A DELAY MODEL FOR A FRAME RELAY SWITCH *

L. Po, P. Sarachik and S. Panwar
Polytechnic University
Center for Advanced Technology in Telecommunications
333 Jay Street
Brooklyn, N.Y. 11201

Abstract

In this paper, we model a frame relay switch using a cyclic server queuing system for the switch module. We present two improved methods for the approximation of the waiting time delay for the cyclic server queuing system. The four traffic parameters (the first and second moments of the packet inter-arrival time and packet length) at the output of the switch are derived, thus the QNA method can be used for the network analysis.

1 Introduction

Frame relay service is an ISDN packet-mode bearer service for data communications. The frame relay switch performs the core function of the LAPD data link layer protocol, including frame validation, error detection, virtual circuit multiplexing/demultiplexing and switching. In a frame relay network, the full LAPD protocol is performed only at the end nodes. Intermediate nodes merely detect errors and perform the switching function for transit packets. Many papers have discussed the architecture and the application of frame relay networks (Chen *et al.* 1989; Marsden 1991; Cole 1988). The modeling of the frame relay switch should be based on this two layer functionality.

A good model for a frame relay switch is important for network implementation. It can be used to analyze and predict the packet delay and packet loss performance of a network, and in routing Permanent Virtual Circuits (PVCs).

Modeling the switching functionality (i.e., the switch module) is the core of the frame relay switch model. For the class of frame relay switches we considered, the switch module could be modeled as a non-exhaustive cyclic service queueing system as was discussed in (Cole 1988). Many papers have contributed to modeling non-exhaustive cyclic service queues. Only approximate results are available.

A basic approach we used in the modeling is the QNA network analysis method (Whitt 1983), which can analyze the delay performance of GI/G/m queues and networks. In the QNA method, the nodes are analyzed as GI/G/m queues characterized by the first moments and the squared coefficients of variation (a variability parameter) of the packet inter-arrival times and service times. Traffic streams, each with those four parameters, arriving to a node are combined and form the combined traffic parameters to the node. The service discipline is first-in first-out. The approximate mean waiting time of the node can then be calculated.

Departure traffic parameters at the output of a given node are estimated, and these are input parameters to the following nodes. If there is no loss in the queueing system, the packet rate, the mean packet length and the variability parameter of the packet length are the same as those of the input traffic parameters, but the variability parameter of the packet inter-departure time is changed. This can be estimated by Marshall's formula (Marshall 1968).

If the departure traffic is split into many traffic streams after the queue service, the QNA method also gives a way to calculate the parameters for each of the split traffic streams.

A frame relay switch can be modeled as a multi-queue cyclic service system. The QNA method cannot solve this type of queue. In this paper, we discuss a systematic analysis for the delay performance of a frame relay switch and present a method to estimate the traffic parameters at the output of the switch, which are the input parameters to the downstream switches.

In the second and third sections of this paper, we describe the switch model and the input traffic parameters. The delay analysis of the input module and the output module of a frame relay switch is discussed in sections 4 and 6. In section 5, we discuss the delay models for the switch module, which is basically a cyclic service queue system.

*This work was supported in part by NYNEX Science and Technology and by the New York State Foundation for Science and Technology as part of its Centers for Advanced Technology Program.

