Performance Modelling of the Cambridge Fast Ring Protocol

M. Zafirovic-Vukotic, I. G. Niemegeers, University of Twente, P.O.B. 217, 7500AE Enschede, The Netherlands

Abstract: The Cambridge Fast Ring is a high speed slotted ring. The features that make it suitable for use at very large transmission rates are the synchronous transmission, the simplicity of the MAC protocol and the possibility of immediate retransmission of erronous packets. A new analytical model of the Cambridge Fast ring with normal slots is presented in this paper. The model shows to be accurate and usable over a wide range of parameters. A performance analysis based on this model is presented.

1. Introduction

Slotted ring protocols are suitable for high speed LANs (HSLANs) [24]. Examples are: the Cambridge Fast Ring (CFR) [20], Orwell [8], Upperbus [7], the proposed IEEE P802.6 standard [18] and FXNET [4]. The CFR has been developed starting from a medium speed ring: the Cambridge ring [10]. It implements acknowledgement at the MAC layer and offers the possibility of immediate retransmission of erroneous packets, which distinguishes it from other slotted ring protocols e.g. Orwell and Upperbus.

The performance of the access mechanism (i.e. the MAC layer

protocol) of CFR is studied in this paper. In the model only normal slots have been considered. The load is of the asynchronous type. Transmission rates in excess of 100 Mbit/s are assumed.

A new approximative analytical model of this access mechanism (AM) and a necessary and a sufficient stability condition are presented in this paper. The model considers the expected packet (i.e. LLC_PDU) delays. Only the queueing delays for access to the medium and the transfer delays are considered. Delays due to processing of the packets are not included. The model has been tested by extensive and detailed simulations in a number of cases typical for HSLANs.

On the basis of the results obtained from the model, a performance analysis of the CFR AM has been done. The sensitivity to the number of stations, the number of slots, the transmission rate, the expected packet length and the slot information field length has been evaluated. One of the main questions is how the mechanism performs in different applications e.g. integrated services local area networks, backbone networks and interconnection networks in multiprocessors.

The paper is organized as follows. The CFR protocol is described in Section 2. The state of the art in analytical modelling of CFR is presented in Section 3. Notation is introduced and the workload model is presented in Sections 4 and 5. In Section 6 the stability condition is derived and a MIGI1 with bulk arrivals and server vacation periods model is presented in detail. In Section 7 this model is tested by comparing to simulations. The performance analysis of the AM is described in Section 8. Finally, some conclusions regarding the model and the performance of the CFR AM are presented in Section 9.

2. The Cambridge Fast Ring Access Mechanism

The ring is partitioned into equal length slots (see Figure 1). We assume that this is achieved by introducing a latency register at the monitor station to virtually lengthen the ring to a multiple of the slot length.

Slots circulate around the ring and can be empty or full. A full slot is occupied by a mini-packet, i.e. a MAC_PDU. Stations are actively coupled to the ring. They repeat or modify the slots. An empty slot may be filled by a mini-packet, if there is one. A full slot circulating around the ring, reaches the destination station which reads it and passes it on to a higher layer. We assume that each station is capable of using every empty slot that arrives and of reading every slot destined to itself.

There are two classes of basic slotted ring AMs depending on which station empties a full slot: the source station, or the destination station. After emptying the slot there are two possibilities: the slot can be used by the station that empties it, or it must be passed to the next downstream station.

The CFR implements the following AM. The source station releases a slot that was full. Further there are two types of slots: normal and channel slots. A channel slot may be reused by the source and the normal slot must be passed on to the next downstream station which may use it (see Figure 2 for an illustration of the function of a normal slot). Only one slot at a time can be carrying mini-packets from the same source.

Channel slots are specially suitable for large bandwidth or bursty sources. However, a fair access to the medium is guaranteed by using the normal slots. Further, using source release of the slots and using only one slot at a time by a station provides the opportunity of retransmitting the erronous mini-packet immediately when a negative ack is received by the source. This in turn implies

that the sequencing of the mini-packets is preserved. This is a

property that other slotted ring protocols e.g. Orwell do not have.

In this paper the performance analysis is done for normal slots only. Analytical modeling of the operation with both normal and channel slots would be more complex. However, we did a simulative performance analysis for that case in [28].

3. State of the Art in Analytical Modelling of the Cambridge

A number of analytical models have been developed for the slotted ring protocols e.g. in [9], [11], [3] and [15]. However, these are either models of a slotted ring protocol where slots are released by the source and more than one slot at a time can be used by a station, or the workload, in particular the arrival process of minipackets, is based on interactive users. In HSLANs however, high throughput users provide most of the load, not the interactive users. So, there is no appropriate analytical model available in the literature which can be used for the CFR.

For a complete overview of the state of the art and a short presentation of new models of the basic AMs of the slotted ring protocols the reader is referred to [25]. Further, peer studies presenting new analytical models we developed for slotted rings are [26] and [27].

4. Notation

Let us introduce the following notation:

- number of stations minus one,

- i-th station in the ring, i=0,1,...n, and for simplicity of notation we assume that station S_0 can also be denoted as S_{n+1} , transmission rate (bit/ μ s or Mbit/s),

- duration of a slot (µs),

- duration of an information field of a slot (µs), such that v<0,

packet arrival rate at station Si (bit/µs or Mbit/s), i=0,...,n,

- the expected duration of packet transmission at Si, if

transmitted at transmission rate w (µs), i=0,...,n,
random variable denoting the bulk size of the arrival process at Si, i.e. the number of mini-packets a packet is split into,

- relative load, or the expected number of mini-packets arriving in the system during of time units, such that

$$\rho = \sum_{i=0}^{n} \lambda_{i \gamma_i \sigma} \ , \tag{1}$$

 $\gamma_i E Z_i^2$ - the first two moments of Z_i , i=0,...,n, X_i - random variable denoting a mini-packet service time in the $model(\mu s)$, i=0,...,n,

 EX_i, EX_i^2 - the first two moments of X_i , i=0,...,n,

 Y_i - random variable denoting a packet service time in the model (μs) , i=0,...,n, EY_i, EY_i^2 - the first two moments of Y_i , i=0,...,n,

