Analysis of timed-token ring protocols

Eitan Altman

UNITÉ DE RECHERCHE
INRIA-SOPHIA ANTIPOLIS

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.: (1) 39 63 55 11

Rapports de Recherche

N° 1570

Programme 1
Architectures parallèles, Bases de données, Réseaux et Systèmes distribués

ANALYSIS OF TIMED-TOKEN RING PROTOCOLS

Eitan ALTMAN

Décembre 1991

Analyse de protocoles pour réseaux à jetons temporisés

Eitan ALTMAN

INRIA Centre Sophia Antipolis
2004 Route des Lucioles
06560 Valbonne, France

16 Octobre, 1991

Résumé

Nous analysons dans cet article les performances de protocoles pour des systèmes d'anneaux à jetons ayant des contraintes sur le temps de cycle. Chaque station sur l'anneau peut avoir une contrainte différente ainsi qu'une taille de tampon différente. Nous considérons deux genres de service: (i) le cas 1-limité, où un paquet au plus peut être transmis par station à chaque arrivée du jeton, et (ii) le service exhaustif, où chaque station transmet des paquets jusqu'à épuisement ou bien jusqu'à ce que la contrainte sur le temps de cycle ne soit plus vérifiée. Le système que nous analysons est une approximation du protocole FDDI (Fiber Distributed Data Interface) et du bus à anneaux IEEE 802.4. L'analyse utilise la méthode de développement des probabilités stationaires en série de puissance en fonction de la charge du système, introduite par Blanc [3] pour l'analyse de systèmes de files d'attente multi-dimensionels. Nous obtenons l'espérance du débit et du délai dans chaque station, ainsi que les deux premiers moments de la taille des files d'attente.

Analysis of Timed-Token Ring Protocols

Eitan ALTMAN

INRIA Centre Sophia Antipolis
2004 Route des Lucioles
06560 Valbonne, France

October 16, 1991

Abstract

We analyze in this paper the performance of Token Ring Protocols with constraints on the cycle times. Each station may have a different cycle time constraint, and a different number of buffers. We consider two types of service discipline in the different stations: (i) the 1-limited case, where at most one packet may be transmitted from each queue at each visit of the token and (ii) the exhaustive service discipline where in each queue, packets are transmitted till either the queue empties or the cycle time constraint in that queue is exceeded. The system we analyze approximates both the FDDI protocol (Fiber Distributed Data Interface) and the IEEE 802.4 Token Bus Standard. The technique of the analysis is based on the power-series expansions of the state probabilities as functions of the load of the system, a method introduced by Blanc [3] for analyzing multi dimensional queueing systems. We obtain the expected throughput and delay in every station, as well as the first two moments of the queues' length.

Keywords: Timed Token Rings, FDDI, power-series expansion.

1 Introduction

In recent years, a growing number of protocols for Lans (local area networls) and Mans (Metropolitain area networks) have been developed and implemented, which use cycle time constraints mechanisms. Such are the IEEE 802.4 Token Bus [1] and the FDDI (Fiber Distributed Data Interference) standards [2],[11]. The need of cycle time constraints arrizes due to the integration of real time traffic such as voice and video (known as isochronous traffic) with non real time traffic, such as data and file transfers (nonisochronous traffic). The role of the cycle time constraints is to prevent nonisochronous traffic to saturate the system, which would cause intolerable transmission delays of the isochronous traffic. Both the IEEE 802.4 and the FDDI standards may have different time constraints at different nodes, which allows prioritization.

A large amount of literature exists on modeling and performance evaluation of token ring protocols, see e.g. the survey by Takagi [13] and references therein. Even in the absence of cycle time constraints, explicit solutions are often unavailable, and approximations, numerical methods or simmulations are the only available approaches for analysis.

Timed Token Protocols have first been suggested by Grow [6] and Ulm [16]. The latter analyzes the influence of different parameters on the utilization of the ring. The cycle time characterization have been analyzed by Johnson [7] and Sevcik and Johnson [12]. Throughput analysis is provided by Dykeman and Bux [5] and Pang and Tobagi [10]. Waiting time analysis for Timed Token Rings with a single buffer are presented by Takagi [14] and Nakamura et al. [9]. Tangemann and Sauer [15] analyze the FDDI protocol and obtain several performance measures such as waiting times, throughput and blocking probabilities for the case of several finite queues. They introduce several approximations such as the assumption that gated service discipline is used in each station.

We consider two different Timed-Token Ring protocols, for which we obtain the expected throughput and delay in every station, as well as the first two moments of the queues' length. (i) the 1-limited case, where at most one packet may be transmitted from each queue at each visit of the token and (ii) the exhaustive service discipline where in each queue, packets are transmitted till either the queue empties or the cycle time constraint in that queue is exceeded. We use an iterative scheme based on the power-series expansions of the state probabilities as functions of the load of the system, a method introduced by Blanc [3] for analyzing multi dimensional queueing systems. In [3] this method was used to analyze the performance of a polling system operating under the Bernoulli regime. Methods for accelerating the rate of convergence of the iterative scheme as well as methods to reduce the amount of memory required are presented there.

In the following Section we introduce the model and notations. In Sections 3 and 4 we then obtain the iterative scheme for the two Timed-Token Ring protocols. Numerical results are finally presented in Section 5.

2 The Model and Notation

We consider a system with s queues. The waiting room in the ith queue, $i = 0, 1, ..., s-1$ is denoted by $m_i > 1$, which may be either finite or infinite. Packets arrive to queue j in accordance with a Poisson process of parameter λ_j. The transmission (service) times in queue j are identical independently distributed (i.i.d.) with an exponential distribution with parameter μ_i. The token visits the queues in a cyclic order, vz. $0, 1, 2, \cdots, s-1, 0, 1, 2, \cdots$. We assume that the walking times, which are the time between the departure of the token from the n-th visited queue and the arrival to the $n+1$-st queue, are negligible. (This assumption could be relaxed, however this would complicate the iterative numerical scheme, where at each step we would have to solve a system of linear equations instead of performing only a few arithmetic operations). Denote by $\rho_j = \lambda_j/\mu_j$, $0 \leq j \leq s-1$ and $\rho = \sum_{j=0}^{s-1} \rho_j$.

