Performance Modeling and Analysis of IP Switching

J. Zheng and V. O. K. Li, Fellow, IEEE

Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong
Email: {jzheng, vli}@eee.hku.hk

Abstract

IP Switching is a new routing technology proposed to improve the performance of IP routers. In this paper, we study the performance modeling and analysis of IP switching. Our proposed performance models can be used to evaluate the percentage of flows switched and the ratio of the switched path delay to the forwarded path delay in an IP switch. Based on these models, the impacts of different system parameters on the performance are also investigated. Our objective is to develop a design tool for IP switches.

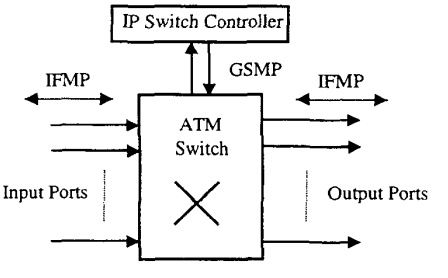
I. Introduction

Due to the explosion of the Internet users as well as emerging multimedia applications, IP traffic has been increasing exponentially for the last few years and IP routers have become the bottleneck of the Internet. To keep pace with the growing demands imposed on IP networks, many new routing technologies have been proposed to improve the performance of IP routers, including IP Switching [1][2], Tag Switching [9], CSR [10] and ARIS [11]. These new technologies share a common label-swapping mechanism that combines the speed of link-level (layer-2) switching with the flexibility of network-level (layer-3) routing, but differ in their ways to bind labels (or tags) to flows. With this label-swapping mechanism, the performance of IP routers can be greatly improved by allowing most datagrams to be switched at the link-level without being forwarded at the network-level. Among all these technologies, IP Switching has received much attention within the IP community. An IP switch integrates the flexibility of IP routing with the speed of ATM switching and thus provides excellent performance. There has already been some research on IP Switching, including system architecture design [1][2], protocol specifications [3][4][5], simulation study [6], laboratory study [7], flow classification algorithms [6][8], etc. Even so, many technical issues remain to be explored. In this paper, we study the performance modeling and analysis of IP switching. Two performance models are proposed to evaluate the performance of an IP switch. The basic model can be used to evaluate the percentage of flows switched and determine the required VC space in an IP switch. The delay model can be used to evaluate the ratio of the switched path delay to the forwarded path delay in an IP switch using the input-queued ATM switch. Based on these models, the impacts of different system parameters on the performance are also investigated.

The remainder of this paper is organized as follows. In Section II, we present our proposed performance models and analyze the system performance based on these models. In Section III, we give some numerical results to investigate the impacts of different system parameters on the performance of an IP switch. In Section IV, we summarize this paper.

II. IP Switching

The general architecture of an IP switch is shown in Fig. 1. It consists of an ATM switch and the IP switch controller. The ATM switch performs ATM switching for the flows selected for switching while the IP switch controller performs IP forwarding for both the newly arrived flows and those flows that are not selected for switching. The IP switch controller also performs flow classification and uses the Ipsilon Flow Management Protocol (IFMP) [3] to send redirect messages to its upstream peers. The control protocol used between the two components is the General Switch Management Protocol (GSMP) [5] that gives the IP switch controller full control of the ATM switch.



IFMP: Ipsilon Flow Management Protocol
GSMP: General Switch Management Protocol

Fig. 1 General Architecture of an IP Switch

At network setup, an IP switch establishes a default ATM virtual circuit (VC) between the local switch and

