one VC on layer 2 because the VC for cut-through is preserved for the IP flow based on its routing information. Therefore, the source that will generate the IP flow, may have two states according to the allowable number of VC for cut-through, namely, it is accepted or rejected for cut-through. So we will define the following states;

- on^a : the number of sources being assigned the cut-through VC;
- on^r : the number of sources not being assigned the cut-through VC.

If the source is in on^a-state, it keeps the cut-through VC until the IP flow terminates, while all datagrams from sources being in on^r-state are lifted up to the routing kernel.

B. Random variables of the system

We define following random variables in the t-th slot:

- $D(t)$: the queue length in datagrams of the buffer at the routing kernel;
- $S_{\text{on}^a}(t)$: the number of sources being in the on^a-state;
- $S_{\text{on}^r}(t)$: the number of sources being in the on^r-state.

Since sources are homogeneous in terms of traffic characteristics, we can completely describe the state in the topology-driven case by above variables. Furthermore, each source generates datagrams according to the IBP described in the previous section, so that the system state $\{D(t), S_{\text{on}^a}(t), S_{\text{on}^r}(t); t = 0, 1, 2, \cdots\}$ can form a discrete time Markov chain. Note that when $S_{\text{on}^a}(t) = j$ and $S_{\text{on}^r}(t) = k$, the number of sources in the off-state is given by $N - j - k$, where N denotes the number of sources in our system. Here, we define S as the state space for $(S_{\text{on}^a}(t), S_{\text{on}^r}(t))$, i.e., $S = \{(j, k)|0 \leq j \leq N_{VC}, 0 \leq k \leq N - j\}$. A state (j, k) is labeled using the following one-to-one mapping function $M(j, k)$:

$$M(j, k) = \frac{1}{2}j(2R - j + 3) + k. \tag{4.1}$$

Therefore, S can be equivalently represented by the state space $S^* = \{0, \cdots, M_s - 1\}$, where $M_s = (N_{VC} + 1)(2N - N_{VC} + 2)/2$.

C. Transition probability of the number of sources being in on-state

We first define the transition probability in terms of the number of sources as follows.

$$P_{M(j,k),M(m,n)} \stackrel{\triangle}{=} \Pr\{S_{\text{on}^a}(t) = m, S_{\text{on}^r}(t) = n | S_{\text{on}^a}(t-1) = j, S_{\text{on}^r}(t-1) = k\} \tag{4.2}$$

We assume that y sources change those states from off at a $(t-1)$-th slot to on at a t-th slot, we classify the following two cases;

1. When $y \leq N_{VC} - j$,

   Datagrams from y sources can be cut through, that is, the state of all y sources becomes the on^a-state;

2. On the other hand, when $y > N_{VC} - j$,

   (a) Datagrams from $N_{VC} - j$ sources can be cut through;

   (b) On the other hand, Datagrams from $y - (N_{VC} - j)$ sources are not cut through, that is, their sources change their states to on^r.

Table 4.1. Transition of the number of sources in each state when $y \leq N_{VC} - j$

	$S_{\text{off}}(t)$	$S_{\text{on}^a}(t)$	$S_{\text{on}^r}(t)$	
$S_{\text{off}}(t-1)$	***N-j-k-y***	***y***	—	***N-j-k***
$S_{\text{on}^a}(t-1)$	***j-m+y***	***m-y***	—	***j***
$S_{\text{on}^r}(t-1)$	***k-n***	—	***n***	***k***
	N-m-n	***m***	***n***	***N***

Table 4.2. Transition of the number of sources in each state when $y > N_{VC} - j$

	$S_{\text{off}}(t)$	$S_{\text{on}^a}(t)$	$S_{\text{on}^r}(t)$	
$S_{\text{off}}(t-1)$	***N-j-k-y***	***N_{VC}-j***	***y+j-N_{VC}***	***N-j-k***
$S_{\text{on}^a}(t-1)$	***N_{VC}-m***	***j-N_{VC}+m***	—	***j***
$S_{\text{on}^r}(t-1)$	***y-N_{VC}+j-n+k***	—	***N_{VC}-j+n-y***	***k***
	N-m-n	***m***	***n***	***N***