Queue j has a predetermined Target Token Rotation Time $TTRT_j$, $j = 0, ..., s-1$. (Since only nonisochronous traffic have cycle time constraints, we shall assume that queues with isochronous trafic have infinite $TTRT$.) A transmission of a packet in queue j may begin at time t only if the time elapsed since the last time that the token arrived to that queue is strictly less than $TTRT_j$. More precisely, let $\tau_j(T) := \sup_{t \leq T}\{$ the token arrived to station j at time $t\}$. Each queue has local timers TRT_j and THT_j. The value of TRT_j at time T is given by

$$TRT_j = T - \tau_j(T), \qquad j = 0, ..., s-1$$

If upon arrival of the token to queue j, $TRT_j < TTRT_j$ and the queue is non-empty then a packet starts to be transmitted; the value of TRT_j is copied to THT_j; TRT_j is then reset to zero, and both THT_j and TRT_j continue counting up. Otherwise the token moves to the next queue and $TTRT_j$ restarts counting.

Two types of service disciplines are considered in Section 3 and 4 respectively. (i) *The 1-limited regime*: when transmission of a packet ends in queue j, the token passes to the next queue. (ii) *The exhaustive regime*: when transmission of a packet ends in queue j, the transmission of a new packet in that queue may begin if $THT_j < TTRT_j$. If upon completion of transmission in queue j it is empty or $THT_j \geq TTRT_j$ then the token passes to the next queue. For the FDDI protocol, if a token rotation time exceeds the target rotation time, the excess is compensated in the next cycle-time. However we shall not include this feature in our model for the simplicity on the analysis.

Let $\mathbf{n} = (n_0, n_2, ..., n_{s-1})$ be a vector of non-negative integer numbers. We shall make use of the following key approximating assumption throughout the paper. Suppose that $\mathbf{n}$ packets have been transmitted in the different queues in the time period $[\tau_i(T), T]$. Then at time T,

$$TRT_i \simeq \sum_{j=0}^{s-1} \frac{n_j}{\mu_j} \tag{1}$$

Similarly, if the token is in queue h, during the current visit there $\tilde{n}_h$ packets have been transmitted, and during the previous cycle (i.e. the time between the last two arrival instants of the token to queue h) $\mathbf{n}$ packets have been transmitted from the different queues, then we assume

$$THT_h \simeq \frac{\tilde{n}_h}{\mu_h} + \sum_{j=0}^{s-1} \frac{n_j}{\mu_j} \tag{2}$$

We see from (1) and (2) that instead of summing the actual transmission times, we assume that we may sum their expectations. *We thus consider a token ring system where instead of having a constraint on the actual cycle time, we have a constraint on a linear combination of number of packets transmitted in the last cycle in the different queues.*

We shall assume throughout the paper that the system is ergodic and in steady state.

$I\{E\}$ denotes the indicator function of the event E, and $\mathbf{e}_j$ is a vector with zero entries except an entry of one at the jth position, $j = 0, 2, ..., s-1$. For any vector $\alpha = \{\alpha_0, ..., \alpha_{s-1}\}$ we shall understand throughout the paper $\alpha_j = \alpha_{(j) mod(s)}$ whenever $j \geq s$ or $j < 0$; Moreover, $|\alpha|$ will denote the sum of absolute values of the components of α.

In order to describe well the system, the state of the system has to include the number of packets in the different queues $\mathbf{N} = \{N_0, ..., N_{s-1}\}$, the location of the token $H \in$

$\{0, ..., s-1\}$, the value of the timers TRT_j in the different queues $j = 0, ..., s-1$, and the value of the timer THT in the queue where the token is currently. As in Blanc [3], it suffices to consider the Markov Chain embedded at the events of an arrival to or the end of service (transmission of a packet) in any queue.

In order to simplify the computations we shall use an equivalent description of the state, which contains $\mathbf{N}$ and H, however instead of including the timers we include (1) the vector $\mathbf{U} = \{U_0, U_1, ..., U_{s-1}\}$, where U_j, $j \neq H$ represents the number of packets transmitted at queue j during the last cycle, i.e. between the last visit of the token to queue H and the current one; U_j, $j = H$ represents the number of packets transmitted from queue H (where the token is currently) in the last visit there plus those transmitted in the current visit of the token. (2) In the exhaustive case, we include the variable Q which represents the number of packets transmitted in the current visit to queue H. (This variable is clearly not necessary in the limited case, where at most one packet may be transmitted in each visit of the token to any queue).

Suppose the token is actually in queue $H = h$. TRT and THT can be recovered in that queue from U and Q by

$$TRT_h = Q/\mu_h.$$

$$THT_h = \sum_{i=0}^{j} \frac{U_i}{\mu_i} + \frac{Q}{\mu_h}$$

Due to the absence of walking times, U, Q and H are not defined when the system is empty. The empty system is thus fully described by $N = \mathbf{0}$.

Following Blanc [3], we express the arrival rates by

$$a_j \rho = \lambda_j, \qquad j = 1, ..., s.$$

We introduce the bilinear mapping of the interval $[0,1]$ to itself:

$$\rho = \rho(\theta) = \frac{\theta}{1 + G - G\theta} \qquad \left(\theta = \frac{(1+G)\rho}{1+G\rho}\right), \qquad G \geq 0 \tag{3}$$

where G is some constant that is chosen to obtain good rate of convergence for the numerical procedure, under high load conditions (see [3]).