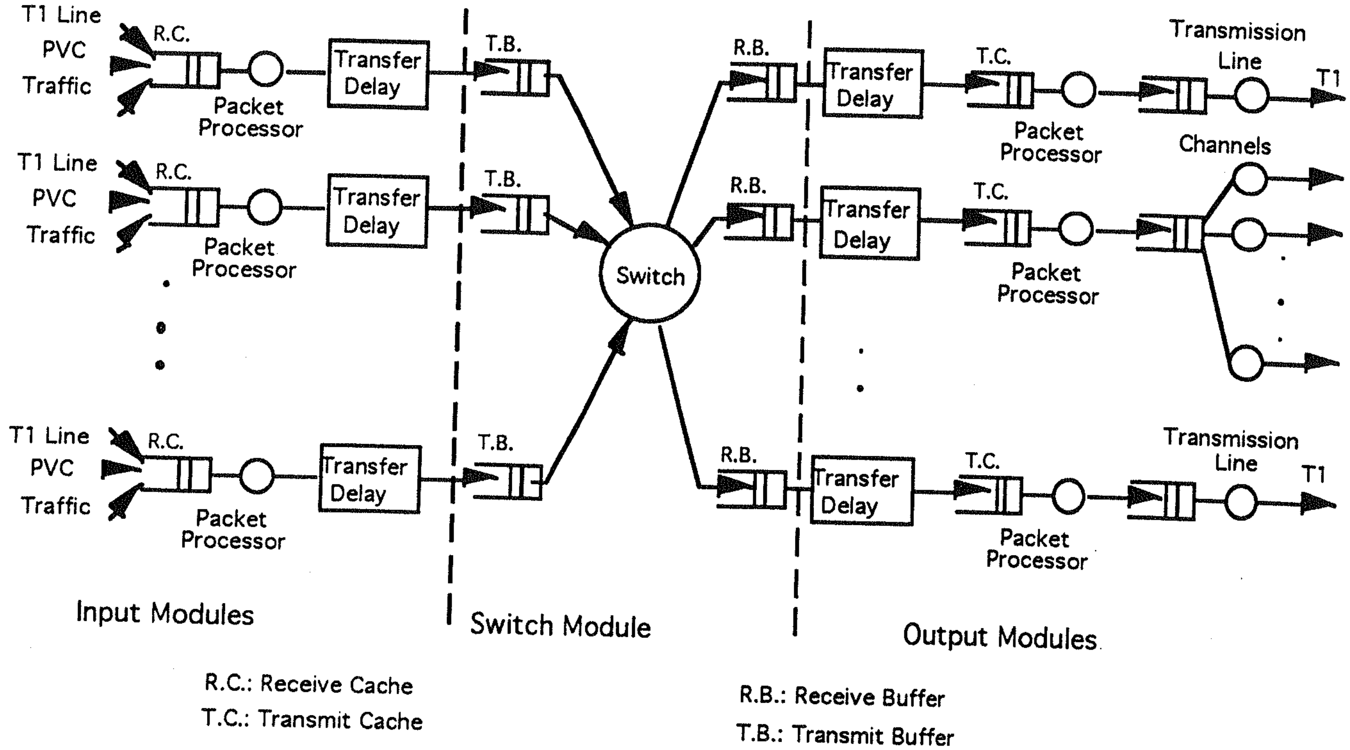

Fig. 1 Equivalent Model for the Frame.Relay Switch

2 Model Description

A typical frame relay switch consists of three parts: an input, a switch and an output module. An input packet (or a frame of data) from a customer premises equipment (CPE) access line or a trunk line to the switch, is detected and verified by the packet processor in the input module. Valid packets are forwarded to the transmit buffer and wait for the switch server to switch them to their destination output module. Packets arriving at the output module are stored in the transmit cache for the packet processor to perform the packet validation. They are then multiplexed and transferred to the output line. All input and output lines are T1 lines, which can be of three different types; 24 channelized 56kbps channels, four partially channelized 384kbps channels or a single 1.344 Mbps channel. The switch discipline is round robin and at most one packet is switched when a queue is polled. A queuing model for the frame relay switch is shown in Fig.1. There are three queues which must be modeled; the input queue, the switch queue and the output queue. There are also some packet transfer delays in or between the modules which are proportional to the packet size and can be placed anywhere in the queueing model. For simplicity, we place all transfer delays which occur before the switch module into the input module and the transfer delays which occur after the switch module into the output module. The transfer delays at the input and output modules are assumed to be equal, because similar data transfers are necessary before and after the switch module.

3 Input Traffic Description

Traffic to a switch node either comes from a previous node through a T1 trunk or enters at the node through a fully or partially channelized or a full T1 line from CPEs. A source CPE sends its traffic through the network via a Permanent Virtual Circuit (PVC) to the destination CPE. PVC k (denoted by PVC_k), which originates at input line (or channel) i of a node and goes to output line (or channel) j, has four parameters: the packet arrival rate α_k, the squared coefficient of variation of packet inter-arrival times $v_{A_k}^2$, the mean packet length $E\{L_k\}$ and the squared coefficient of variation of the packet length $v_{L_k}^2$. Those four parameters are observed before the traffic is transmitted to the input module, hence, when many PVCs exist in an input line, the input traffic is the PVCs' combined traffic with four parameters: the packet arrival rate λ_{1i}, the squared coefficient of variation of the packet inter-arrival time $v_{T_{1i}}^2$, the mean packet length $E\{L_{1i}\}$ and the squared coefficient of variation of the packet length $v_{L_{1i}}^2$. A simple way to estimate the parameters of the combined traffic is by the QNA method (Whitt 1983).

If an input line is a trunk or a full T1 access line, all PVCs coming through the line are merged in the line with a line service rate of 1.344 Mbps.