0-7803-5428-1/99/$10.00©1999 IEEE

each of its neighbor peers. The default channel is used for the forwarding of IP datagrams in the traditional hop-by-hop manner. All datagrams are encapsulated into AAL5 frames and segmented into ATM cells before they are sent on their way. The first datagrams of a flow are sent and received on the default VC. When they arrive at the IP switch, these datagrams are submitted to the IP switch controller through the ATM switch. The IP switch controller routes these datagrams normally as in conventional IP routing and forwards them to their destination through the ATM switch again. Meanwhile, the IP switch controller reassembles these datagrams and performs flow classification. The flow classification determines whether the flow should be switched directly in the ATM hardware or should continue to be forwarded hop-by-hop by the IP switch controller. This can be achieved by inspecting the values of the header fields and making a decision based upon a local policy. Generally, long-duration flows with a large number of datagrams should be selected for ATM switching while short-duration flows with a small number of datagrams should be handled by normal hop-by-hop forwarding. If a flow is selected for ATM switching, the IP switch controller must establish a VC for the flow so that it can be switched directly in the ATM hardware. For this purpose, the IP switch controller will first select a free label or virtual circuit identifier (VCI) from the VC space of the corresponding input port and then send an IFMP redirect message to inform the upstream node of the association between the flow and the label (VCI). If the upstream node accepts the redirect message, it will send all further datagrams of the flow on the new VC. This flow labeling process runs independently and concurrently on each link of the IP switch. When the IP switch receives a redirect message from its downstream node, all further datagrams of the flow can be switched directly from the input port to the output port through the ATM switch, and no longer processed by the IP switch controller. Flows not selected for ATM switching will continue to be forwarded by the IP switch controller hop-by-hop.

III. Performance Models

While there are many important performance metrics for an IP switch, we focus on the percentage of flows switched and the ratio of the switched path delay to the forwarded path delay in this study.

3.1 Basic Model

We first assume that an IP switch has I input ports and all the input ports are identical. Since there are a large number of sources in an IP network and these sources generate flows independently, the flow arrival process to the IP switch on input port i can be modeled as a Poisson process with rate λ_i $(i = 1,2,\cdots,I)$. The first one or few datagrams of a flow will be processed by the IP switch controller for flow classification. Since the delay incurred by this processing is negligible compared to the flow duration and each flow can get served (either forwarded or switched), the IP switch may be thought of as equivalent to an infinite number of virtual servers for flows arriving an input port. The service time of a virtual server for one flow is just the duration of the flow. Considering the variable duration of different flows, we assume that the service time of a virtual server for one flow is exponentially distributed with mean $1/\mu_i$, where μ_i is the mean flow service rate. Therefore, for flows arriving on input port i, the IP switch can be modeled as an $M/M/\infty$ system with arrival rate λ_i and service rate μ_i. Further, the IP switch controller performs flow classification on all flows and decides whether or not they should be switched in the ATM hardware. When a flow is selected for ATM switching, the IP switch controller will request a free VCI from the VC space. Each input port maintains a dedicated VC space. If there are free VCIs available in the VC space, one will be selected to label the flow so that the flow can be switched directly in the ATM hardware. Otherwise, the flow will be blocked for ATM switching and continue to be forwarded hop-by-hop by the IP switch controller. Therefore, the request process to the VC space of input port i can be modeled as a Poisson process with rate $P_s\lambda_i$, where P_s denotes the probability that a flow is selected for ATM switching. The probability P_s depends on the local flow classification policy and can generally be estimated from empirical data. If we take the VC as a virtual server, the service time of a VC for one flow on input port i is approximately equal to the duration of the flow. Accordingly, we assume that it is also exponentially distributed with mean $1/\mu_i'$, where μ_i' is the mean flow service rate of a VC. Generally, μ_i' is less than μ_i as the mean duration of the flows selected for ATM switching is longer than that of all the flows on input port i. Therefore, the VC space of input port i can be modeled as an $M/M/n/n$ loss system with arrival rate $P_s\lambda_i$ and service rate μ_i', where n is the VC space size.

The queuing model for flows arriving on input port i is shown in Fig. 2, which we refer to as the basic model. From this model, we can evaluate the average number of flows N_i, the probability that a flow is blocked for ATM switching P_{iB} and the average number of flows switched N_{is} on input port i, i.e.

$$N_i = \frac{\lambda_i}{\mu_i},$$

$$P_{iB} = \frac{1}{n!}\left(\frac{\lambda_i}{\mu_i'}\right)^n P_s^n p_{i0}$$

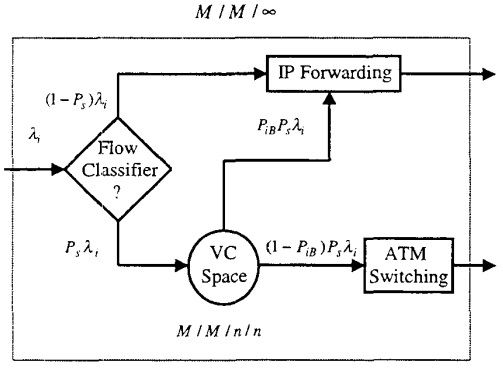
$$N_{is} = \sum_{k=1}^{n} \frac{1}{(k-1)!}\left(\frac{\lambda_i}{\mu_i'}\right)^k P_s^k p_{i0}$$

where

$$p_{i0} = \left[\sum_{k=0}^{n} \frac{1}{k!}\left(\frac{\lambda_i}{\mu_i'}\right)^k P_s^k\right]^{-1}$$