- an element of the packet communication source_to_destination matrix, $||p_{ij}||_{n+1 \times n+1}$ that represents the relative traffic intensity from source S_i to destination S_j , and

$$0 \leq p_{ij} \leq 1, \qquad \text{and} \qquad \sum_{j=0}^{n} p_{ij} = 1, \qquad \qquad i,j = 0,...,n. \quad (2)$$

- the number of slots in the ring, - propagation time from S_i to S_j including the latency at station S_j (μ s), i,j=0,...,n, - slot rotation time or τ_{ii} for all i (μ s), such that

$$\tau = \sum_{i=0}^{n} \tau_{i,i+1}, \text{ and } \tau = s\sigma,$$
 (3)

 τ_i - the expected propagation time of a packet or a mini-packet sent by S_i from S_i to the destination (μs), i=0,...,n,

EV_i - the expected mini-packet waiting time at S_i (μs), i=0,...,n, EW_i - the expected packet waiting time at S_i (μs), i=0,...,n, ET_i - the expected packet delay or the expected duration of MAC layer service per packet of S_i i.e. the expected packet delay from arrival at S_i till its complete delivery at the destination

5. Workload Model

We assume that packets arrive at the MAC layer of station Si according to a Poisson process with intensity I_i. Packets (LLC_PDUs) are segmented and MAC protocol control information (PCI) is added to form a number of mini-packets (MAC_PDUs) of

F2.1 163 which the expected value is gi. So, the arrival process of minipackets at Si can be considered to be a bulk Poisson process

Packet lengths are assumed to be independent identically distributed (i.i.d.). Note that often when talking about packet lengths it is assumed that the length is expressed in time units, i.e. that it represents the duration of packet transmission if transmitted at rate

The information field of a MAC_PDU (mini-packet) in the slotted ring protocol has a constant length v x w (bits). The PCI of a mini-packet has a constant length too. So, the length of a minipacket is constant and equal to $\sigma x w$ (bits) i.e. a slot length.

The random variable Zi, denoting the number of mini-packets a packet is split into, has a distribution which is derived from the packet length distribution.

We allow each station to send to any other station including itself i.e. pii can take any value such that relation (2) holds.

6. M|G|1 Model with Bulk Arrivals and Server Vacation Periods

Model Description

A MIGI1 model with bulk arrivals of mini-packets and individual service (corresponding to one mini-packet sent per slot visit) [5] can be used for modelling queueing at S_i. A mini-packet is a service unit in this model. This model has to be extended to take into account the periodical inavailability of the server at the queue. The server vacation periods are introduced therefore. Server vacation models have been studied in [5], [6] and [17]

It will be explained in the section on stability conditions that the CFR can also be represented as a multiple cyclic servers system with a limited service discipline, with the restriction that only one server at a time can be serving each queue. An approximative analytical model for a single cyclic server system with a limited service discipline is given in [13] and an exact solution for a symetric case in [19]. However, we cannot use these results because they treat a single cyclic server case and because of the restriction that each queue can be served only by one server at a time. This argument does not hold if s=1. For the s=1 case the reader is referred to [27] which presents a modelling study of the CFR where the restriction that only one slot at at time can be used by a station is released. However, it is interesting to notice that the results of [13] are the same as for a server vacation model where a cycle duration is interpreted as a vacation period (for a view of a cyclic server system as a server vacation model see [6]). A server vacation model is used here as well

A $M^B[G]1$ queue with server vacation has been studied in [1]. However, as will be explained later on, we use a $M^B[G]1$ model only to make a companion MIGI1 model with server vacation and use the latter one for analysis. So, we do not use the results for the bulk arrival case.

Bulks arrive according to a Poisson arrival process with

intensity λ_i and with a bulk size Z_i .

The mini-packet service time X_i is evaluated as the time between the following two moments: the beginning of transmission of a mini-packet at Si, and after this mini-packet has returned back to Si the next arrival of an empty slot that can be used by Si. When a server finds the queue at Si empty it takes a vacation of duration Oi. It is assumed that Xi as well as Oi are i.i.d. and mutually independent.

A companion model of this MB|G|1 bulk arrival model can be made. It is a MIGI1 model where a packet is a service unit [5]. The server takes vacation the same way as in the former model. The intensity of the arrival process is λ_i . The distribution and moments of the packet service time Y_i can be evaluated from the distribution of Z_i (the bulk size) and of X_i . Again, the server vacation time is O_i . The model is illustrated in Figure 3.

In the case that the number of slots is larger than the number of stations in the ring, i.e. if s>n+1, a part of the system capacity is wasted since at most n+1 slots can be used. We do not model this

case. So, we assume s≤n+1.

It is assumed in this model that all the slots are independent i.e. that the state of one slot does not give information about the state of other slots. Since Oi and Xi are assumed to be i.i.d. and mutually

independent, queue lengths at all stations are also independent.

In the following the model is presented. The stability conditions are derived first. The probability π_i that a slot arriving at S_i cannot be used, given that S_i is presently not using a slot, is evaluated. The distribution and the first two moments of the server vacation time O: and the mini-packet service time X_i are derived. Then the first two moments of the packet service time Yi are determined. Finally, the expected packet and mini-packet waiting time and the expected packet delay are evaluated.

Stability Conditions

Let us first analyse the system in the following two cases: (1) all the queues are instable and (2) all the queues are stable.

Let us consider the case where all the queues are instable i.e.

each station is empty with probability zero. System capacity is wasted because stations have to pass on empty slots unused whenever they have already a slot in use. The server overhead in which this results depends on the pattern that the ring operation follows. These patterns differ in the order of empty slots that visit each particular station. In the case which has the best performance it never happens that a station has a full slot in the ring when an empty slot arrives. So, the order of empty slots which visit each particular station is the same as the physical order of slots in the ring. In the case which has the worst performance each empty slot has to skip (s-1) stations since the other (s-1) slots are occupied by packets of these stations. It can be shown that both patterns which have been described are such that if they are entered they will not be changed with probability one.

Let us now look at the system when each queue is in the stable state, so, each station is empty with a nonzero probability. In this case the system is randomly changing patterns of operation due to visiting empty stations.