First, we consider $Q_{M(j,k),M(m,n)}$ which is the transition probability of the number of sources in the $(S_{\text{on}^a}(t), S_{\text{on}^r}(t))$ state in case that $y \leq N_{VC} - j$. Table 4.1 shows the transition of the sources in this case. From this table, we thus have $Q_{M(j,k),M(m,n)}$ as

$$Q_{M(j,k),M(m,n)} = \sum_{y=\max(0,m-j)}^{\min(m,N-j-k,N_{VC}-j)} \binom{N-j-k}{y} \beta^y (1-\beta)^{N-j-k-y} \times \binom{j}{m-y} \binom{k}{n} \alpha^{j+k-m-n+y} (1-\alpha)^{m+n-y}, \quad \text{if } k \geq n. \tag{4.3}$$

Next, we consider $R_{M(j,k),M(m,n)}$ which is the transition probability of the number of sources in case that $y > N_{VC} - j$. Table 4.2 shows the transition of the sources in this case. The same way as Eq.(4.3), we thus have $R_{M(j,k),M(m,n)}$ as

$$R_{M(j,k),M(m,n)} = \sum_{y=\max(N_{VC}-j+n-k,N_{VC}-j+1)}^{\min(N-j-k,N_{VC}-j+n)} \binom{N-j-k}{y} \beta^y (1-\beta)^{N-j-k-y} \times \binom{j}{N_{VC}-m} \binom{k}{N_{VC}-j+n-y} \alpha^{j+k-m-n+y} (1-\alpha)^{m+n-y}, \quad \text{if } j+m \geq N_{VC}. \tag{4.4}$$

From Eq.(4.3) and (4.4), $P_{M(j,k),M(m,n)}$ is given by

$$P_{M(j,k),M(m,n)} = Q_{M(j,k),M(m,n)} + R_{M(j,k),M(m,n)} \quad . \tag{4.5}$$

D. Definition of datagram arrival probability at the routing kernel

By letting $a(t)$ denote a random variable that represents the number of datagrams arriving at the routing kernel, we define the datagram arrival probability $a_{M(j,k),M(m,n)}(x)$ and obtain it as follows.

$$\begin{aligned} a_{M(j,k),M(m,n)}(x) &= \Pr\{a(t) = x, S_{\text{on}^a}(t) = m, S_{\text{on}^r}(t) = n | S_{\text{on}^a}(t-1) = j, S_{\text{on}^r}(t-1) = k\} \\ &= \begin{cases} P_{M(j,k),M(m,n)} & \text{if } k = 0, x = 0 \\ 0 & \text{if } k = 0, x > 0 \\ P_k(x) P_{M(j,k),M(m,n)} & \text{if } k > 0 \end{cases} \end{aligned} \tag{4.6}$$

where $P_i(x)$ is the same of Eq.(3.4).

E. Steady state probability

We define the system state probability at the t-th slot as

$$P_t(b, M(j,k)) \stackrel{\triangle}{=} \Pr\{D(t) = b, S_{\text{on}^a}(t) = j, S_{\text{on}^r}(t) = k\}, \tag{4.7}$$

and the steady state probability $x(b, M(j,k))$ and its vector representation as

$$\begin{aligned} x(b, M(j,k)) &\stackrel{\triangle}{=} \lim_{t\to\infty} P_t(b, M(j,k)) \\ \boldsymbol{x}_b &= [\cdots, x(b, M(j,k)), \cdots] \\ & \qquad 0 \le M(j,k) \le M_s - 1 \end{aligned} \tag{4.8}$$

We next show the transition probability matrix $\boldsymbol{A}(x)$ representing that x datagrams arrive at the routing kernel with changing state of sources, as follows:

$$\boldsymbol{A}(x) = \begin{pmatrix} a_{0,0}(x) & \cdots & a_{0,M(m,n)}(x) & \cdots & a_{0,M_s-1}(x) \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{M(j,k),0}(x) & \cdots & a_{M(j,k),M(m,n)}(x) & \cdots & a_{M(j,k),M_s-1}(x) \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{M_s-1,0}(x) & \cdots & a_{M_s-1,M(m,n)}(x) & \cdots & a_{M_s-1,M_s-1}(x) \end{pmatrix}. \tag{4.9}$$

Since the system state forms Markov chain of M/G/1 type queueing system, the steady state probability vector $\boldsymbol{x}_b$ can be shown by Eq.(3.8).

F. Derivation of performance measures

By using the steady state probability obtained in the previous subsection, we can get some performance measures. However, the mis-order does not occur while the IP flow is transmitted because the cut-through VC is assigned per IP flow in case of topology-driven model.