When PVCs come through channel m of input module i, similarly, the combined four parameters, λ_{1im}, $E\{L_{1im}\}$, $v^2_{T_{1im}}$ and $v^2_{L_{1im}}$ for that channel can be found by considering each channel as an input line with the service rate of the channel. The channel traffic streams are further combined and form the traffic of input module i. Thus, there are two merges for the input traffic through channelized channels.

4 Delay Model of the Input Module

An input module i has a packet processor, which processes each incoming packet with a constant processing time $E\{X_{1i}\}$. The load on the queue is $\rho_{1i} = \lambda_{1i}E\{X_{1i}\}$. The waiting time of the packet processor is (Whitt 1983)

$$E\{W_{1i}\} = g_{1i}\frac{\rho_{1i}E\{X_{1i}\}(v^2_{T_{1i}} + v^2_{X_{1i}})}{2(1-\rho_{1i})} \tag{1}$$

where $g_{1i} = \exp\left(-\frac{2(1-\rho_{1i})(1-v^2_{T_{1i}})^2}{3\rho_{1i}(v^2_{T_{1i}} + v^2_{X_{1i}})}\right)$ if $v^2_{T_{1i}} < 1$ and $g_{1i} = 1$ if $v^2_{T_{1i}} > 1$ and the variability parameter $v^2_{X_{1i}} = 0$ in this case.

The transfer delay is proportional to the packet size. Let the transfer rate be R_t, which depends on the switch fabric and can be measured or calculated on a real system. Because the transfer rate is very high, there is no queuing delay but there is a delay in addition to the packet delay. If there are M transfers in the input module, the total input transfer delay is ML/R_t, where L is the packet length.

The packet departure rate and the packet length parameters after the processor are the same as for the input to the processor, but the squared coefficient of variation of the inter-arrival time has been changed. By the QNA method, the variability parameter, $v^2_{T_{2i}}$, is estimated by

$$v^2_{T_{2i}} \approx (1-\rho^2_{1i})v^2_{T_{1i}}. \tag{2}$$

The four parameters for the input to the switch are then determined.

5 The Switch Module

The switch module is modeled as a non-exhaustive cyclic queue with at-most-one packet served per polling of the queue. In this part, we combine some basic methods and try to develop a better delay model. We first discuss the switch model for Poisson input traffic and then modify it for general input traffic.

5.1 The Conditional Cycle Times

Define C''_i to be the conditional cycle time random variable for queue i, given that a packet from queue i is served in the cycle; C'_i to be the conditional cycle time random variable for queue i given that no packet from queue i is served in the cycle.

Kuehn 1979 had an approximation method for mean cyclic times, $c''_i = E\{C''_i\}$ and $c'_i = E\{C'_i\}$. Let $c''_{\sim i}$ and $c'_{\sim i}$ be Kuehn's approximate values, then

$$c''_{\sim i} = \frac{s_0 + h_i}{1-\rho_0+\rho_i} \tag{3}$$

$$c'_{\sim i} = \frac{s_0}{1-\rho_0+\rho_i} \tag{4}$$

where $\rho_i = \lambda_{2i}h_i$ is the load contributed to the switch by the queue i traffic and $\rho_0 = \sum_{i=1}^{N}\rho_i$ is the total switch load. λ_{2i} is the packet arrival rate to the switch input queue i, which is equal to λ_{1i}, if there is no packet loss at the input module; h_i is the average packet service time, which is equal to the packet length $E\{L_{1i}\}$ over the switch rate R_w; and s_0 is the sum of the mean polling times s_i for each queue i.

Equations (3) and (4) give extreme approximations for the conditional cycle times. If $c_0 = s_0/(1-\rho_0)$ is the average cycle time, the relationship among the cycle times is

$$c''_{\sim i} \geq c''_i \geq c_0 \geq c'_i \geq c'_{\sim i}. \tag{5}$$

We can find another approximation method for c''_i and c'_i. Let $c''_{\simeq i}$ and $c'_{\simeq i}$ be these approximate values where

$$c''_{\simeq i} = s_0 + h_i + \sum_{j\neq i}\delta''_j h_j \tag{6}$$

$$c'_{\simeq i} = s_0 + \sum_{j\neq i}\delta'_j h_j. \tag{7}$$

Here $\delta''_j = \lambda_{2j}c''_{\simeq j}$ is the probability that a packet from queue j is served in the cycle, given that a packet from queue j is served in the previous cycle; $\delta'_j = \lambda_{2j}c'_{\simeq j}$ is the probability that a packet from queue j is served in the cycle, given that no packet from queue j is served in the previous cycle.