We give here a necessary condition and a sufficient condition for all the queues to be stable (see Figure 4). For determining the necessary condition we assume the best case pattern of passing empty slots. For determining the sufficient condition we assume a worst case pattern. Note that both of the patterns are not realistic (though possible) under a random arrival process unless all the

Consider now the best case pattern. So, each slot after being emptied is used by the first downstream station. In this case we can represent the system as a multiple cyclic server system as follows. The service time is equal to a mini-packet transmission time i.e. σ .

The switchover time is equal to $\tau/(n+1)$.

According to the results available for a multiple cyclic server system (see [16]) which are based on the pseudo-work conservation law [2], we have the following stability condition. Let ER₁ denote the expected server rotation time in this case. We have

$$ER_1 = \frac{\tau}{1-\rho}, \tag{4}$$

where ρ is given in (1). For all the queues to be stable it is required

$$\rho < 1, \tag{5}$$

 $\lambda_i \gamma_i ER_1 < s$,

$$\lambda_{j}\gamma_{j}ER_{1} < s$$
, (6) or from (4) and (6) we have

$$\lambda_{j}\gamma_{j}\sigma < 1-\rho \tag{7}$$

for each j=0,...,n. Relation (7) implies (5) and it represents the necessary stability condition.

Assume now the worst case pattern. So, each slot after being emptied has to skip (s-1) stations since they are busy sending. In that case we can again use a multiple cyclic server model as follows. The service time is equal to a mini-packet transmission time i.e. σ and the switchover time is equal to $s\tau/(n+1)$. In a similar way as for (7), one can derive the sufficient stability condition:

$$\lambda_j \gamma_j \tau < 1 - \rho$$
 (8)

for each j=0,...,n.

Note that if s=1 the necessary and the sufficient stability condition are the same. Note also that the CFR can be represented as a multiple cyclic servers system with a limited service discipline which has been specified for determining the necessary stability condition and with the restriction that each queue can be served by only one server at a time.

Further on we assume that all the queues are stable.

Probability π_i that a Slot Arriving at S_i cannot be Used given that Si has no Full Slot in the Ring

We proceed as follows. At first, the probability (qi) that when an empty slot arrives at S_k , station S_k ($k \neq i$) is empty and there is a nonempty station S_i ($j\neq i$) that has no full slot in the ring, or S_k is nonempty but has a full slot in the ring with n#0 is determined. Next, the expected duration of service overhead per mini-packet (θ_i) due to protocol properties is determined. Finally, the probability (π_i) that a slot arriving at S_i cannot be used given that S_i has no full slot in the ring is evaluated. The probability π_i will be used later on when estimating the mini-packet service time distribution. In this section we assume that the number of stations is larger than one i.e. n>0, except when determining π_i in formulas (15) and (16). Note that when deriving expressions, explanations of the steps are given within brackets just after the step.

Let q_i denote the probability that, when an empty slot arrives at S_k , either station S_k , $k \in \{0,...,n\}$, $k \neq i$ is empty and there is a nonempty station S_j ($j \neq i$) that has no full slot in the ring, or S_k is

nonempty but has a full slot in the ring. We approximate q_i as follows.

 $q_i = P$ (on arrival of an empty slot at some station S_k , $k \neq i$, either station S_k is empty and there is a nonempty station S_j that has no full slot in the ring $(j \neq i, j \neq k)$, or station S_k is nonempty and has a full slot in the ring }, $k \in \{0,...,n\}$, $k \neq i$

$$= P_{1,i} + P_{2,i},$$
 $k \in \{0,...,n\}, k \neq i,$ (9)

with

and

 $P_{2,i} = P$ { station S_k is nonempty and has a full slot in the ring | an empty slot arrives at S_k }, $k\epsilon\{0,...,n\}, k\neq i$.

Further, we have

$$\begin{array}{ccccc} P_{1,i} = \begin{matrix} 1 & n & n \\ - & \Sigma & (& P\{S_k \text{ is empty}\}.(& 1 - \prod & P\{S_j \text{ is empty or is nonempty} \\ n & k=0 & j=0 & \text{and has a full slot in the} \\ & k \neq i & j \neq i & ring \} \) \) \\ & & i \neq k \end{array}$$

(an approximation because of taking the average over the probabilities concerning each S_k , $k\neq i$)

$$\begin{array}{l} 1 & n \\ = \cdots & \sum\limits_{k=0}^{\tau} \left\{ \begin{array}{l} (1-\lambda_k \gamma_k EX_k) \cdot (1-\prod\limits_{j=0}^{\tau} (1-\lambda_j \gamma_j EX_j + \lambda_j \gamma_j EX_j, -\cdots) \right\} \\ k \neq i & j \neq i \\ j \neq k & j \neq k \end{array} \right.$$

(probability that S_k is nonempty is approximated by $\lambda_k \gamma_k EX_k$, $k \neq i$)

Note that if n=1, i.e. there are two stations in the ring expression (10) takes a value of 0.

Let us now evaluate P { station S_k is nonempty and has a full slot in the ring | an empty slot arrives at S_k }, $k \in \{0,...,n\}$, $k \neq i$. When discussing this probability the reader is referred to the discussion about the stability conditions. It is argued there that a number of stable patterns of operation may exist when all the queues are instable. Since a number of patterns is stable, there is no limiting value of P { station S_k is nonempty and has a full slot in the ring | an empty slot arrives at S_k }, $k \in \{0,...,n\}$, $k \neq i$, when the load converges to the limiting value to be carried by the network. In the state that has the best performance, this probability is zero (note that it is assumed that $s \leq n + 1$). If all the queues are stable, the system can stay in one pattern at very high loads for a long time relative to the duration of a packet transmission. This has a large influence on the variance and on the expected value of the time a packet spends in the system and it may eventually lead to instability of all the queues. However, this effect is incorporated only by its mean into the model. We approximate $P_{2,i}$ as follows:

$$\begin{array}{ll} P_{2,i} = \begin{matrix} 1 & n \\ - & \Sigma \end{matrix} \text{ (P (station S_k has a full slot in the ring that is not the} \\ n & k = 0 & last of the bulk \} + P \left(station S_k has a full slot in the ring that is the last in the bulk \right) . P \left(station S_k is nonempty \right) \end{array}$$

(an approximation because of taking the average over the probabilities concerning each S_k , $k\neq i$)

$$\begin{array}{l} 1 & n \\ = - & \sum_{k=0}^{\infty} \left(\left(\lambda_{k} (\gamma_{k} - 1) \tau \right) + \lambda_{k} \tau . \lambda_{k} \gamma_{k} EX_{k} \right) \\ n & k = 0 \\ k \neq i \end{array}$$

(probability that S_k is nonempty is approximated by $\lambda_k \gamma_k EX_k$, $k \neq i$)

From (9), (10) and (11), we have that

with i=0,...,n, n≠0. This way we finish the estimate of qi.