F.1 Cut-through rate, R_c

In the same way of Eq.(3.9), the cut-through rate R_c is given by

$$R_c = \frac{\sum_{b=0}^{\infty}\sum_{j=1}^{N_{VC}}\sum_{k=0}^{N-j} x(b, M(j,k))\sum_{a=0}^{j} aP_j(a)}{\sum_{b=0}^{\infty}\sum_{j=0}^{N_{VC}}\sum_{k=0}^{N-j} x(b, M(j,k))\sum_{i=0}^{j+k} iP_{j+k}(i)} \quad . \tag{4.10}$$

F.2 Average datagram waiting time, W_d

In the same way of Eq.(3.11), the average datagram waiting time W_d is given by

$$W_d = \frac{\sum_{b=1}^{\infty} b\sum_{j=0}^{N_{VC}}\sum_{k=0}^{N-j} x(b, M(j,k))}{N\lambda} \times L_p \quad \text{(cells)}, \tag{4.11}$$

where L_p is datagram length in cells.

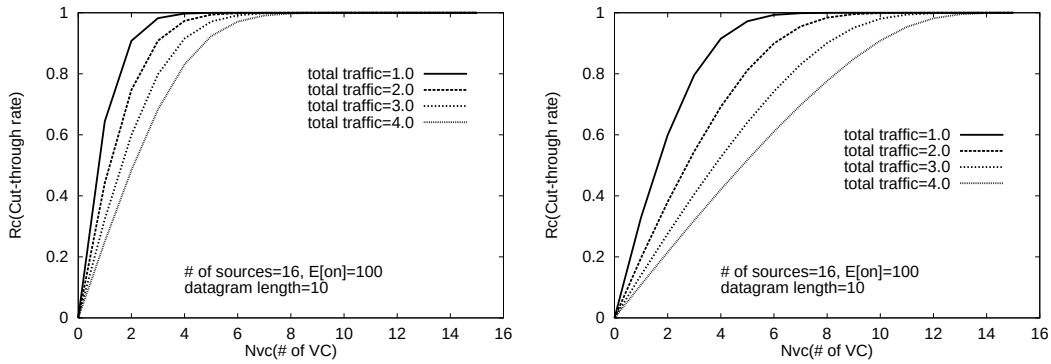
V. Numerical Results and Discussions

In this section, we will show some numerical results based on performance measure derived in the preceding sections and compare performance of the flow-driven case with that of the topology-driven one. Throughout this section, we basically set the number N of sources to 16 and the probability λ_D that one datagram arrives during a slot in a flow to 0.5.

A. Influence of total arrival rate

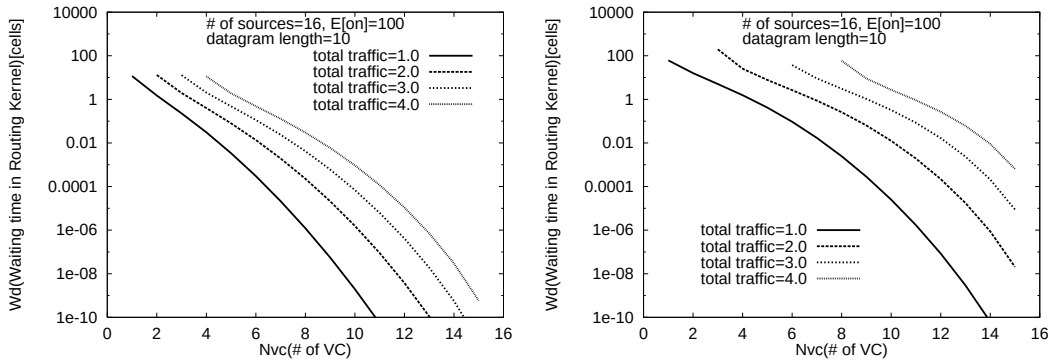
First, we will show the impact of the total arrival rate $N\lambda$ at each layer 3 switch on performance. We set $N\lambda$ to 1.0, 2.0, 3.0 and 4.0. Furthermore, we suppose the datagram length, L_p to 10 cells and set the average flow length, $E[\text{on}]$ to 100 cells.