Therefore, the percentage of flows switched on input port i can be calculated as follows

$$\alpha_i = \frac{N_{is}}{N_i} = [\sum_{k=1}^{n} \frac{P_s^k}{(k-1)!} (\frac{\lambda_i}{\mu_i})^k p_{i0}] / (\frac{\lambda_i}{\mu_i})$$



λ_i: the flow arrival rate on input port i

P_s: the probability that a flow is selected for ATM switching

P_{iB}: the probability that a flow is blocked for ATM switching

Fig. 2 Basic Model

3.2 Delay Model

The datagrams of a flow may take two different paths in an IP switch. The first datagrams of the flow take the forwarded path. They arrive at the IP switch through the default VC and then are submitted to the IP switch controller for forwarding and classification through the ATM switch. After being processed by the IP switch controller, they will be switched to an output port through the ATM switch again. If the flow is selected for ATM switching, further datagrams of the flow will take the switched path and will be switched directly from the input port to the output port in the ATM hardware. To develop a performance model for the evaluation of the delay that a datagram takes to go through an IP switch, we assume that the IP switch uses the input-queued ATM switch.

To simplify the analysis, the datagram arrival process of each flow is approximated as a Poisson process and the total datagram arrival process on one input port may also be approximated as a Poisson process as it is a merging of a random number of Poisson processes. According to the results in Section 3.1, the average number of flows on input port i is N_i, in which N_{if} flows take the forwarded path and N_{is} flows take the switched path. If we use γ_i to denote the mean datagram arrival rate of all the flows on input port i and γ_i' to denote the mean datagram arrival rate of one single flow, we have $\gamma_i = N_i \gamma_i'$. In IP switching, a datagram is segmented into a number of ATM cells for transmission. These ATM cells are sent and received consecutively. Each datagram may contain a different number of cells. Due to the variable lengths of different datagrams, we assume that the service time of the ATM switch for one datagram is exponentially distributed with mean $1/\eta_A$, where η_A is the mean datagram service rate. Therefore, for datagrams arriving on input port i, the ATM switch can approximately be modeled as an $M/M/1/S+1$ [12] delay system with arrival rate γ_i and service rate η_A, where S is the buffer size of the input port in terms of the number of datagrams and is generally determined by the desired datagram blocking probability P_{isB} at the input port. In reality, however, a datagram is buffered in terms of cells. Therefore, we need to find n_b, the buffer size in terms of number of cells, which also satisfies the desired blocking probability P_{isB}. Since the service time for one datagram is approximately supposed to be exponentially distributed, it can be proved that the number of cells corresponding to S datagrams (n_S) is a Pascal distribution [13], i.e.

$$P(n_S = k) = \binom{k-1}{S-1} p^S (1-p)^{k-S} \qquad k \geq S$$

and $$p = p\{t_d \leq t_c\} = 1 - e^{\eta_A t_c}$$

where t_d is the service time for one datagram and t_c is the service time for one single cell. Therefore, n_b, which ensures that the desired datagram blocking probability at an input port is satisfied with probability $(1-\delta)$, can be calculated as follows