Each mini-packet occupies a slot for a time τ . However, the system capacity is used somewhat longer per mini-packet transfer i.e. there is some service overhead due to the protocol properties. Extra capacity is occupied after sending a mini-packet because of the following: (1) passing the slot empty to the first downstream station, and (2) passing the slot around empty among all stations except S_i and starting from the first downstream station from S_i . The latter can occur because of two reasons: the fact that the queue at a visited station can be empty given that there is a nonempty station S_j that can use an empty slot $(j\neq i)$, and the fact that a visited station can be nonempty and have a full slot in the ring. The former element in the overhead has the expected value $\tau/(n+1)$. The latter element in the overhead is evaluated starting from the assumption of slot independence. Let θ_i denote the expected duration of this overhead. We evaluate θ_i as follows.

$$\theta_i = \frac{\tau}{n+1} + \sum_{j=0}^{\infty} j, \frac{\tau}{n+1} \cdot q_i j \ (1-q_i) = \frac{1}{1-q_i} \cdot \frac{\tau}{n+1}, \quad (13)$$

with q_i given in (12), and i=0,...,n, $n\neq 0$. This way we finish the estimate of θ_i .

Let π_i denote the probability that a slot arriving at S_i can not be used by S_i given that there are no packets in the ring originating from S_i . In our evaluation of this probability we incorporate the overhead due to the protocol properties.

$$\pi_{i} = \begin{cases}
1 & n \\
- & \sum_{i=0} \lambda_{j} \gamma_{j} (\tau + \theta_{i}), \\
s & j=0 \\
j \neq i
\end{cases}$$

$$i=0,...,n, n \neq 0. \tag{14}$$

For n=0, we define

$$\pi_i = 0$$
, $i=n, n=0.$ (15)

Finally, from (13), (14) and (15), we have

$$\pi_{i} = \begin{pmatrix} n \\ \sum_{j=0}^{n} \lambda_{j} \gamma_{j} \sigma & (1 + \frac{1}{\dots - 1}, \frac{1}{1 - q_{i}}), & i = 0, \dots, n, n \neq 0 \\ j \neq i & & (16) \\ 0, & i = n, n = 0 \end{pmatrix}$$

where q_i is given in relation (12). This way we finish the estimate of π_i .

Server Vacation Time Oi and Mini-packet Service Time Xi

Let us now evaluate the distribution of O_i and X_i , i=0,...,n. It is assumed that O_i , as well as X_i , are i.i.d. This assumption may turn out to be unacceptable under high or asymmetric loads.

Each station has to pass an empty slot to the first downstream station. After passing an empty slot, station S_i will certainly receive this very slot within $(n+1)\tau$. This is because of the following. The maximum time each station can keep a slot occupied is equal to the time a slot carries its mini-packet i.e. τ . There are n stations to be visited by this slot when empty before it returns to S_i . Moreover, the slot has to be passed from one station to another in total duration τ . So, the slot appears at S_i empty certainly within $(n+1)\tau$. Since there are s slots in the ring an empty slot arrives at S_i certainly within $(n+1)\tau$ -(s-1) σ . Let us introduce m_i denoting

$$m_i = (n+1)s - (s-1) = ns+1$$
 (17)

The time between two consecutive arrivals of a slot at S_i that can be used by S_i is a multiple of σ . We assume the following distribution of the server vacation period O_i , recalling the assumption of independent slots:

$$P\{O_i = k\sigma\} = \begin{cases} \pi_i^{k-1}(1-\pi_i), & k=1,2,...,m_{i-1} \\ \pi_i^{m_{i-1}}, & k=m_i \end{cases}$$
 (18)

with i=0,...,n.

A mini-packet service time consists of the slot rotation time and the waiting time for a slot that can be used by the station. The former time is a constant and equal to τ. For the latter time the distribution (18) is taken. So, a mini-packet service time distribution is approximated by

$$\begin{array}{ll} D \\ X_i = O_i + \tau \; , & i = 0, ..., n \end{array} \eqno(19)$$

where = denotes equality in the distribution. The first two moments of Oi are given by

$$EO_{i} = \frac{\sigma}{1 - \pi_{i}} (1 - \pi_{i}^{m_{i}}), \tag{20}$$

$$EO_{i}^{2} = \frac{\sigma^{2}}{(1-\pi_{i})^{2}} \left\{ 1 + \pi_{i} - (2m_{i}+1)\pi_{i}^{m_{i}} + (2m_{i}-1)\pi_{i}^{m_{i}} + 1 \right\}$$
(21)

with i=0,...,n.

The first two moments of Xi are given by

$$EX_i = s\sigma + EO_i, \qquad i=0,...,n, \qquad (22)$$

$$EX_i^2 = s^2\sigma^2 + 2s\sigma EO_i + EO_i^2$$
, $i=0,...,n$. (23)

Packet Service Time Yi

Since a packet consists of Zi mini-packets, the packet service time Yi is equal to

$$Y_{i} = \sum_{j=1}^{Z_{i}} X_{i}^{(j)}$$

$$(24)$$

where random variable Xi(i) denotes the service time of the j-th mini-packet within a packet arriving at S_i , j=1,2,..., i=0,...,n. We assume that $X_i(0)$, j=1,2,... are i.i.d. with the distribution (19).

Let us now evaluate the first two moments of Y_i. From (24) we get (see also [12]):

$$EY_i = \gamma_i EX_i , \qquad (25)$$

$$EY_{i}^{2} = \gamma_{i}EX_{i}^{2} + (EZ_{i}^{2} + 2\gamma_{i} + 1).(EX_{i})^{2}, \qquad (26)$$

with i=0,...,n.