The solutions to the linear equations (6) and (7) are

$$c''_{\simeq i} = \frac{s_0 + h_i(1-d_1) + d_3}{1-\rho_0+\rho_i(1-d_1)+d_2} \tag{8}$$

$$c'_{\simeq i} = \frac{s_0}{1-\rho_0+\rho_i(1-d_1)+d_2} \tag{9}$$

where $d_1 = \sum_{j=1}^{N}\frac{\rho_j}{1+\rho_j}$; $d_2 = \sum_{j=1}^{N}\frac{\rho_j^2}{1+\rho_j}$ and $d_3 = \sum_{j=1}^{N}\frac{h_j\rho_j}{1+\rho_j}$. Equations (8) and (9) also are upper and lower bound approximations for the conditional cycle times.

By combining Kuehn's approximations and our approximations, a better approximation for the conditional cycle times can be acquired. Let $\epsilon''_{ij} = \min(\lambda_{2i}c''_{\sim j}, \lambda_{2i}c''_{\simeq j}, \delta''_i, 1)$ and $\epsilon'_{ij} = \min(\max(\lambda_{2i}c'_{\sim j}, \lambda_{2i}c'_{\simeq j}, \delta'_i), 1)$, which gives a better approximation for the conditional probabilities that a packet from queue j is served in a cycle, given that a packet from queue i is served in the cycle or given that no packet from queue i is served in the cycle. The approximations for the cycle times are

$$c''_{\simeq i} = s_0 + h_i + \sum_{j \neq i} \epsilon''_{ij} h_j \quad (10)$$

$$c'_{\simeq i} = s_0 + \sum_{j \neq i} \epsilon'_{ij} h_j \quad (11)$$

which give better approximations for the conditional cycle times, and can improve the queuing delay approximation.

Let $c''^{(2)}_{\simeq i}$ and $c'^{(2)}_{\simeq i}$ be the second moments of these conditional cycle times, then (Kuehn 1979)

$$c''^{(2)}_{\simeq i} = \sum_{j=1}^{N}(s_j^{(2)} - s_j^2) + h_i^{(2)} - h_i^2 + \sum_{j \neq i}(\epsilon''_{ij} h_j^{(2)} - \epsilon''^2_{ij} h_j^2) + c''^2_{\simeq i} \quad (12)$$

$$c'^{(2)}_{\simeq i} = \sum_{j=1}^{N}(s_j^{(2)} - s_j^2) + \sum_{j \neq i}(\epsilon'_{ij} h_j^{(2)} - \epsilon'^2_{ij} h_j^2) + c'^2_{\simeq i} \quad (13)$$

where $s_j^{(2)}$ is the second moment of the polling time and $h_i^{(2)}$ is the second moment of the packet service time for the ith queue of the switch, which is the second moment of L_{1i}/R_w.

In the above discussion, we assume that the switch is stable. For unstable systems, an additional step must be taken to identify the unstable queues and treat them differently.

5.2 The Mean Waiting Time of The Cyclic Queues

The mean waiting time of Poisson input traffic queuing systems depends on the mean residual time and the traffic load. A packet upon arrival to an input queue will find either that there is a head-of-line (HOL) packet or there is none. It sees different residual times in those two cases. Based on the renewal theory and M/G/1 queue theory, we can derive Kuehn's formula for the mean waiting time for queue i,

$$E\{W_{2i}\} = \frac{c_i'^{(2)}}{2c_i'} + \frac{\lambda_{2i} c_i''^{(2)}}{2(1 - \lambda_{2i} c_i'')}. \quad (14)$$

Kuehn used his approximations for the first and second moment of the cycle times and had a delay approximation formula.

Another method developed by Boxma and Meister (1986) is based on two assumptions:
1. $p_i = \lambda_{2i} c''_{\sim i}$ is the utilization observed at queue i;
2. All arrival packets see approximately the same mean residual time r.
The second assumption is good if the traffic load is light or the queues are not very unbalanced, otherwise a large error exists. Based on those assumptions, the mean waiting time is approximated by

$$E_{BM}\{W_{2i}\} = \frac{r}{1 - \lambda_{2i} c''_{\sim i}} \quad (15)$$

where r is a parameter to be determined. Boxma and Meister used the conservation law, which was first developed by Watson (1984), to evaluate r. This gives r and the waiting time approximation for queue i as

$$r \approx \frac{1 - \rho_0}{(1 - \rho_0)\rho_0 + \sum_{i=1}^{N} \rho_i^2} C_{NE} \quad (16)$$

$$E_{BM}\{W_{2i}\} \approx \frac{1 - \rho_0 + \rho_i}{1 - \rho_0 - \lambda_{2i} s_0} r \quad (17)$$

where

$$C_{NE} = \rho_0 \frac{\sum_{i=1}^{N} \lambda_{2i} h_i^{(2)}}{2(1 - \rho_0)} + \rho_0 \frac{s_0^{(2)}}{2s_0} + \frac{s_0}{2(1 - \rho_0)}(\rho_0 + \sum_{i=1}^{N} \rho_i^2).$$