$$\delta = P(n_S > n_b) = \sum_{k=n_b+1}^{\infty} P(n_S = k) \qquad n_b \geq S$$

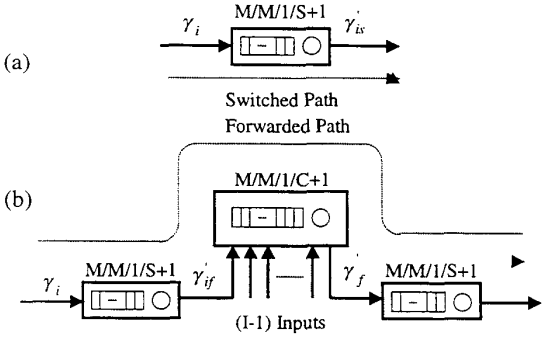


Fig. 3 Delay Model

The queuing model for the switched path is shown in Fig. 3 (a), in which $\gamma_{is}' = (1-P_{isB})\gamma_{is}$ and γ_{is} is the mean datagram arrival rate of flows taking the switched path on input port i. From this model, we can evaluate the switch path delay, i.e. the delay for a datagram to cross the ATM switch from the input port to the output port

$$D_s = \frac{1 + [\sum_{k=1}^{S+1} (k-1) a_s^k p_{s0}] - p_{s0}}{\gamma_i [1 - a_s^{S+1} p_{s0}]}, \qquad a_s = \frac{\gamma_i}{\eta_A}$$

where $p_{s0} = \begin{cases} [1-a_s]/[1-a_s^{S+2}] & a_s \neq 1 \\ 1/(S+2) & a_s = 1 \end{cases}$

For the IP switch controller, the flows come from all the input ports of the ATM switch. Since the departure process of an $M/M/1/S+1$ system is approximately a Poisson process when S is very large or the ratio of γ_i to η_A is very small, the datagram arrival process to the IP switch controller from each input port can be approximated as a Poisson process. Accordingly, the total datagram arrival process to the IP switch controller can also be approximated as a Poisson process as it is a merging of I Poisson processes. Since only an average number of N_{if} flows take the forwarded path, the mean datagram arrival rate from input port i is $\gamma_{if}' = (1-P_{isB})\gamma_{if}$, where P_{isB} is the probability that a datagram is blocked at input port i and γ_{if} is the mean datagram arrival rate of flows taking the forwarded path. Therefore, the mean datagram arrival rate to the IP switch controller is thus $\gamma_f = \gamma_{1f}' + \gamma_{2f}' + \cdots + \gamma_{If}'$. Similarly, we assume that the service time of the IP switch controller for one datagram is exponentially distributed with mean $1/\eta_f$, where η_f is the mean datagram service rate. Therefore, the IP switch controller can approximately be modeled as an $M/M/1/C+1$ delay system, where C is the buffer size of the IP switch controller in terms of the number of datagrams.

The datagrams leaving the IP switch controller will be switched again to their respective output ports in the ATM hardware. Similarly, the datagram departure process can be approximated as a Poisson process and the mean rate becomes $\gamma_f' = (1-P_{fB})\gamma_f$, where P_{fB} is the probability that a datagram is blocked at the IP switch controller. Therefore, for datagrams leaving the IP switch controller, the ATM switch can also be modeled as an $M/M/1/S+1$ delay system with arrival rate γ_f' and service rate η_A.

The queuing model for the forwarded path delay is shown in Fig. 3 (b). From this model, we can evaluate the delay for a datagram to be switched to the IP switch controller

$$T_{s1} = \frac{1+[\sum_{k=1}^{S+1}(k-1)a_{s1}^k p_{s10}] - p_{s10}}{\gamma_i[1-a_{s1}^{S+1}p_{s10}]}, \qquad a_{s1} = \frac{\gamma_i}{\eta_A}$$

and the probability that a datagram is blocked at the input port of the ATM switch

$$P_{isB} = a_{s1}^{S+1} p_{s10}$$

where $p_{s10} = \begin{cases} [1-a_{s1}]/[1-a_{s1}^{S+2}] & a_{s1} \neq 1 \\ 1/(S+2) & a_{s1} = 1 \end{cases}$

Similarly, the delay for a datagram to go through the IP switch controller is given by

$$T_f = \frac{1+[\sum_{k=1}^{C+1}(k-1)a_f^k p_{f0}] - p_{f0}}{\gamma_f[1-a_f^{C+1}p_{f0}]}, \qquad a_f = \frac{\gamma_f}{\eta_f}$$

and the probability that a datagram is blocked at the IP switch controller is

$$P_{fB} = a_f^{C+1} p_{f0}$$

where $p_{f0} = \begin{cases} [1-a_f]/[1-(a_f)^{C+2}] & a_f \neq 1 \\ 1/(C+2) & a_f = 1 \end{cases}$