The cut-through rate R_c is illustrated in Fig. 5.1. We can see that the smallest N_{VC} achieving that R_c = 1 becomes larger as $N\lambda$ is larger in both flow-driven and topology-driven model. For example, if $N\lambda$ = 1.0, the value of N_{VC} becomes 4, namely a 25% of the total numbers of



(a) Flow-driven model (b) Topology-driven model

Fig. 5.1. Cut-through rate, R_c: Influence of total arrival rate, $N\lambda$



(a) Flow-driven model (b) Topology-driven model

Fig. 5.2. Average datagram waiting time W_d: Influence of total arrival rates, $N\lambda$

VC in case of flow-driven model, whereas the value becomes 6, namely a 38% of them in case of topology-driven model. Therefore, we can find that the flow-driven model is superior to the topology-driven one with respect to the cut-through rate.

Figure 5.2 shows the average datagram waiting time W_d. Note that the queue length distribution obtained from the steady state probability $\boldsymbol{x}_b$ is unstable when the actual arrival rate at the routing kernel $N\lambda(1 - R_c)$ exceeds 1.0 of the capacity of the routing kernel, thus we omit the values in this range of N_{VC}. However, in both models, W_d exponentially increases as the total traffic rate increases, the performance of topology-driven model becomes worse remarkably. Furthermore we can say datagrams in flow-driven model are transmitted about ten times faster

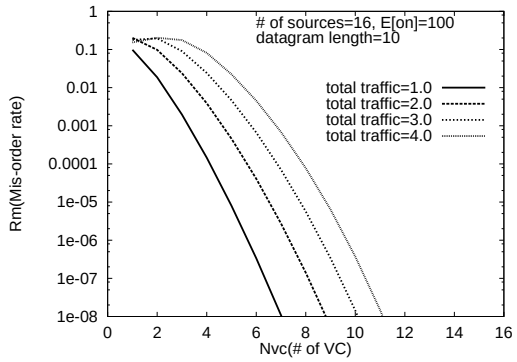


Fig. 5.3. Mis-order rate, R_m: Influence of total arrival rates, $N\lambda$

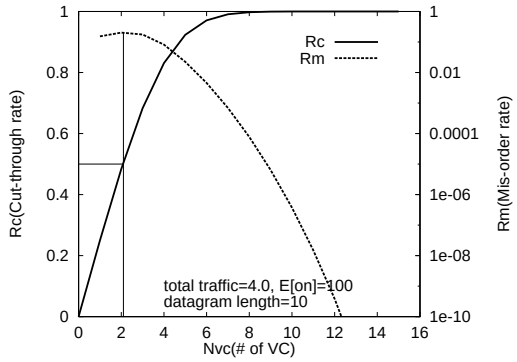


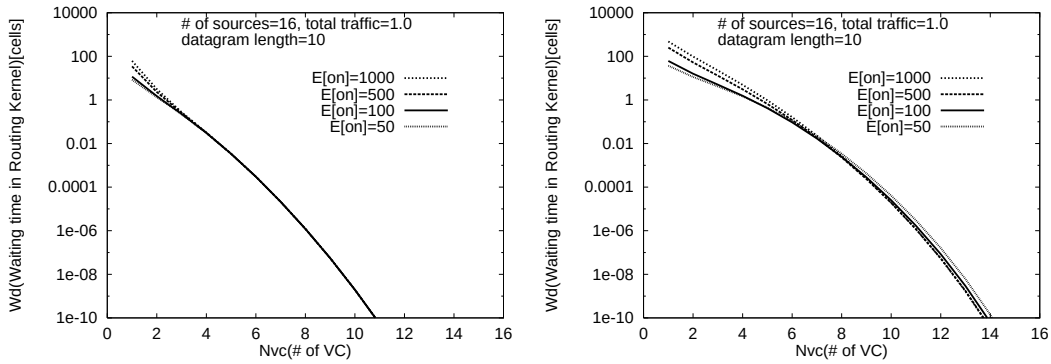
Fig. 5.4. Correlation of R_c and R_m (total arrival rates=4.0)

than in topology-driven model.

Figure 5.3 illustrates the mis-order rate R_m in case of flow-driven model. If N_{VC} is 0, the mis-order transmission of datagrams does not occur, that is R_m is 0. We can see that R_m exponentially increases as the total traffic rate increases. Furthermore, when the total arrival rate is 4.0, we can see that R_m has the maxmum value. Then, we show the correlation of R_c and R_m in that case in Fig. 5.4. In [11] which deals with 3-state MMBP as the source model, R_m is in proportion to $R_c(1 - R_c)$ which is maximized when R_c becomes 0.5. In our analytical model which makes reduction of the number of states, we can say the same feature.