Equation (17) gives a closed form formula for the queue waiting time approximation which, for moderate switch loads and slightly unbalanced queues, gives more accurate results than the approximation for (14) found by Kuehn's method. Later, some numerical results will show that when the switch load is high and queues are very unbalanced (the load of the maximally loaded queue is more than twice the load of the minimally load queue), large errors can exist, which are even worse than Kuehn's approximation. The errors produced are due to the errors of the two assumptions. If either one could be improved, the errors would be reduced.

The first method is to use $c''_{\simeq i}$ as the approximation for the conditional cycle time c''_i instead of $c''_{\sim i}$ in (15), then the residual time and the mean waiting time are approximated by

$$r_0 \approx \frac{(1 - \rho_0) C_{NE}}{\sum_{i=1}^{N} \frac{\rho_i (1 - \rho_0 - \lambda_{2i} s_0)}{1 - \lambda_{2i} c''_{\simeq i}}} \quad (18)$$

$$E_1\{W_{2i}\} \approx \frac{r_0}{1 - \lambda_{2i} c''_{\simeq i}}. \quad (19)$$

Equation (19) gives a better approximation than (17), especially when the switch load is high and queues are highly unbalanced, due to a more accurate approximation for the mean cycle time c''_i.

A second method is to use equation (14) as the approximation for the mean waiting time by substituting c''_i, c'_i, $c''^{(2)}_i$ and $c'^{(2)}_i$ with $c''_{\simeq i}$, $c'_{\simeq i}$, $c''^{(2)}_{\simeq i}$ and $c'^{(2)}_{\simeq i}$. Let the waiting time be $E_K\{W_{2i}\}$. For different queues, we can have different residual times instead of approximating all by r as in Boxma-Meister's method. Let $E_2\{W_{2i}\}$ be the approximation of the mean waiting time by this method, then

$$E_2\{W_{2i}\} = b \cdot E_K\{W_{2i}\} \quad (20)$$

where b is a parameter to be determined for matching the job estimated by $E_2\{W_{2i}\}$ with the total job calculated by the

Load	Simulation	Method 1	Method 2	BM	Kuehn
0.2	.4047(.012)	.4077(.74)	.4135(2.1)	.4073(.64)	.3860(-4.6)
0.4	.9959(.029)	.9975(.16)	1.026(3.0)	.9932(-.27)	.8683(-13.)
0.6	2.548(.068)	2.479(-3.)	2.562(.54)	2.436(-4.4)	1.941(-3.1)
0.8	11.13 (.25)	10.40 (-3.)	10.62(-2.)	9.602(-10.)	6.917(-35.)
0.85	22.18 (.65)	21.62(-2.5)	21.94(-.1)	18.82(-15.)	13.43(-39)
0.9	139.9(23)	163.2 (17.)	164.4(18.)	95.7(-32)	91.7(-34)

Table 1: Waiting Times for the First Queue

Load	Simulation	Method 1	Method 2	BM	Kuehn
0.2	.3814(.013)	3.800(-.5)	.3766(-1.3)	.3801(-.34)	.3576(-6.2)
0.4	.8225(.031)	.8383(1.9)	.8219(-.07)	.8411 (2.3)	.7282(-11)
0.6	1.651(.044)	1.735(5.1)	1.692(2.48)	1.768(6.7)	1.421(-16.)
0.8	3.469(.086)	4.325(25.)	4.272(23.5)	4.935(42.)	2.800 (8.7)
0.85	4.255(.038)	5.959(40.)	5.931(39.0)	7.790(83.1)	5.856(37.6)
0.9	5.335(.056)	9.006(69.)	9.069(70.)	22.83(328.)	11.88(123.)

Table 2: Waiting Times for the Second Queue

conservation law.

$$b = \frac{C_{NE}}{\sum_{i=1}^{N} \rho_i (1 - l_{2i} c_0) E_K\{W_{2i}\}}. \quad (21)$$

When b is found, the mean waiting time $E_2\{W_{2i}\}$ is given by equation (20). In most situations, the second method gives better approximations than the other methods.