Packet Waiting Time

Now the results for the MIGI1 queue with server vacation are used to obtain the moments of the packet waiting time. We have

$$EW_{i} = \frac{\lambda_{i}EY_{i}^{2}}{2(1-\lambda_{i}EY_{i})} + \frac{\lambda_{i}EO_{i}^{2}}{2EO_{i}}, \qquad i=0,...,n. \quad (27)$$

The expected packet delay consists of the following components: (1) the expected packet waiting time (EWi), (2) the expected service time of all but the last mini-packet of a packet $((\gamma_i-1)EX_i)$, (3) the transmission time of the last mini-packet (σ), and (4) the expected propagation time of the last mini-packet from S_i to the destination (τ_i) . So, we have

$$ET_i = EW_i + (\gamma_{i-1})EX_i + \sigma + \tau_i$$
, $i=0,...,n.$ (28)

Mini-packet Waiting Time

To determine the expected mini-packet waiting time EV_i i.e. the waiting time of a customer in the MB|G|1 model with a server vacation, the following exact formula derived in [13] and [23] can

EV_i = EW_i +
$$\frac{EZ_i^2}{2}$$
 (----- - 1), i=0,...,n. (29)

Discussion

The model provides an implicit estimate of EO_i (and thus EX_i) and of π_i . An explicit solution would demand solving a system of n+1 equations each of the order mi, i=0,...,n. In a symmetric case this reduces to solving one equation of the same order. So, a numerical procedure has to be used in the general case to determine values of π_i and of EO_i. The convergence of the procedure is

ensured if π_i in the first iteration is chosen the smallest possible, i.e. with qi=0. Since the procedure consists of a straightforward usage of the formulas derived here, it is an efficient one. We do not elaborate further on the numerical properties of such a procedure.

Let us now summarize the basic modelling assumptions:

- the load from all the stations S_i (j≠i) can be represented solely by the load intensity $\lambda_j \gamma_j$. This way only the first moments of interarrival times and packet lengths at S_j ($j\neq i$) are used;
- the protocol overhead due to passing empty slots can be adequately modelled by including it in π_i ;
- slots are independent i.e. it is assumed that the state of one slot does not give any information about the states of the other slots; and
- O_i and X_i are i.i.d., and consequently the queue lengths at S_i are independent, i=0,...,n.

These assumptions may cause inaccuracy especially at very high or asymmetric loads, where an underestimate of the expected packet delays is expected.

The assumption s<n+1 has been used for determining the stability conditions and for the recognition of the best and the worst performance pattern if all the queue are instable. However, it is not an essential one for the delay estimate. It has been used only to obtain mi.

The model is accurate in the limit, when $\lambda_i > 0$, i=0,...,n. Note that at high loads the variance of the packet delay can be large because of the large variance of the duration of the protocol overhead due to passing empty slots in the ring. Note also that if the sufficient stability condition is not satisfied and the necessary one is, the model provides an estimate of the expected delays in the most probable pattern of operation given that all the queues are stable. In that case not only the variance of the delay is large, but all the queues may become unstable and the delays infinite. The larger the load, the larger the probability that this happens.

7. Testing and Analysis of the Model

The simulation model of the CFR protocol against which the analytical model is checked is a detailed one. It is written in SIMULA and is documented in [14] and [21]. The analytical model has been tested by comparing the expected packet delays to the results of simulations. 90% confidence intervals have been obtained except for the runs where the correlation between the samples was too large. In those cases only a point estimate of the delay is shown in the figures.

Configurations, system parameters and workload models expected to be typical for HSLANs have been used. We present them here. In all the examples an equal distance between the neighbouring stations has been assumed, i.e. $\tau_{i,i+1} = \tau_{i,j+1}$, i,j=0,...,n. An exponential and a bimodal packet length distribution have been assumed.

We consider the following configurations, system parameters and workload: configuration: cable length = 5 and 1 km, number of stations = 40 and 10; system parameters: transmission rate = 140 Mbit/s, slot information field = 512 bits, overhead in slot = 48 bits, latency register = 24 bits; and workload: expected packet length = 7100 and 3000 bits, a symmetric load, and a symmetric traffic

Note that in the CFR [20] the slot information field length is 256 bits. We have taken another information field because of the following. We are presently conducting a comparative performance analysis of a number of slotted ring protocols. In order to be able to compare them a common information field length of 512 bits has been chosen. Such a choice does not change the qualitative behaviour of the protocol.

The expected packet delay of the CFR AM vs load is depicted in Figures 5 through 9. An exponential packet length distribution has been used except for the case of Figure 9 where the following bimodal distribution has been taken:

P { packet length =
$$x$$
 } =
$$\begin{cases} 0.78, & x = 512 \text{ (bit)} \\ 0.22, & x = 30000 \text{ (bit)} \end{cases}$$
 (30)

This distribution has been chosen because we would like to test the model using a distribution with the same expected value of 7100 bits and a second moment about twice as large as in the case of the exponential distribution. Moreover, a bimodal distribution was thought to be a realistic one for the packet lengths [22]. Let us now

discuss the accuracy of the model.

The MIGI1 with bulk arrivals and server vacation periods model has the following properties. Its estimates fall within the confidence intervals (halfwidths of which are less than 10% of the mean obtained by the simulations) of the expected delays for low and moderate loads until a utilization of about 0.7. At very high loads the model has an underestimate of the expected delays. The cause of the underestimate is that the packet interarrival times and packet length distributions at S_j , $j \neq i$ are represented only by their first

moments and because of the assumption that Oi and Xi are i.i.d. (this is strongly related to the assumption of independent slots).

The model is more accurate for a larger number of stations (see Figures 5 and 7). We attribute this to the smaller correlation between activities at S_i and other stations in these cases. So, the assumption that Oi and Xi are i.i.d. is less justifiable in this case (and so is the assumption of independent slots).

The model is expected to be more accurate if there are more slots in the ring (see Figures 6 and 7). The smaller accuracy when there are less slots in the ring appears because of independent modelling of the slot occupancy, which is more justifiable when there are more slots in the ring. However, this property is not clearly visible from the figures.

Note that if n=0 and s=1 the solution for ET_i is exact. It is however, not exact for s=1 and n+0. An exact solution for that case

under a symmetric load is given in [27].

The accuracy of the model seems not to depend very much on the expected packet lengths in the experiments conducted (see Figures 5 and 6). The model is however, slightly more accurate with shorter packets. The accuracy of the model seems also to be very good for different packet length distributions (see Figures 5 and 9).