3 1-limited service discipline

According to the previous section, the state of the system is given by the tuple $(\mathbf{N}, \mathbf{U}, H)$. Denote $V = \{0,1\}^s$. We now present the balance equations for the steady state probabilities. For $h = 0, ..., s-1$, $0 < n_h \leq m_h$ and $\sum_{i=0}^{s-1} u_i/\mu_i < TTRT_h$

$$(\sum_{j=0}^{s-1} \lambda_j I\{n_j < m_j\} + \mu_h) p(\mathbf{n}, \mathbf{u}, h) = \tag{4}$$

$$= \sum_{j=0}^{s-1} \lambda_j p(\mathbf{n} - \mathbf{e}_j, \mathbf{u}, h) I\{n_j > 0, n_j > 1 \text{ if } j = h\} \tag{5}$$

$$+ \lambda_h p(\mathbf{0}) I\{\mathbf{n} = \mathbf{e}_h\} I\{\mathbf{u} = \mathbf{0}\} \tag{6}$$

$$+ \sum_{j=1}^{s} \mu_{h-j} I\{n_{h-j} < m_{h-j}\} \tag{7}$$

$$\times \sum_{\mathbf{v} \in V} p(\mathbf{n} + \mathbf{e}_{h-j}, \mathbf{v}, h-j) \tag{8}$$

$$\times \Bigg[I\{u_i = 0 : i = h-j+1, ..., h-1\} I\{u_{h-j} = 1\} \tag{9}$$

$$\times \prod_{r=1}^{j-1} I \left\{ \sum_{i=h}^{h+s-j} \frac{u_i}{\mu_i} + \sum_{i=h-r}^{h-1} \frac{v_i}{\mu_i} \geq TTRT_{h-r} \text{ or } n_{h-r} = 0 \right\} \tag{10}$$

$$\times\ I\{v_i = u_i : i = h, h+1, ..., h-j-1 \text{ or } j = s\} \tag{11}$$

$$+ I\{\mathbf{u} = \mathbf{0}\} \tag{12}$$

$$\times I\{n_{h-r} = 0,\ r = 1, 2, ..., j-1\} \tag{13}$$

$$\times\ I \left\{ n_{h+r} = 0 \text{ or } \mu_{h-j}^{-1} + \sum_{i=h+r}^{h-j+s-1} \frac{v_i}{\mu_i} \geq TTRT_{h+r},\ r = 0, 1, ..., s-j) \right\} \Bigg]. \tag{14}$$

$$p(\mathbf{n}, \mathbf{u}, h) = 0 \text{ if } n_h = 0 \tag{15}$$

$$\sum_{j=0}^{s-1} \lambda_j p(\mathbf{0}) = \sum_{j=0}^{s-1} \sum_{\mathbf{u} \in V} \mu_j p(\mathbf{e}_j, \mathbf{u}, j) \tag{16}$$

(4) represents the flow of probability out of state $(\mathbf{n}, \mathbf{u}, h)$, whereas the right hand of the equation represents the flow of probability into state $(\mathbf{n}, \mathbf{u}, h)$. (5) corresponds to arrivals

to the different queues, when $|\mathbf{n}| \geq 2$ (note that (4) is defined for $n_h > 0$). (6) corresponds to an arrival to the queue h when $\mathbf{n} = \mathbf{e}_h$; note that this is the only case that need to be considered when arrivals occur and $|\mathbf{n}| \leq 1$. (7) corresponds to the end of service in (transmission from) the different queues. If the last event was the end of service in queue $h - j$, then the number of packets there before the end of that service was $n_{h-j} + 1$ which cannot be larger than the $h - j$th buffer's size, from which the indicator in (7). The variables $\mathbf{v}$ on which the summation is taken in (8) correspond to the state variable U before the end of the service in queue $h - j$ (whereas $\mathbf{u}$ corresponds to U at the beginning of service of queue h). When the last queue to be served was $h - j$, we distinguish two cases: (i) when the token left queue $h - j$ then TRT_h was smaller than $TTRT_h$. (ii) when the token left queue $h - j$ then TRT_h was greater or equal to $TTRT_h$.

In the first case, it follows that $u_{h-j} = 1$ and then $u_i = 0$, $i = h - j + 1, ..., h - 1$ (if $j = s$ then $u_i = 0$, $i \neq h$) and hence (9). Suppose that $j \neq 1$. The fact that service ended in queue $h - j$ and that the token is now in queue h rather than in queues $h - j + r$, $r = 1, ..., j - 1$, means that those queues were either empty or that TRT_{h-j+r} exceeded $TTRT_{h-j+r}$ when the token visited them. This is expressed by (10). (11) places restrictions on the v's on which we sum in (8) is taken. It states that some values of the state variable U need not have changed when the token moved from queue $h - j$ to queue h. If the last queue that was served was the same queue that is served now (i.e. $j = s$), it means that a whole cycle has elapsed since, and therefore the $\mathbf{v}$ may be completely different than $\mathbf{u}$. Otherwise, less than a whole cycle has elapsed, and some values are therefore common.

In the second case, more than a whole cycle elapsed since the token left queue $h - j$, which explains (12). Suppose that $j \neq 1$. The fact that service ended in queue $h - j$ and that the token is now in queue h rather than in queues $h - j + r$, $r = 1, ..., j - 1$, means that those queues were empty (note that TRT_{h-j+r} was zero when queue $h - j + r$ was last visited). This is expressed in (13). The fact that after service ended in queue $h - j$, the token is now in queue h rather than in queues $h + r$, $r = 0, ..., s - j$, means that those queues were either empty or that TRT_{h+r} exceeded $TTRT_{h+r}$ when the token visited them. This is expressed by (14).

Next we express the steady state probabilities using the following power series:

$$\begin{aligned} p(\rho(\theta); \mathbf{n}, \mathbf{u}, h) &= \theta^{|\mathbf{n}|} \sum_{k=0}^{\infty} \theta^k b(k; \mathbf{n}, \mathbf{u}, h), \qquad \mathbf{n} \neq \mathbf{0}, \\ p(\rho(\theta); \mathbf{0}) &= \sum_{k=0}^{\infty} \theta^k b(k; \mathbf{0}). \end{aligned} \tag{17}$$