Tables 1, 2, 3 and 4 show some numerical results for delay in a 4 queue system with constant polling times $s_i = 0.05$ and exponential packet service times with means $h_i = 1$. The arrivals to the queues are Poisson with average rates $l_1 = 2l_2 = 4l_3 = 4l_4$. These values were chosen to yield an unbalanced system for which the load at queue 1 is twice that at queue 2 and four times the load at queues 3 and 4. Tables 1, 2 and 3 give the waiting time of the first queue, the second queue, and the third and fourth queues respectively. The first column is the total switch load. Simulation results from a special purpose simulation program are given in the second column, with the 95th percentile confidence interval range shown in parentheses. The mean waiting time delay for each method and their percentage errors (compared with the simulation results) are shown in the remaining columns. The results show that improved methods one and two give better waiting time approximations than both Boxma-Meister's method and Kuehn's method. The difference between method 1 and 2 is not large.

In the above discussion, the mean waiting times of the cyclic server queues are for Poisson arrival traffic. If the input traffic is general with two parameters, l_{2i} and $v^2_{T_{2i}}$, the formula for the waiting time must be modified. According to the GI/G/1 theorem (Whitt 1983), the mean waiting time, $E_G\{W_{2i}\}$, for general input traffic can be approximated by

$$E_G\{W_{2i}\} = g_{2i}\left(\frac{c''_{\cong i}\rho_{2i}(v^2_{T_{2i}} - 1)}{2(1 - \rho_{2i})} + E_2\{W_{2i}\}\right) \quad (22)$$

where $g_{2i} = \exp\left(-\frac{2(1-\rho_{2i})(1-v^2_{T_{2i}})^2}{3\rho_{2i}(v^2_{T_{2i}} + v^2_{L_{2i}})}\right)$ if $v^2_{T_{2i}} < 1$ and $g_{2i} = 1$ if $v^2_{T_{2i}} > 1$; $\rho_{2i} = l_{2i}c''_{\cong i}$ is the utilization observed at queue i. Equation (22) gives us the waiting time for general input traffic.

Table 4 shows the results for the same system as Tables 1 to 3 except that the arrivals are not Poisson. The interarrival times have a squared coefficient of variation equal to 2. For simulation, a hyperexponential distribution was used to generate the interarrival times. No comparison can be made to the Kuehn and Boxma-Meister results since these are not applicable to non-Poisson traffic. The results are not as accurate as for the Poisson arrival case, but are adequate for engineering purposes.

6 Model of the Output Module

Packets coming to the output module go though two tandem queues for service. One is for packet processing and the other is at the output line. Packet processors perform the same function as in the input queue module. The output line is a T_1 line, which may be partitioned into channels. Therefore, there are three possible transmission rates, 56 kbps, 384 kbps or 1.344 Mbps for the output transmission. The queueing model is simple. The only problem is to calculate the four incoming traffic parameters to each output module. We will estimate the traffic parameters to the output queues in the following subsections.

6.1 Traffic Parameters for the Output Modules

Let:

l_{3j} be the packet arrival rate to the packet processor queue of the output module j;

Load	Simulation	Method 1	Method 2	BM	Kuehn
0.2	.3780(.013)	.3665(-3.0)	.3584(-5.2)	.3671(-2.9)	.3413(-9.7)
0.4	.7487(.024)	.7655(-2.2)	.7277(-2.8)	.7709(-2.9)	.6478(-13.)
0.6	1.307(.023)	1.441(10.3)	1.336(-2.2)	1.484(13.5)	1.140(-13.)
0.8	2.259(.022)	2.906(28.6)	2.655(17.5)	3.422(54.8)	2.567(13.6)
0.85	2.569(.024)	3.621(40.9)	3.314(29.)	4.911(91.2)	3.653(42.2)
0.9	2.910(.027)	4.723(62.)	4.351(50.)	12.31(323.)	6.270(115.)