So, the bulk arrival model is very accurate in all the cases studied except at very high loads and for s=1.

8. Performance Analysis of the Cambridge Fast Ring AM

The results of the performance analysis of the CFR AM using our analytical model are shown in Figures 5 through 15. Figures 5 through 9 also show simulation results and have already been discussed in the previous section. Figures 10 through 14 are used to study the sensitivity of the CFR AM with respect to the following parameters: the expected packet delays with respect to the number of slots, the number of stations, the transmission rate, the expected packet length and the slot information field length, respectively. All the other parameters are kept unchanged and are the same as for Figure 5. A load intensity of 80 Mbit/s is used except in Figure 14 where a relative load of 0.65 is used. The same relationship is shown in Figure 15 as in Figure 5 except that the slot information field is 256 bits as in the original CFR. In some of the figures the delays of the CFR where the restriction that only one slot at a time can be used by a station is released are shown for a comparison (this protocol is denoted by CFRV). The results for the CFRV have been obtained using the model of [27]. Let us now evaluate the performance of the CFR AM.

The stability conditions of the CFR AM (see relations (7) and (8)) show that the maximum carried load depends on the number of stations, the slot duration, and the first moments of arrival process and bulk size. The sufficient condition also depends on the number of slots. The necessary condition does not depend on the number of

slots. Hence, it also does not depend on the ring latency.

Let us now analyse the stability condition expressed in relations (7) and (8). For all the queues to be stable it is necessary that the arrival rate of mini-packets at each S_i per slot duration ($\lambda_i \, \pi \, \sigma$) is less than one minus the relative load (ρ). For all the queues to be stable it is sufficient that the arrival rate of mini-packets at each Si stable it is sufficient that the arrival rate of mini-packets at each S_1 per ring latency ($\lambda_i \gamma_i \tau$) is less than one minus the relative load (p). The relative load (p) depends on the arrival rate of mini-packets at each station ($\lambda_i \gamma_i$, i=0,...,n). This arrival rate depends on the arrival rate of packets at each station ($\lambda_i \gamma_i$) and on the expected number of mini-packets included in a packet (γ_i) which depends on the slot information field length (v) and on the packet length distribution. From (7) and (8), we have that the relative load or the total arrival rate of mini-packets in σ time units (ρ) is bounded by 1, if the system stable.

The lower bound on the maximum carried load is given by the sufficient stability condition and the upper bound on the maximum carried load by the necessary stability condition (see Figures 5 through 9). The CFR AM proves to be capable of carrying loads that are not far off from its transmission rate e.g. about 110 Mbit/s in the case of Figure 5.

The CFR AM shows a steep increase of the delays when the load is at about 0.7 of the maximum carried load estimated by the model (see Figures 5 through 9). This indicates that this AM is very sensitive to sudden changes of the load in that region of the load values. So, a sudden and unexpected increase of the delays may be observed by a user. Some other slotted ring protocols e.g. Orwell do not have this effect so strongly [26].

The expected delay is quite sensitive to the number of slots in

the ring (see Figures 7, 8 and 10 and Table 1). It performs much better if the number of slots is very small (e.g. 2) than if it is large (e.g. 40), see Figure 10. (This is opposed to e.g. Orwell where the performance is better with a larger number of slots e.g. 7 than with the smaller one e.g. 2, see [26].) The main reason is in the fact that only one slot at a time can be occupied by a station. This property has a larger influence on the performance if the number of slots is larger (i.e. the ratio between the number of stations and the number of slots decreases). In that case the delays are larger.

The expected delays depend on the number of stations in the ring (see Figures 5, 7 and 11). The delays decrease asymptotically, with the increase of the number of stations. The performance differs significantly for the smaller and medium number of stations (e.g. between 8 and 60 in Figure 11). As already mentioned if n+1<s a part of the ring capacity is wasted since at most n+1 slots can be occupied at a time. There are two main reasons for the increase of the delays when n+1 gets close to s. The first one is the fact that only one slot at a time can be occupied by a station. This effect has already been explained. The second reason is queueing at S_i which takes a larger share in the total delay than waiting for the access to the medium. Because of the constant load queueing at S_i is larger if the number of stations is smaller. The latter effect can be observed with some other slotted ring protocols, see [26] and [27]

The expected packet delay varies with respect to the expected packet length (see Figures 5, 6 and 13). Note that when the expected packet length changes, the relative load on the network also changes. This happens because of the change in the expected number of mini-packets in a packet which causes a change in the the expected number of mini-packets that are the last ones in a packet. These mini-packets are only partially filled in by data. Table 3 shows the relative load ρ vs the expected packet length (see relation (1)). This is why the performance strongly degrades when the expected packet length is smaller than the slot information field length. However, this is not to be expected in a typical slotted ring.

Let us now analyse the case when the expected packet length is larger than the slot information field. The larger the expected value the larger the packet delay. Note that when the expected packet length increases, the load on the network decreases. The increase of the delays in Figure 13 is mainly due to the fact which permits that only one slot at a time can be used by a station.

Since a packet is split into a number of mini-packets each

having its own PCI, the packet transmission time is approximately σ/ν times the transmission time of an unsegmented packet.

The sensitivity of the CFR AM with respect to the slot information field length is depicted in Figure 14 for load values 60, 80 and 100 Mbit/s (see also Figures 5 and 15). The delay function is discontinuous because of the change in the number of slots with the increase of the information field length (see Table 4). The relative load ρ is shown in Table 5.

This AM performs the best with the information field length between 1024 and 2048 bits. This is because of the change in relative load and the number of slots in the ring. If the information field length decreases the relative load increases since the overhead gets large in relation to the slot length. If the information field increases the relative load also increases. This happens because of the decrease of the expected number of mini-packets in a packet which causes the same effect as already explained in the case of a change in the expected packet length. The sharp increase in delays in Figure 14 can be compared to the increase in delays when the offered load is changed (e.g. in Figures 5 through 9) which happens if the utilization is larger than about 0.7. This effect is however combined with the effect of changing the number of slots. If the slot information field is between 1024 and 2048 bits the relative load is small and the number of slots as well, so the performance of the CFR is the best.