B. Influence of the average IP flow length

We show the impact of the average IP flow length, $E[\text{on}]$ in Fig. 5.5. In this figure, we set $E[\text{on}]$ to 50, 100, 500 and 1000. Furthermore, the datagram length, L_p is set to 10 in cells, respectively. We set the average arrival rate λ to 0.0625, that is, total arrival rate $N\lambda$ equals 1.



(a) Flow-driven model (b) Topology-driven model

Fig. 5.5. Average datagram waiting time, W_d: Influence of the average IP flow length, E[on]

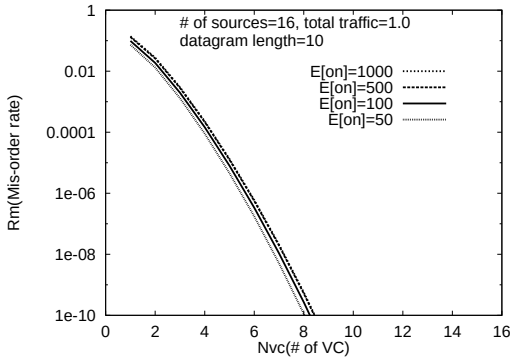


Fig. 5.6. Mis-order rate, R_m: Influence of the average IP flow length, E[on]

Note here that, if the average arrival rate λ is the same value in any cases of IP flow length, the cut-through rate R_c is also the same although some IBP parameters (λ_D, α, β) are quite different. In this subsection, since we suppose λ to the same value, we omit the figure of R_c that is the same as Fig. 5.1.

From Fig. 5.5, W_d becomes larger when E[on] is larger if N_{VC} is set to comparatively small values, otherwise W_d is the same for each E[on].

Figure 5.6 illustrates the mis-order rate R_m in case of flow-driven model. From Fig. 5.6, R_m is smaller as E[on] is set to a smaller value, the differences among them are quite little, namely, R_m is insensitive to N_{VC}.

C. *Influence of the number N of sources: case of flow-driven model*

We can deal with about ten times larger number of sources than in [11] for the flow-driven model since the number of system states can shrink in this analysis. So, we will show the impact of the number N of sources on performance in Fig. 5.7. In this figure, N is set to 32, 64, 96 and 128 and the x axis shows the ratio of N_{VC} to N. Note that the average arrival rate λ of each source is smaller as N becomes larger since $N\lambda$ is fixed, i.e., set to 10 for each N. Therefore, as N_{VC} increases, R_c becomes smaller and both W_d and R_m get larger when N is smaller, however, from a viewpoint in terms of N_{VC}/N, we have an opposite tendency. This implies that besides the number N_{VC}, the assigned bandwidth of each VC is very important factor for improving performance of the layer 3 switch. Moreover, R_m is in proportion to $R_c(1 - R_c)$ even if N becomes larger.

VI. Concluding Remarks

The cut-through technique on the layer 3 switch has been considered as a way to transmit IP datagrams in the Internet backbone fast. Therefore, in this paper, we have analyzed the layer 3 switch and derived the following three performance measures, namely, the cut-through rate, the average IP datagram waiting time at the routing kernel and the mis-order rate. By using IBP, we can deal with about ten times larger number of sources in case of the flow-driven model than in [11], and also analyze the topology-driven model.

Through some numerical results, we have obtained the following results in case of flow-driven model.

- The mis-order rate R_m is in proportion to $R_c(1 - R_c)$ where R_c indicates the cut-through rate.
- R_c becomes 1 if there is a half of the total numbers of VC and over.

In comparison with the case of topology-driven model, we have obtained some results as follows.

- The number of required VCs achieving that R_c = 1 becomes smaller.
- The average IP datagram waiting time W_d is better approximately 10^3 times.
- The mis-order transmission of datagrams occurs while it does not in the topology-driven model.

The above means that flow-driven model applies VC connection more efficiently than topology-driven one. However, we can say that the topology-driven model is superior to the flow-driven one in respect of the mis-order transmission. When the average arrival rate λ is the same value at each source, the performance is hardly affected by the average IP flow length in both models.