Table 3: Waiting Times for the Third and Fourth Queues

Load	Q1 Sim	Method 1	Method 2	Q2 Sim	Method 1	Method 2	Q3,4 Sim	Method 1	Method 2
0.2	.560(.015)	.510(-8.9)	.516(-7.9)	.446(.016)	.433(-2.9)	.430(-3.7)	.407(.024)	.393(-3.4)	.385(-5.4)
0.4	1.66(.067)	1.32(-21)	1.35(-19)	1.11(.023)	1.01(-9.0)	.990(-11)	.887(.016)	.850(-4.1)	.812(-8.4)
0.6	5.74(.226)	3.39(-41)	3.47(-40)	2.74(.084)	2.19(-20)	2.14(-22)	1.71(.032)	1.66(-2.9)	1.55(-9.1)
0.8	29.7 (6.17)	14.4 (-52)	14.6(-51)	6.98(.628)	5.73(-18)	5.67(-19)	3.24(.094)	3.47(7.1)	3.22(-.62)
0.85	61.3 (8.31)	29.8(-51)	30.1(-51)	9.48(.451)	7.96(-16)	7.93(-16)	3.80(.089)	4.36(15)	4.05(6.6)

Table 4: Waiting Times for Non-Poisson Arrivals

$E\{L_{3j}\}$ be the average packet length of the traffic to the queue;
$v^2_{L_{3j}}$ be the squared coefficient of variation of the packet length;
$v^2_{T_{3j}}$ be the square coefficient of variation of the packet inter-arrival time to the queue.

Still assuming that all packets destined to output queue j arrive there without loss, then (Whitt 1983)

$$l_{3j} = \sum_{PVC_k \in \text{ output } j} \alpha_k \quad (23)$$

$$E\{L_{3j}\} = \sum_{PVC_k \in \text{ output } j} E\{L_k\}\alpha_k / l_{3j} \quad (24)$$

$$v^2_{X_{3j}} = v^2_{L_{3j}} = \sum_{PVC_k \in \text{ output } j} \frac{\alpha_k E^2\{L_k\}(1+v^2_{L_k})}{l_{3j} E\{L_{3j}\}^2} - 1 \quad (25)$$

These three parameters are calculated independently of the previous queue information, but $v^2_{T_{3j}}$ depends on that information and requires more calculation. It is derived in the following steps.
a). Let $v^2_{SW_i}$ be the squared coefficient of variation of the input queue i packet inter-departure time at the output of the switch module. Then,

$$v^2_{SW_i} = \rho^2_{2i} v^2_{X_{2j}} + (1-\rho^2_{2i}) v^2_{T_{2i}}. \quad (26)$$

b). Traffic from input queue i will split among the output queues. The probability, γ_{ij}, that a packet from queue i goes to output queue j is, approximately,

$$\gamma_{ij} = \frac{\sum_{PVC_k \in (i,j)} \alpha_k}{l_{2i}} \quad (27)$$

where (i,j) is defined to be a path from input queue i to output queue j. Thus, the variability parameter, $v^2_{SW_{ij}}$, of the packet inter-arrival time from input queue i to output queue j is,

$$v^2_{SW_{ij}} = 1 - \gamma_{ij} + \gamma_{ij} v^2_{SW_i}. \quad (28)$$

c). The traffic from input queues to the output queue j is combined and forms the traffic to the output module. If we assume that the traffic from different queues are independent, we can use the QNA's traffic-merge method to combine the traffic.

Let θ_{ij} be the fraction of traffic from input queue i to output queue j, that is

$$\theta_{ij} = \frac{\sum_{PVC_k \in (i,j)} \alpha_k}{l_{3i}}. \quad (29)$$

Hence, the variability parameter of inter-arrival to the output module j is

$$v^2_{T_{3j}} = (1 - w_{3j}) + w_{3j} \sum_{i=1}^{N} (\theta_{ij} v^2_{SW_{ij}}) \quad (30)$$

where $w_{3j} = 1/[1 + 4(1-\rho_{3j})^2((\sum_{i=1}^{N} \theta^2_{ij})^{-1} - 1)]$ and $\rho_{3j} = l_{3j}E\{X_{3j}\} = l_{3j}E\{X_{1j}\}$ is the load of the output processor j. The input traffic parameters to the output packet processor are then completely specified by formulas (23) through (30).

6.2 Delay of the Output Module

There are five sources of delay for an output module: packet processing delay, waiting time delay in the receive cache, packet transfer delay, waiting time delay for multiplexing and transmission and the transmission delay.

The first queue in the output module is the packet processor queue, which performs the same function as in the input queue. Changing the subscript $1i$ of (1) into $3j$, gives the formula for the mean waiting time, $E\{W_{3j}\}$, of the output packet processor. The transfer delay is the same as that of the input module. The packet inter-departure variability parameter, $v^2_{T_{4j}}$ after the packet processor can be calculated, which is an input parameter to the transmission queues. Due to the lossless assumption, the other parameters for transmission queue j (i.e., l_{4j}, $E\{L_{4j}\}$ and $v^2_{L_{4j}}$) are the same as those for the input to the output packet processor (i.e., the values given by (23) to (25)). $E\{W_{4j}\}$ is then calculated using (1) with these parameter values.