We now replace ρ by θ in the balance equations (4) according to (3), and substitute the power-series (17) into the expressions for the steady state probabilities. Equating the coefficiences of corresponding powers of θ in the balance equations leads to the following equations which are used later as the basic step in the iterative numerical scheme that computes the coefficiences in (17). For $h = 0, ..., s-1$, $0 < n_h \leq m_h$ and $\sum_{i=0}^{s-1} u_i/\mu_i < TTRT_h$ we obtain:

$$(1+G)\mu_h b(k; \mathbf{n}, \mathbf{u}, h) = \tag{18}$$

$$\begin{aligned}
&= \left[G\mu_h - \sum_{j=0}^{s-1} a_j I\{n_j < m_j\} \right] b(k-1; \mathbf{n}, \mathbf{u}, h) I\{k > 0\} \\
&+ \sum_{j=0}^{s-1} a_j b(k; \mathbf{n} - \mathbf{e}_j, \mathbf{u}, h) I\{n_j > 0, n_j > 1 \text{ if } j = h\} \\
&+ a_h b(k; \mathbf{0}) I\{\mathbf{n} = \mathbf{e}_h\} I\{\mathbf{u} = \mathbf{0}\} \\
&+ \sum_{j=1}^{s} \mu_{h-j} I\{n_{h-j} < m_{h-j}\} \\
&\quad \times \sum_{\mathbf{v} \in V} [(1+G) b(k-1; \mathbf{n} + \mathbf{e}_{h-j}, \mathbf{v}, h-j) I\{k > 0\} - G b(k-2; \mathbf{n} + \mathbf{e}_{h-j}, \mathbf{v}, h-j) I\{k > 1\}] \\
&\quad \times \Bigg[I\{u_i = 0 : i = h-j+1, ..., h-1\} I\{u_{h-j} = 1\} \\
&\quad \times \prod_{r=1}^{j-1} I \left\{ \sum_{i=h}^{h+s-j} \frac{u_i}{\mu_i} + \sum_{i=h-r}^{h-1} \frac{v_i}{\mu_i} \geq TTRT_{h-r} \text{ or } n_{h-r} = 0 \right\} \\
&\quad \times I\{v_i = u_i : i = h, h+1, ..., h-j-1 \text{ or } j = s\} \qquad\qquad (19) \\
&\quad + I\{\mathbf{u} = \mathbf{0}\} \\
&\quad \times I\{n_{h-r} = 0,\ r = 1, 2, ..., j-1\} \\
&\quad \times I \left\{ n_{h+r} = 0 \text{ or } \mu_{h-j}^{-1} + \sum_{i=h+r}^{h-j+s-1} \frac{v_i}{\mu_i} \geq TTRT_{h+r},\ r = 0, 1, ..., s-j) \right\} \Bigg].
\end{aligned}$$

The coefficiences of $p(\rho(\theta); \mathbf{0})$ are computed by substituting the power series expansion (17) of the steady state probabilities into the law of total probability, which yields

$$\begin{aligned} b(0; \mathbf{0}) &= 1, \\ b(k; \mathbf{0}) &= -\sum_{|\mathbf{n}| \le k} \sum_{h=0}^{s-1} \sum_{\mathbf{u} \in V} b(k - |\mathbf{n}|; \mathbf{n}, \mathbf{u}, h), \qquad k = 1, 2, \ldots \end{aligned} \tag{20}$$

4 Performance Measures

We describe below the procedure through which we obtain the moments of the queues' length, the expected waiting times and the expected throughputs. The moments of the length of queue j, $j = 0, 1, \ldots, s-1$ are given by

$$E\left\{N_j^{\nu}\right\} = \sum_{k=1}^{\infty} \theta^k f_{\nu}(k; j), \qquad \nu = 1, 2, \ldots \tag{21}$$

where $f_{\nu}(k; j)$ are obtained from (17) and (21) and are given by

$$f_{\nu}(k; j) = \sum_{|\mathbf{n}| \le k} \sum_{h=0}^{s-1} \sum_{\mathbf{u} \in V} n_j^{\nu} b(k - |\mathbf{n}|; \mathbf{n}, \mathbf{u}, h), \qquad k = 1, 2, \ldots \tag{22}$$

In the actual numerical calculations, some techniques may be used to improve the rate of convergence of the partial sums in (21) such as the epsilon algorithm and extrapolations suggested in [3]. For the latter, suppose that we wish to calculate only the first M terms $f_{\nu}(k; j), j = 0, \ldots, s-1, k = 1, \ldots, M, \nu = 1, 2$. Then instead of approximating (21) by the truncated sum till the Mth element, a better approximation would be [3]:

$$E\left\{N_j^1\right\} \simeq \sum_{k=1}^{M} \theta^k f_1(k; j) + f_1(M, j) \frac{\theta^{M+1}}{1-\theta} \tag{23}$$

and

$$E\left\{N_j^2\right\} \simeq \sum_{k=1}^{M} \theta^k f_2(k; j) + \left[f_2(M, j) - \frac{f_2(M, j) - f_2(M-1, j)}{1-\theta}\right] \frac{\theta^{M+1}}{1-\theta} \tag{24}$$

The throughput λ_i^* of queue i, $i = 0, \ldots, s-1$ is given by the probability that it is non empty and $h = i$, times μ_i. Note that for an infinite queue, $\lambda_i^* = \lambda_i$.

The expected soujourn time in the jth queue is then given by Little's formulae $E\left\{N_j^1\right\}/\lambda_j^*$, and the expected waiting time there is given by

$$E[W_j] = E\left\{N_j^1\right\}/\lambda_j^* - 1/\mu_j. \tag{25}$$

5 Exhaustive service discipline

According to the previous section, the state of the system is given by the tuple $(\mathbf{N}, \mathbf{U}, Q, H)$. Denote

$$V = \left[\prod_{i=0}^{s-1}\{0, 1, ..., \lceil\mu_i TTRT_i\rceil\}\right] \times \{0, 1, ..., \max_{0\leq i<s}\lceil\mu_i TTRT_i\rceil\}$$

where $\lceil x\rceil$ is the smallest integer larger or equal to x. We now present the balance equations for the steady state probabilities. For $h = 0, ..., s-1$, $0 < n_h \leq m_h$, $q \leq u_h$, and $\sum_{i=0}^{s-1} u_i/\mu_i < TTRT_h$

$$(\sum_{j=0}^{s-1}\lambda_j I\{n_j < m_j\} + \mu_h)p(\mathbf{n}, \mathbf{u}, q, h) = \tag{26}$$

$$= \sum_{j=0}^{s-1}\lambda_j p(\mathbf{n} - \mathbf{e}_j, \mathbf{u}, q, h)I\{n_j > 0, n_j > 1 \text{ if } j = h\} \tag{27}$$

$$+ \quad \lambda_h p(\mathbf{0})I\{\mathbf{n} = \mathbf{e}_h\}I\{\mathbf{u} = \mathbf{0}\} \tag{28}$$