Our study can show that the difference of performance for layer 3 switch. However, in the actual network, each source has different arrival rate and traffic parameters. Furthermore, IETF has recently proposed Multi-Protocol Label Switching (MPLS)[12] as generalized and expanded the layer 3 switching mechanism with some QoS provisioning. Thus, clarifying these topics will be left for further study.

Acknowledgment

This work was supported in part by Research for the Future Program of Japan Society for the Promotion of Science under the Project "Integrated Network Architecture for Advanced Multimedia Application Systems"(JSPS-RFTF97R16301).

References

[1] P. P. White, "ATM Switching and IP Routing Integration: The Next Stage in Internet Evolution?" *IEEE Communications Magazine*, Vol.36, No. 4, pp. 79–83, Apr. 1998.

[2] M. Laubach, "Classical IP and ARP over ATM," RFC2225, Apr. 1998.

[3] "Multi-protocol Over ATM Version-1," af-mpoa-0087.000, July 1997.

[4] P. Newman, T. Lyon and G. Minshall, "Flow Labeled IP: A Connectionless ATM Under IP," *Networld+Interop*, Apr. 1996.

[5] TOSHIBA Corporation, "Cell Switch Router," whitepaper, Nov. 1996.

[6] Y. Rekhter, B. Davie, D. Katz, E. Rosen and G. Swallow, "Cisco Systems' Tag Switching Architecture Overview," RFC2105, Feb. 1997.

[7] A. Viswanathan et al., "ARIS: Aggregate Route-Based IP Switching," *Internet draft*, Mar. 1997.

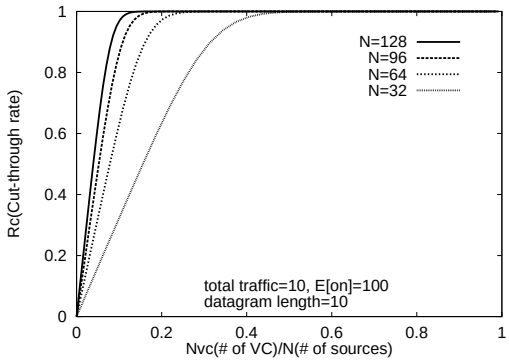
[8] S. Lin and N. McKeown, "A Simulation Study of IP Switching," *ACM SIGCOMM'97*, Sept. 1997.

[9] M. F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, (Marcel Dekker, New York, 1989).

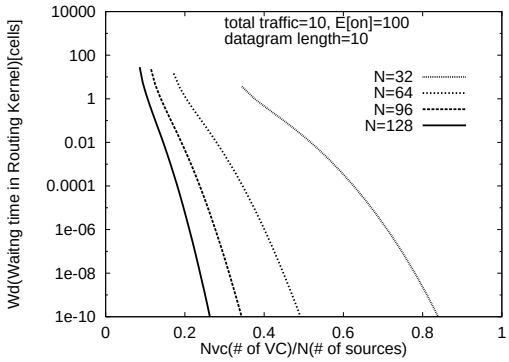
[10] V. Ramaswami, "Stable Recursion for the Steady State Vector in Markov Chains of M/G/1 Type," *Stochastic Models*, vol. 4, pp. 183–188, 1988.

[11] K. Kawahara, S. Nakazawa, T. Takine, Y. Oie, "Performance Analysis of Layer3 switch : Case of Flow-driven Connection Setup," *Internet Workshop '99*, TS1-2, Feb. 1999.

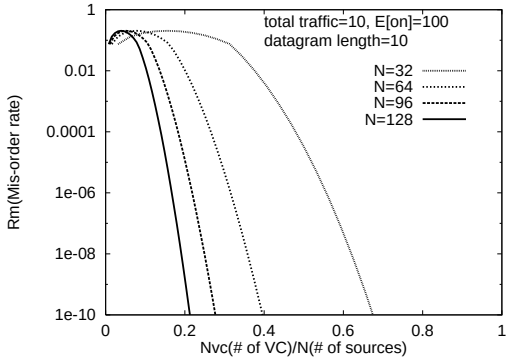
[12] A. Viswanathan, N. Feldman, Z. Wang and R. Callon, "Evolution of Multiprotocol Label Switching" *IEEE Communications Magazine*, Vol.36, No. 5, pp. 165–173, May 1998.



(a) Cut-through rate, R_c



(b) Average datagram waiting time, W_d



(c) Mis-order rate, R_m

Fig. 5.7. Influence of the number of sources, N: Case of Flow-driven model