Note however, that the choice of the information field length in practice is to a large extent determined by the packet length distribution. Namely, if the most dominant traffic class has a constant packet length (e.g. voice) the best performance of the protocol could be expected if the packet length fits into an integer number of slots. The slot information field length should be chosen such that on one hand, the ratio between the relative load and the offered load is small, and on the other hand the number of slots in AM with 256 bits information field length as proposed in [20] has much worse performance than with 512 bits (see Figures 5 and 15). Figure 14 shows that information field lengths between 512 bits and 3072 bits provide good performance. However, when the

information field length is in this range the delays for eventual packet transmission becomes unacceptable. A uniframe PCM like scheme over the CFR protocol would improve the performance for voice in this case. However, the study in [28] shows that in that case the delays for long packets belonging to the asynchronous traffic become large

The packet delay vs transmission rate is shown in Figure 12 (see also Table 2 which shows the number of slots in the ring vs the transmission rate for the case of Figure 12). The relative load is held constant at 0.65, e.g. 80 Mbit/s load at 140 Mbit/s transmission rate. Above 300 Mbit/s the CFR has increasing delays, mainly because of the increase of the number of slots.

Applications: Because of the properties of the CFR AM, we can conclude that it can perform well when the number of slots is small or moderate relative to the number of stations. This means that in a backbone application with a small number of stations and a large cable length with a large number of slots, this protocol

performs worse than in e.g. an application as a multiprocessor or packet switch or interconnection structure where the number of slots is small relative to the number of stations. The CFR AM could also perform well as an integrated services network, provided that the number of stations is large enough relative to the number of slots which is in general to be expected. However, fulfillment of the QoS requirements for the different traffic classes e.g. synchronous and asynchronous traffic is a subject of the study in [28].

9. Conclusion and Further Work

The CFR AM has been analysed. A new analytical model has been developed and a performance analysis has been done.

The conclusions concerning the analytical model can be summarised as follows:

- the exact necessary and sufficient stability condition for the CFR AM has been derived starting from the pseudo-work conservation law used for multiple cyclic server models;
- the MiGI1 with bulk arrivals and server vacation periods model provides a good estimate of the expected packet delays over a wide range of parameters under a symmetrical load and traffic pattern for s+1; it gives a good qualitative insight into the delays and at lower and medium loads an accurate delay estimate as well; and
- the model is more accurate with more stations in the ring.

The conclusions concerning the performance of the CFR AM can be summarised as follows:

- the CFR AM shows to be able to carry loads that are not far off from its transmission rate;
- the CFR AM performs better if the ratio between the number of slots and the number of stations is smaller;
- the performance of the CFR AM is better with smaller expected packet lengths:
- the CFR AM shows a good performance at higher transmission rates only if the number of slots is small relative to the number of
- the following two parameters have the strongest influence on the performance of the CFR if the system parameter slot information field length is changed: the ratio between the relative and the offered load, and the number of slots; the smaller they are the better the performance is; a slot information field length between 512 bits and 3072 bits provides good performance for representative system parameters and workloads; and
- the performance of the CFR AM makes it interesting for applications where the ratio between the number of stations and the number of slots is large or moderate e.g. as multiprocessor or packet switch or interconnection structure.

We are in the process of using this model as well as other analytical models we developed [25] in a comparative analysis of slotted ring protocols at high speeds (e.g. CFR, Orwell, their variants and uniframe slotted ring). The asymmetrical cases and workloads consisting of synchronous and asynchronous traffic are included (see e.g. [28]).

Acknowledgement

The authors would like to thank Dr.M.Sposini for useful discussions and supplying literature concerning HSLANs. The authors would also like to thank Dr.E.van Dooren and Ir.B.van Arem for reading the manuscript and giving some usefull remarks. Finally, we are gratefull to Prof.O.J.Boxma for his interest in the work and providing us with literature.

[1] Baba Y., "On the M^X/G/1 queue with vacation time", Oper. Res. Letters, Vol.5.,No.2., pp. 93-98, Elsevier Sc.Pb.B.V., July 1986. [2] Boxma O.J., Groenendijk W.P., "Pseudo-conservation laws in cyclic-service systems", to appear in Journ. Appl. Prob., Vol.24, December 1987.

[3] Bux W., "Local area subnetworks: a performance comparison", IEEE Trans. on Comm., Vol. Com.-29, No.10, pp. 1465-1473, October 1981.

[4] Casey L.M., Dittburner R.C., Gamage N.D., "FXNET: a backbone ring for voice and data", IEEE Comm. Mag., Vol.24, No.12, December 1986.

[5] Cohen J.W., The Single Server Queue, North Holland, 1982. [6] Doshi B.T., "Queueing systems with vacations - a survey", Queueing Systems 1, pp. 29-66, 1986. [7] Giloi W.K., Behr P., Zuber G., "Upperbus, a high speed backbone for metropolitan area networks", Proc. EFOC/LAN'86,

pp. 286-289, Amsterdam, June 1986. [8] Falconer R.M., Adams J.L., "Orwell: a protocol for an integrated services local network", British Tel. Tech.J., Vol.3, No.4, pp.27-35, October 1985.

[9] Hayes J.F., Sherman D.N., "Traffic analysis of a ring switched

[3] Italyes J., Siletinal D.N., Trajic analysis of a ring switched data transmission", Bell System Techn.J., Vol.50, No.9, pp. 2947-2978, November 1971.
[10] Hopper A., "The Cambridge ring - a local network", Advanced Techniques for Microprocessor Systems, Hanna F.K. (ed.), Peter

Pergrinus Ltd., Stevenage, U.K., pp.67-71, 1980.
[11] King P.J.B., Mitrani I., "Modelling a Slotted Ring Local Area Network", IEEE Trans. on Comp., Vol. C-36, No.5, pp. 554-561, May 1987.