The last queue of the frame relay switch is at the output transmission line. The transmission line can be a full T1 line or channelized. In the former case, the arrival traffic is known, it is a simple GI/G/1 queue; in the later case, the traffic to output module j is split onto the channels. We consider that it is randomly split. The QNA method can be applied to calculate the parameters and delays of the traffic streams.

This completes our discussion of the delay model for the output modules.

7 Switch Delay of PVC Packets

The PVC packet delay through the switch is defined to be the interval from the time a PVC packet arrives to the input module of the switch till it leaves the output module of the switch. If PVC_k goes through the switch via input module i and the output module j, the average switch delay, $E\{D_k\}$, of the PVC_k packets is the sum of the waiting time delay and the service time delay of the passed queues in the switch. The average waiting time delay is the same for any PVC going through the same path; but the service time delay is different for PVCs whose traffic has different average packet lengths. Let $E\{W_{ij}\}$ be the waiting time delay of the path from input module i to output module j and $E\{S_k\}$ be the average service time delay for PVC_k packets going through the switch. Then, if PVC_k is via path (i, j), the PVC_k packet delay is,

$$\begin{aligned} E\{D_k\} &= E\{W_{ij}\} + E\{S_k\} \\ &= E\{W_{1i}\} + E_G\{W_{2i}\} + E\{W_{3j}\} + E\{W_{4j}\} \\ +E\{X_{1i}\} &+ E\{X_{3j}\} + E\{L_k\}(\frac{2M}{R_t} + \frac{1}{R_w} + \frac{1}{R_{4j}}) \quad (31) \end{aligned}$$

The first four terms are the waiting time delay; the remaining terms are the service time delays. $E\{X_{1i}\}$ and $E\{X_{3j}\}$ are due to the input and the output packet processors; R_t is the packet transfer rate of the input or output modules; R_w is the service rate of the cyclic queue switch and R_{4j} is the service rate of the transmission line.

8 Conclusion

Frame relay service is PVC-oriented. PVC traffic goes through a predetermined path. A frame relay switch contains three different type of modules, the input module, the switch module and the output module.

We modeled each module by using infinite buffer queues and analyzed the delay performance of each queue. The basic approach used in the analysis is the QNA method, which is based on two traffic parameters, the packet rate and the squared coefficient of variation of the packet inter-arrival time and two service parameters, the mean packet service time and the squared coefficient of variation of the packet service time. The main problem is how to approximate the variability parameter of the packet inter-departure time. We have combined several methods and developed a delay model for non-exhaustive cyclic queues, which gives a better approximation for the delay performance in a variety of situations, especially when the load is very unbalanced and high. In addition, we modified the model to get the delay performance of cyclic queues with general input traffic.

For the cyclic queue switching system, we give a method to analyze the packet inter-departure parameters at the output of the switch, which is an important part of the frame relay switch model.

We will further develop the network analysis models for a frame relay network. The main problem is to determine the packet interarrival time variability parameters for each trunk of the network. After the network model is perfected, network optimization can be performed.

References

K.J. Chen, K.K.Y. Ho and V.R. Saksena 1989, "Analysis and Design of a Highly Reliable Transport Architecture for ISDN Frame-Relay Networks," IEEE J. on Selected Areas in Commun. **7**, No. 8 (Oct.): 1231-1242.

P. Marsden 1991, "Internetworking IEEE 802/FDDI LAN's Via the ISDN Frame Relay Bearer Service," Proceedings of the IEEE **79**, No.2 (Feb.): 233-229.

R. G. Cole 1988, "Performance Models of Binary Synchronous Communications Multi-point Circuits over Virtual Private-Line Frame Relay Networks," AT & T Technical Journal, (Sept.-Oct.): 41-56.

W. Whitt 1983, "The Queueing Network Analyzer" BSTJ **62**, No. 9 (Nov.): 2779-2815.

K. T. Marshall 1968, "Some Inequalities in Queueing," Oper. Res. **16**, No.3 (May): 651-665.

P. J. Kuehn 1979, "Multiqueue System with Nonexhaustive Cyclic Service," BSTJ **58**, No. 3(Mar.): 671-698.

O. J. Boxma and B. Meister 1986, "Waiting Time Approximations for Cyclic-Service Systems with Switch-over Times," Performance Eval. Rev. **14**: 254-262.

K. S. Watson 1984, "Performance Evaluation of Cyclic Service Strategies–a Survey," *Performance'84*, ed. E. Gelenbe, 521-533, North-Holland, Amsterdam.