$$+ \quad \mu_h p(\mathbf{n} + \mathbf{e}_h, \mathbf{u} - e_h, q-1, h)I\{n_h < m_h\}I\{q > 0\} \tag{29}$$

$$+ \quad \sum_{j=1}^{s}\mu_{h-j}I\{n_{h-j} < m_{h-j}\}I\{q = 0\} \tag{30}$$

$$\times \sum_{(\mathbf{v},Q)\in V} p(\mathbf{n} + \mathbf{e}_{h-j}, \mathbf{v}, Q, h-j) \tag{31}$$

$$\times \left[I\{u_i = 0 : i = h-j+1, ..., h-1\}I\{u_{h-j} > 0\}\right. \tag{32}$$

$$\times \prod_{r=1}^{j-1} I\left\{\sum_{i=h}^{h+s-j}\frac{u_i}{\mu_i} + \sum_{i=h-r}^{h-1}\frac{v_i}{\mu_i} \geq TTRT_{h-r} \text{ or } n_{h-r} = 0\right\} \tag{33}$$

$$\times I\left\{\mu_{h-j}^{-1} + \sum_{i=0}^{s-1}\frac{v_i}{\mu_i} \geq TTRT_{h-j} \text{ or } n_{h-j} = 0\right\} \tag{34}$$

$$\times\ I\{v_i = u_i : i = h, h+1, ..., h-j-1 \text{ or } j = s\} \tag{35}$$

$$\times\ I\{u_{h-j} = Q+1\} \tag{36}$$

$$+I\{\mathbf{u} = \mathbf{0}\} \tag{37}$$

$$\times I\{n_{h-r} = 0,\ r = 1, 2, ..., j-1\} \tag{38}$$

$$\times\ I\left\{ n_{h+r} = 0 \text{ or } (Q+1)\mu_{h-j}^{-1} + \sum_{i=h+r}^{h-j+s-1} \frac{v_i}{\mu_i} \geq TTRT_{h+r},\ r = 0, 1, ..., s-j) \right\}\Bigg] \tag{39}$$

$$p(\mathbf{n}, \mathbf{u}, q, h) = 0 \text{ if } n_h = 0 \tag{40}$$

$$\sum_{j=0}^{s-1} \lambda_j p(0) = \sum_{j=0}^{s-1} \sum_{(\mathbf{u},q)\in V} \mu_j p(\mathbf{e}_j, \mathbf{u}, q, j) \tag{41}$$

Most of the terms in the above balance equations are the same as in the 1-limited case. We describe the new trems, and the changes in existing terms. If transmission of a packet ends in queue h, it is possible to go on transmitting packets in the same queue. This is expressed in (29). Both here and in the 1-limited case, a service (transmission) was initiated in queue h after it ended in queue $h-j$ if all the queues in between, i.e. $h-j+r$, $r = 1, ..., j-1$, were either empty or TRT_{h-j+r} exceeded $TTRT_{h-j+r}$ when the token visited them. This is expressed by (33). Here, however, another condition has to be fulfilled, namely that queue $h-j$ is also either empty, or TRT_{h-j} exceeded $TTRT_{h-j}$ when the transmission ended in that queue.

Next we express the steady state probabilities using the following power series:

$$\begin{aligned} p(\rho(\theta); \mathbf{n}, \mathbf{u}, q, h) &= \theta^{|\mathbf{n}|} \sum_{k=0}^{\infty} \theta^k b(k; \mathbf{n}, \mathbf{u}, q, h), \qquad \mathbf{n} \neq \mathbf{0}, \\ p(\rho(\theta); \mathbf{0}) &= \sum_{k=0}^{\infty} \theta^k b(k; \mathbf{0}). \end{aligned} \tag{42}$$

As in the previous Section, we equate the coefficiences of corresponding powers of θ in the balance equations to obtain the following equations that are later used as the basic step in the iterative numerical scheme that computes the coefficiences in (42). For $h = 0, ..., s-1$, $0 < n_h \leq m_h$, $q \leq u_h$, and $\sum_{i=0}^{s-1} u_i/\mu_i < TTRT_h$ we obtain:

$$(1+G)\mu_h b(k; \mathbf{n}, \mathbf{u}, q, h) = \tag{43}$$

$$
\begin{aligned}
= & \left[G\mu_h - \sum_{j=0}^{s-1} a_j I\{n_j < m_j\}\right] b(k-1;\mathbf{n},\mathbf{u},q,h) I\{k>0\} \\
+ & \sum_{j=0}^{s-1} a_j b(k;\mathbf{n}-\mathbf{e}_j,\mathbf{u},q,h) I\{n_j > 0, n_j > 1 \text{ if } j = h\} \\
+ & \; a_h b(k;\mathbf{0}) I\{\mathbf{n} = \mathbf{e}_h\} I\{\mathbf{u} = \mathbf{0}\} \\
+ & \; \mu_h [(1+G) b(k-1;\mathbf{n}+\mathbf{e}_h,\mathbf{u}-e_h,q-1,h) I\{k>0\} \\
& \quad -G b(k-1;\mathbf{n}+\mathbf{e}_h,\mathbf{u}-e_h,q-1,h) I\{k>1\}] \\
& \times I\{n_h < m_h\} I\{q>0\} \\
+ & \sum_{j=1}^{s} \mu_{h-j} I\{n_{h-j} < m_{h-j}\} I\{q=0\} \\
& \times \sum_{(\mathbf{v},Q)\in V} [(1+G) b(k-1;\mathbf{n}+\mathbf{e}_{h-j},\mathbf{v},Q,h-j) I\{k>0\} \\
& \qquad -G b(k-2;\mathbf{n}+\mathbf{e}_{h-j},\mathbf{v},Q,h-j) I\{k>1\}] \\
& \times \Bigg[I\{u_i = 0 : i = h-j+1,\ldots,h-1\} I\{u_{h-j} > 0\} \\
& \times \prod_{r=1}^{j-1} I\left\{ \sum_{i=h}^{h+s-j} \frac{u_i}{\mu_i} + \sum_{i=h-r}^{h-1} \frac{v_i}{\mu_i} \geq TTRT_{h-r} \text{ or } n_{h-r} = 0 \right\} \\
& \times I\left\{ \mu_{h-j}^{-1} + \sum_{i=0}^{s-1} \frac{v_i}{\mu_i} \geq TTRT_{h-j} \text{ or } n_{h-j} = 0 \right\} \\
& \times \; I\{v_i = u_i : i = h, h+1, \ldots, h-j-1 \text{ or } j = s\} \\
& \times \; I\{u_{h-j} = Q+1\} \\
& + I\{\mathbf{u} = \mathbf{0}\} \\
& \times I\{n_{h-r} = 0,\ r = 1,2,\ldots,j-1\} \\
& \times \; I\left\{ n_{h+r} = 0 \text{ or } (Q+1)\mu_{h-j}^{-1} + \sum_{i=h+r}^{h-j+s-1} \frac{v_i}{\mu_i} \geq TTRT_{h+r},\ r = 0,1,\ldots,s-j) \right\} \Bigg].
\end{aligned}
$$