The delay for a datagram leaving the IP switch controller to be switched to its output port is given by

$$T_{s2} = \frac{1+[\sum_{k=1}^{S+1}(k-1)a_{s2}^k p_{s20}] - p_{s20}}{\gamma_f'[1-a_{s2}^{S+1}p_{s20}]}, \qquad a_{s2} = \frac{\gamma_f'}{\eta_A}$$

where $p_{s20} = \begin{cases} [1-a_{s2}]/[1-(a_{s2})^{S+2}] & a_{s2} \neq 1 \\ 1/(S+2) & a_{s2} = 1 \end{cases}$

With the above results, we have the forwarded path delay

$$D_f = T_{s1} + T_f + T_{s2}$$

and the ratio of the switched path delay to the forwarded path delay $\beta = \frac{D_s}{D_f} = \frac{T_{s1}}{T_{s1}+T_f+T_{s2}}$

IV. Numerical Results

In this section, we give some numerical results to investigate the impacts that different parameters have on the system performance of an IP switch. For simplicity, we assume that $\lambda_i = \lambda$, $\mu_i = \mu$, $\mu_i' = \mu'$ and $\gamma_i' = \gamma$.

The impact of n on α_i is shown in Fig. 4. When λ_i is very small, n has no impact on α_i since no flow is blocked for ATM switching. With λ_i increasing, α_i drops quickly because of the increased blocking probability. As expected, α_i increases as n increases. This is because a larger VC space size will decrease the blocking probability and thus increase α_i.

The impact of η_A and η_f on β is shown in Fig. 5, in which we let $\xi = \eta_A/\eta_f$. A larger value of ξ produces a smaller value of β. This is because, when η_f is fixed, a larger value of ξ means a larger value of η_A. When η_A is larger, less time is required for a datagram to go through the ATM switch. This results in a smaller value of β. When λ is very small, ξ has less impact on β. This is because little queuing delay is incurred in this case and the queuing delay is largely affected by η_A. Furthermore, β varies as λ increases. When λ is very small, β drops dramatically as λ increases. This is because the slow processing speed of the IP switch controller increases the queuing delay at the IP switch controller quickly while the queuing delay at the input port remains almost unchanged.