 [12] Kleinrock L., Queueing Systems, Vol.I, J.Wiley, 1979.
 [13] Kuehn P.J., "Multiqueue systems with nonexhaustive cyclic service", Bell Syst. Tech. J., Vol.58, No.3, pp.671-698, March 1979. [14] Lande v.d. K., Design and Implementation of a Cambridge Fast Ring Variant Simulation Model, Report no. 080-87-06, University of

Twente, The Netherlands, February 1987. [15] Loucks W. M., Hamacher V. C., Preiss B. R., Wong L., "Short packet transfer performance in local area ring networks", IEEE Trans. on Comp., Vol. C-34, No.11, pp.1006-1014, November 1985. [16] Morris R.J.I., Wang Y.I., "Some results for multi-queue systems with multiple cyclic servers", Performance of Computer Communication Systems, H.Rudin and W.Bux (ed.), Elsvier Sc.Pb.

pp. 4.2A-5-1 - 4.2A-5-7, 1984. [17] Scholl M., Kleinrock L., "On the M/G/1 queue with rest periods and certain service-independent queueing disciplines", Oper. Research, Vol.31, No.4, pp.705-719, July-August 1983. [18] Sze D.T.W., "A metropolitan area network", IEEE Journ. on

Sel. Areas in Comm., Vol. Sac-3, No.6, pp.815-824, Novemb. 1985.

[19] Takagi H., Analysis of Polling Systems, The MIT Press,
Cambridge, Massachusetts, 1986.

[20] Temple S., "The design of the Cambridge ring", Ring
Technology Local Area Networks, Dallas I.N. and Spratt E.B.(ed.),

Elsevier Sc.Pb.B.V., pp.79-88, 1984.

Elsevier Sc.F.D. V., pp. 17-60, 1764.
[21] Valk D., Performance Analysis of Integrated Services Slotted Ring Protocols in Gathering HSLANs, Report no.080-87-08, University of Twente, The Netherlands, March 1987.
[22] Vyncke E., Danthine A., "A realistic simulation of a wideband backbone network", Proc.ICCC'86, North Holand, pp. 176-181, 1986.

[23] Whitt W., "Comparing batch delays to customer delays", Bell Syst. Tech. J., Vol.62, pp. 2001-2009, September 1983.
[24] Zafirovic-Vukotic M., Niemegeers I.G., "An evaluation of high speed local area network access mechanisms", Proceedings

Communication in Distributed Systems, Aachen, in Informatik - Fachbericht no 130, Springer-Verlag, pp. 426-440, February 1987. [25] Zafirovic-Vukotic M., Niemegeers I.G., "Analytical models of the slotted ring protocols in HSLANS", IFIP WG 6.4 Workshop High Speed Local Area Networks, A.Danthine, O.Spaniol (ed.), Aachen, North Holland, pp. 115-134, 1987.
[26] Zafirovic-Vukotic M., Niemegeers I.G., "Performance

modelling of the Orwell basic access mechanism", ACM SIGCOMM '87: Frontiers in Computer Communication

Technologies, Stowe, Vermont, August 1987.

[27] Zafirovic-Vukotic M., Niemegeers I.G., "Performance modelling of a HSLAN slotted ring protocol", Report no 080-47-48, submitted for publication, 1987.

[28] Zafirovic-Vukotic M., Niemegeers I.G., "Performance analysis of slotted ring protocols in HSLANs", Memorandum INF/87/31, submitted for publication, 1987.

submitted for publication, 1987.

no of slots		2	I	10	ı	20	ı	30	1	40	1
cable length (km) l	0	Т	6.63	ī	14.63	T:	22.6	31	30.63	П

Table 1. The cable length vs the number of slots in the case of Figure 9.

transmission rate (Mbit/s)					
no of slots	17	8	1 27	<u> 38 </u>	40

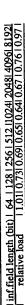
Table 2. The number of slots vs transmission rate in the case of Figure 12.

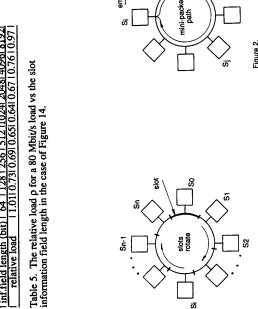
							_
exp. packet length (103bit)	0.5	1 2	I 7.1	1 14	1 28	1 56	1
relative load	1.00	10.71	10.65	10.64	10.63	10.63	ī

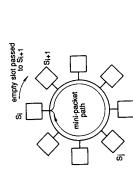
Table 3. The relative load ρ vs the expected packet length in the case of Figure 12.

inf. field length (bit)												
no of slots	33	32	16	8	Т	5	1 :	3	2	Т	1	٦

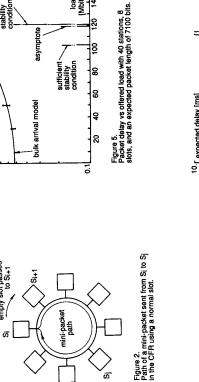
Table 4. The number of slots vs the slot information field length in the case of Figure 14.







ij



load [Mbivs] 120 140

9

20 40

load [Mbit/s]

8

6

sufficient stability condition

] = 90% conf.int. = simulations

Figure 7. Packet delay vs offered load with 10 stations. 7 slots, and an expected packet length of 7100 bits.

necessary stability condition

trans. rate = 140 Mbit/s no. of stations = 10 no. of stofts = 7 exp. pack. Heng. = 7100 bits stot inf. field = 512 bits

transmission rate = 140 Mbit/s no. of stations = 40 no. of stofis = 8 exp. packet length = 7100 bits slot inf. field = 512 bits

10.f expected delay [ms]

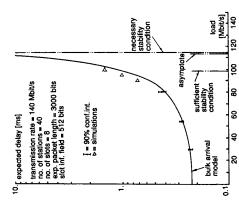
[= 90% conf.int.
• = simulations

10. F expected delay [ms]

sufficient stability condition

bulk arrival model

asymptote

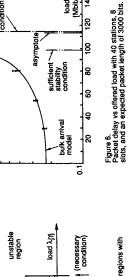


necessary stability condition

transmission rate = 140 Mbit/s no. of stations = 10 no. of stors = 2 exp. packet length = 7100 bits stot inf. field = 512 bits

10. Fexpected delay [ms]

I = 90% conf.int. P = simulations



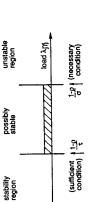


Figure 3. A MiGI1 model with bulk arrivals of mini-packets and server vacation periods.

mini- bulk packets arrivals

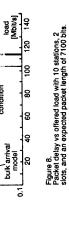
server vacation

individual packets arrivals

Figure 1. The slotted ring structure.

F2.7

Figure 4.
An illustration of stable and instable regions with respect to the load at Sj, j=0,...,n.



bulk arrival model