The coefficiences of $p(\rho(\theta);\mathbf{0})$ are computed by substituting the power series expansion

(42) of the steady state probabilities into the law of total probability, which yields

$$\begin{aligned} b(0;\mathbf{0}) &= 1, \\ b(k;\mathbf{0}) &= -\sum_{|\mathbf{n}|\le k}\sum_{h=0}^{s-1}\sum_{(\mathbf{u},q)\in V} b(k-|\mathbf{n}|;\mathbf{n},\mathbf{u},h), \qquad k=1,2,\ldots \end{aligned} \tag{44}$$

6 Numerical results

The results which we obtained were validated using several tools, which are applicable for both the limited and exhaustive service disciplines. For the case of infinite queues, a way of checking the results is by checking whether the well known conservation law for expected waiting times (e.g. [4]) is obtained, i.e.

$$\sum_{h=0}^{s-1}\rho_h E[W_h] = \rho\frac{\sum_{h=0}^{s-1}\lambda_h\beta_h^{(2)}}{2(1-\rho)} = \frac{\rho}{1-\rho}\sum_{h=0}^{s-1}\frac{\rho_h}{\mu_h}. \tag{45}$$

($\beta_h^{(2)}$ is the second moment of the service (transmission) time in queue h). For equal service rates in the different queues (and infinite buffers), we can check the correctness of the results by verifying that the probability that there are k customers all together in the system equals to the steady state probability of finding k customers in an M/M/1 queue whose service rate is the same, and the arrival rate is the sum of the arrival rates to the different queues.

To check the correctness of the cycle constraints mechanism, we used the following model. We consider two infinite queues, where queue zero has no cycle time constraints where as the constraint on the cycle time of queue one satisfies $0 < TTRT_1 \le 1/\mu_1$. In that case, it can be easily verified that the resulting model behaves exactly as a standard M/M/1 queue with two priority classes, where queue zero has non-preemptive priority over queue one. (This does not generalize for more than two queues). For that model, results were thus compared to the corresponding performance measures of non-preemptive priority M/M/1 queues. In figures 1 and 2 we present the expected queueing length as a function of the load ρ. We chose $\lambda_i = 0.5r$, $\mu_i = 10.0$, $i = 0,1$, $r = 0,1,\ldots 9$. We used $G = 0.5$, and ran 10 iterations. The truncated version of the sum (21) and the extrapolated sum (23) are compared with the explicit solution, and it is seen that for the low priority queue, the extrapolated sum is very close to the true value and does significantly better than the truncated sum, whereas in the high priority queue, both sums are quite close to the exact value, but the extrapolated sum does slightly worse than the truncated sum.

Next we describe some results for the 1-limited case. The influence of the TTRT on the expected waiting times is depicted in figure 3. We consider a system with three queues with infinite buffers. The arrival rates to the queues are 1.5, 1, and 0.5 respectively, and the service (transmission) rates are 20, 10 and 5 respectively. Only queue 1 is assumed to have a cycle time constraint, equal to $0.1 \cdot r$ where r is a parameter, $r = 1, 2, 3, 4, 5$; queues 0 and 2 have no constraints. (In practice it may mean that in the first queue there are non-isosynchronous traffic, whereas in the other queues the traffic is isosynchronous). We chose $G = 0.5$ and performed 8 iterations. Figure 3 displays the expected waiting times obtained by the extrapolation approach, i.e. by substituting (23) into (25). To check the correctness of the results we compared $\sum_{i=0}^{s-1} \rho_i E[W_i]$ to the value obtained through (45). Indeed, we obtained *complete correspondence*, namely in both cases we obtained $\sum_{i=0}^{s-1} \rho_i E[W_i] = 0.012802$. When applying the simple truncation method (i.e. summing the first M terms in (21) instead of using (23)), the difference that we obtained was of 0.4%. Note that the expected waiting time in queue 1 is the longest among the queues when it has $TTRT_1 = 0.1$, and it becomes the shortest when $TTRT_1 = 0.4$.

Figure 6 shows the influence of the load on the throughput in a system with 3 queues, where the size of the first queue is two, the second is one, and the third queue is infinite. We worked with the 1-limited model. We chose the arrival rate to the first and second queues to be r where r is a parameter, $r = 1, 2, ..., 7$, and the arrival rate to the third queue is constant and equal to one. The service rates are all equal to 10. We chose $TTRT = 0.1$ in the first queue, and infinite in the others; $G = 0.5$ and the number of iterations is 12.

Next we describe some results obtained for the exhaustive case. In figure 5 we show how the TTRT can be used to adjust priorities between two infinite queues. The TTRT in queue 0 was chosen as $0.1r$, where r is a parameter: $r = 1, 2, ..., 6$. The TTRT in queue 1 was chosen to decrease with r, i.e. $TTRT_1 = 0.7 - 0.1r$. The arrival rates are $\lambda_0 = \lambda_1 = 1.5$, and the service rates are $\mu_0 = \mu_1 = 10$. We ran 8 iterateions, and used $G = 0.5$. Figure 5 shows the influence of the parameter r on the expected waiting times in the queues. Here again, we display the expected waiting times as obtained by the extrapolation scheme. We check the correctness of the results by comparing $\sum_{i=0}^{s-1} \rho_i E[W_i]$ to the value obtained through (45). Indeed, we obtained full correspondence, namely $\sum_{i=0}^{s-1} \rho_i E[W_i] = 0.012857$. whereas 0.5% of difference is obtained by using the truncated scheme.