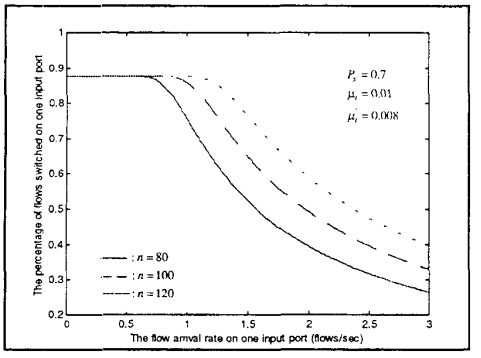


Fig. 4 Impact of n on α_i

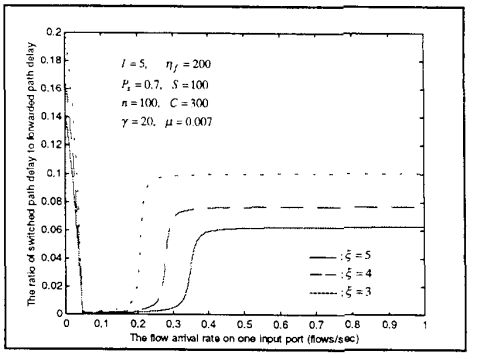


Fig. 5 Impact of ξ on β

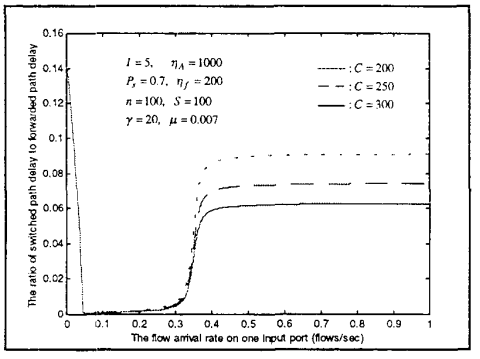


Fig. 6 Impact of C on β

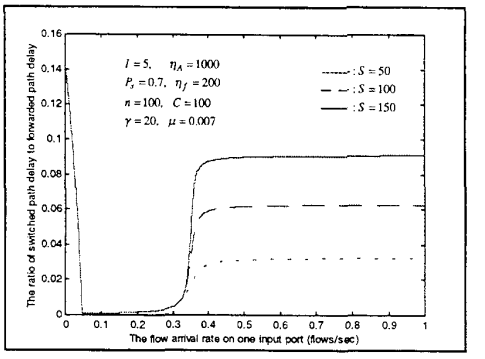


Fig. 7 Impact of S on β

When λ reaches a certain value, the buffer of the IP switch controller will become full, keeping the queuing delay at the IP switch controller relatively unchanged and β very small within a certain range of λ. With λ increasing further, the queuing delay at the input port will begin to increase and β will thus increase sharply. Similarly, when the buffer of the input port also becomes full, the queuing delay at the input port and β will become relatively unchanged again.

The impacts of S and C on β are illustrated in Fig. 6 and Fig. 7. A larger value of S causes a larger value of β while a larger value of C causes a smaller value of β. This is because a larger buffer size will result in a longer queuing delay, which will further affect β. When λ is small, β cannot benefit much from the buffer size since there is almost no queuing delay at the input port. Only when λ increases beyond a certain value will the buffer size produce any impact on β.

V. Conclusion

In this paper, we developed two performance models to analyze the performance of IP switching. The basic model can be used to evaluate the percentage of flows switched in an IP switch and the delay model can be used to evaluate the ratio of the switched path delay to the forwarded path delay in an IP switch using the input-queued ATM switch. Given the desired blocking probability that a flow is blocked for ATM switching, the basic model can also be used to determine the required VC space in an IP switch. To simplify the analysis, we assumed Possion arrival processes and exponentially distributed service times. The validity of these assumptions is as yet unjustified. For this reason, the performance models proposed in this paper are only approximate models. Further work is under way to improve these two performance models.

Reference

[1] Peter Newman, Greg Minshall and Tom L. Lyon, "IP Switching—ATM under IP," IEEE/ACM Transactions on Networking, Vol. 6, No. 2, April 1998, pp. 117-129.

[2] P. Newman, *et al.*, "IP Switching and Gigabit Routers," IEEE Communication Magazine, Jan. 1997, pp. 64-69.

[3] P. Newman, *et al.*, "Ipsilon Flow Management Protocol Specification for IPv4," IETF RFC 1953, May 1996.

[4] P. Newman, *et al.*, "Transmission of Flow Labelled IPv4 on ATM Data Links," IETF RFC 1954, May 1996.

[5] P. Newman, *et al.*, "Ipsilon's General Switch Management Protocol Specification," IETF RFC 1987, Aug. 1996.
[6] Steven Lin and Nick Mckeown, "A Simulation Study of IP Switching," ACM SIGCOMM' 97, pp.15-24.
[7] J. K. Lucek, *et al.*, "Laboratory Study of IP Switching," Proceedings of the 1997 IEE Colloquium on IP Routing Versus ATM Switching-What are the real issues, Nov. 1997, n 334, London, pp. 8/1-8/4.
[8] Hao Che, San-qi Li and Arthur Lin, "Adaptive Resource Management for Flow-Based IP/ATM Hybrid Switching Systems," IEEE/ACM Transactions on Networking, Vol. 6, No. 5, Oct. 1998, pp. 544-557.
[9] Y. Rekhter *et al.*, "Cisco Systems' Tag Switching Architecture Overview," IETF RFC 2105, Feb. 1997.
[10] S. Matsuzawa, *et al.*, "Architecture of Cell Switch Router and Prototype System Implementation," IEICE Transaction on Communication, Vol. E80-B, No.8, Aug. 1997, pp.1227-1237.
[11] R. Woundy, "ARIS: Aggregate Route-Based IP Switching," IETF Internet Draft, draft-woundy-aris-ipswitching-00.txt, Nov. 1996
[12] Donald Gross and Carl M. Harris, "Fundamentals of Queuing Theory," John Wiley & Sons, 1998.
[13] Alvin W. Drake, "Fundamentals of Applied Probability Theory," McGraw-Hill, 1967.