Finally, we present the influence of $TTRT$ on the expected queues' length in figure 6. We consider the exhaustive model with three queues, each containing two buffers. The arrival rates are 1.5, 1.5 and 1 respectively, and the service rates are all 10. The TTRT in queues 1 and 2 is 0.2, whereas in queue 0 it is given by $0.1r$, $r = 1, 2, ..., 6$. We ran 8 iterations, and chose $G = 0.5$.

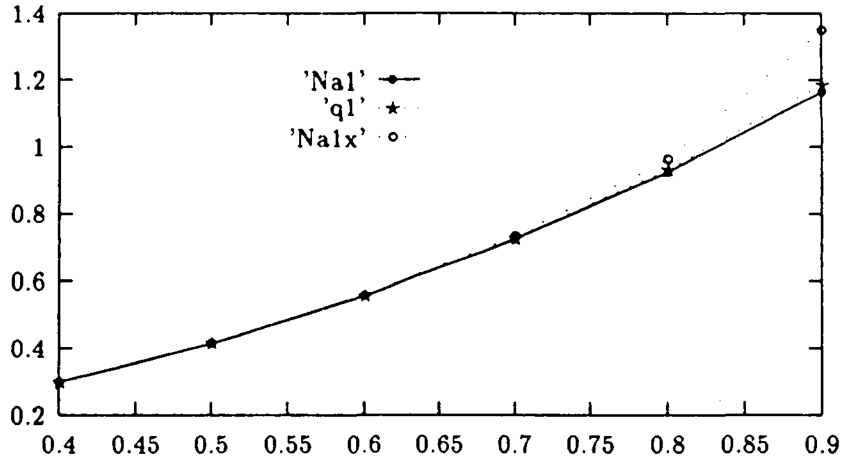

Figure 1: expected length of the queue with non-preemptive priority. **Na1** is the truncated sum of first 10 summands, **Na1x** is the extrapolated sum, and **q1** is the exact value.

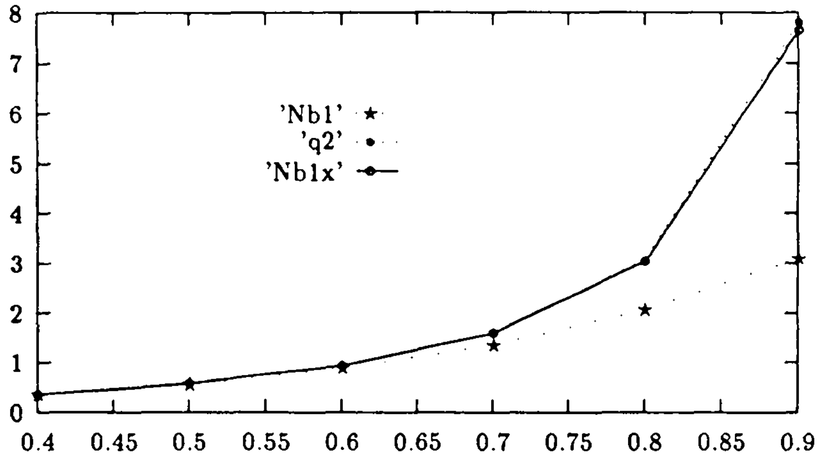

Figure 2: expected length of the queue with low priority. **Nb1** is the truncated sum of first 10 summands, **Nb1x** is the extrapolated sum, and **q2** is the exact value.

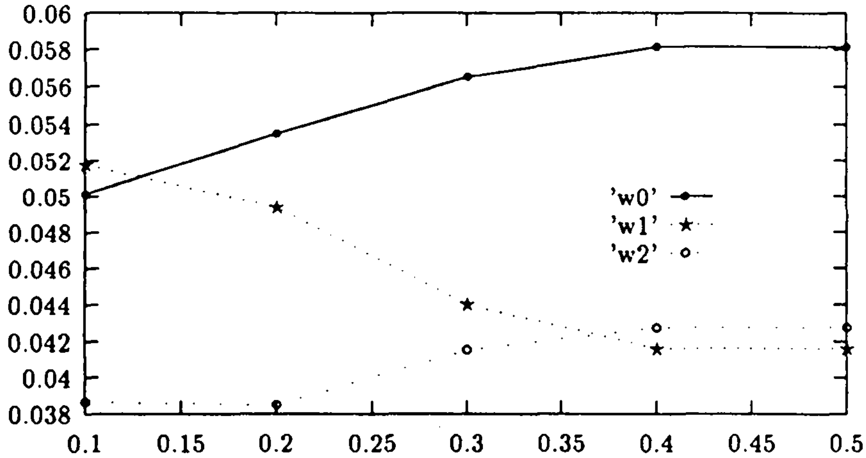

Figure 3: expected waiting times as function of $TTRT$ in the second queue (i.e. in queue 1). w_i denotes the expected waiting time in queue i, $i = 0, 1, 2$.

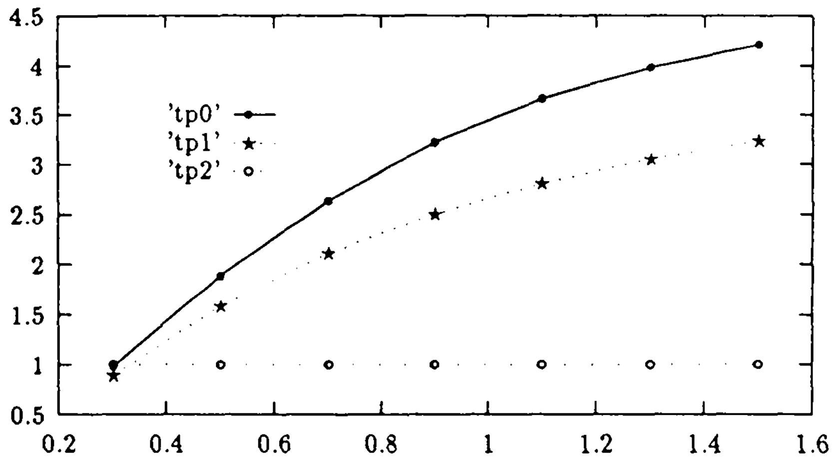

Figure 4: expected throughput as a function of the load ρ. tp_i denotes the expected throughput of queue i, $i = 0, 1, 2$.

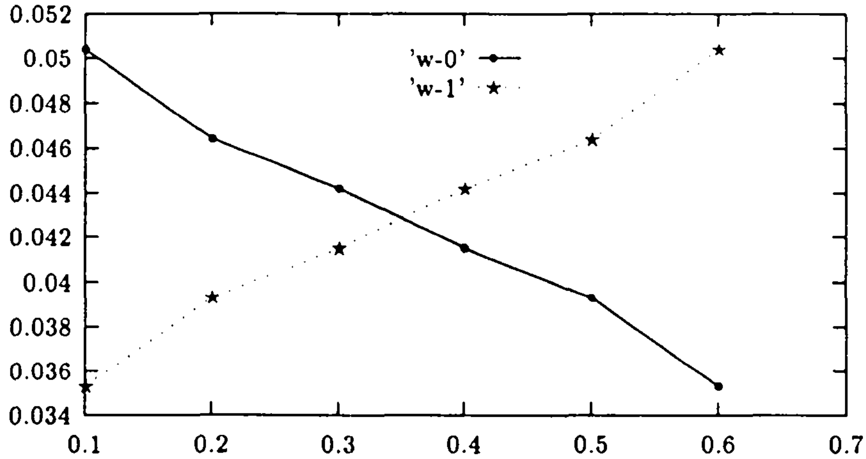

Figure 5: influence of TTRT on expected waiting times in 2 priority queues. $w - i$ denotes the expected waiting time in queue i, $i = 0, 1$.

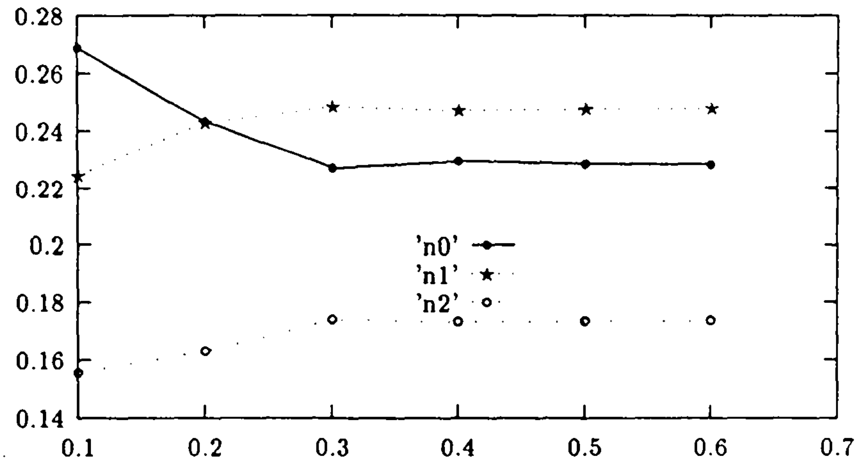

Figure 6: influence of TTRT on expected queues' length. ni denotes the expected waiting time in queue i, $i = 0, 1$.

References

[1] ANSI/IEEE Standard 802.4-1985 "Token-passing bus access method and physical layer specifications", The institute of Electrical and Electronic Engineering, Inc., New York, 1985.

[2] ANSI Standard X3T9.5 "FDDI token ring media access control", 1988.

[3] J. P. C. Blanc, "A numerical Approach to Cycle-Service Queueing Models", *QUESTA* **6**, pp. 173-188, 1990.

[4] O. J. Boxma, W. P. Groenendijk, "Pseudo-Conservation Laws in Cyclic-Service Systems", *Journal of Applied Prob.* **24**, 949-964, 1987.

[5] D. Dykeman and W. Bux, "Analysis and Tuning of the FDDI Media Access Control Protocol", *IEEE J. Selec. Areas in Com.* **6**, No. 6, pp. 997-1010, 1988.

[6] R. M. Grow, "A timed token protocol for local area networks", presented at Electro 82, *Token Access Protocols* (17/3), May 1982.

[7] M. J. Johnson, "Proof that Timing Requirements of the FDDI Token Ring Protocol are Satisfied", *IEEE Trans. Communications*, **COM-35**, No. 6, pp. 620-625, 1987.

[8] R. O. LaMaire, "An M/G/1 Vacation Model of an FDDI Station", *IEEE J. Select. Areas Commun.* **9**, no. 2, pp. 257-264, Feb. 1991.

[9] K. Nakamura, T. Takine, Y. Takahashi and T. Hasegawa, "An Analysis of an Asymmetric Polling Model with Cycle-time Constraint", Proceedings of NATO Advanced Res. Workshop, Sophia Antipolis, France, June 1990.

[10] J. W. M. Pang and F. A. Tobagi, "Throughput Analysis of a Timer Controlled Token Passing Protocol Under Heavy Load", *IEEE Trans. on Communications*, **37**, No. 7, pp. 694-702, 1989.

[11] F. E. Ross, "An overview of FDDI: the fiber distributed data interface", *IEEE J. Select. Areas Commun.* **7**, no. 7, pp. 1043-1051, Sept. 1989.

[12] K. C. Sevcik and M. J. Johnson, "Cycle Time Properties of the FDDI Token Ring Protocol", *IEEE trans. Software Eng.*,**SE-13**, No. 3, 376-385, 1987.

[13] H. Takagi, "Queueing Analysis of Polling Models: an Update", *Stochastic Analysis of Computer and Communication Systems*, Elsivier Science Publishers B.V., 1990.

[14] H. Takagi, "Effects of the Target Token Rotation Time on the Performance of a Timed-Token Protocol", *Proc. Performance '90*, Edinburgh, Scotland, pp. 363-370, Sept. 1990.

[15] M. Tangemann and K. Sauer, "Performance Analysis of the Timed Token Protocol of FDDI and FDDI-II", *IEEE J. Select. Areas Commun.* **9**, no. 2, pp. 271-278, Feb. 1991.

[16] J. M. Ulm, "A timed token ring local area network and its performance characteristics", *Proc. 7th Conf. Local Comput. Networks*, Minneapolis, MN, pp. 50-56, 1982.

ISSN 0249-6399