(6)

Modelling of a Multi-Queue Polling System with Arbitrary Server Interrupts for the Idle-Slot-Concatenation Packet Switching Principle in a Hybrid CS/PS Node

Manfred N. HUBER, Paul J. KUEHN

Institute of Communications Switching and Data Technics
University of Stuttgart, F.R.G.

ABSTRACT

For future communication requirements in the inhouse area, a new system integrating circuit switching (CS) and packet switching (PS) is suggested. A substantial part of this system is a hybrid switching node, which is realized as a bus system using the Idle Slot Concatenation Principle for the integration of CS- and PS-traffic. The subject of this paper is the performance modelling and analysis of this switching node. Due to the packet access protocol, this system has been modelled as a polling system with arbitrary server interrupts and without any switching overhead. For some regular CS-time-slot distributions the mean waiting time of packets has been estimated by an inflated service time approximation, whereas for the arbitrary CS-time-slot distribution a new random service interrupt analysis, based on a Markov chain has been developed. The obtained results are validated by means of simulations.

1 INTRODUCTION

For future communication requirements in the inhouse area, we suggest a new system integrating circuit switching with variable and adaptable bandwidth as well as packet switching with high throughput rate [1]. This system consists of small CS/PS-LANs [2], CS/PS-PBXes and CS/PS-Links for interconnecting the elements.

These new LANs, on a ring basis, will only be installed in areas with higher communication requirements. Several of these LANs could be interconnected via these new PBXes. The mass of terminals, assumed to be of the ISDN-type, may also be interconnected through existing subscriber lines to these CS/PS-PBXes.

2 HYBRID SWITCHING NODE

2.1 Node Architecture

The CS/PS-PBX is realized as a bitparallel, synchronous bus system (Figure 1) with several access modules and one master module. ISDN-terminals, CS/PS-LANs and CS/PS-Links are connected to the bus system through special interface units within the access modules. The master module is responsible for clock and frame generation, CS-connection management and routing for virtual PS-connections.

2.2 CS/PS-Integration

The integration of CS- and PS-traffic is based on a synchronous pulse frame, partitioned into equal-sized time-slots (Figure 2). Each time-slot can either be used by CS-traffic or by PS-traffic and all time-slots, not being used by CS-traffic are concatenated to one remaining PS-channel (*Idle Slot Concatenation*).

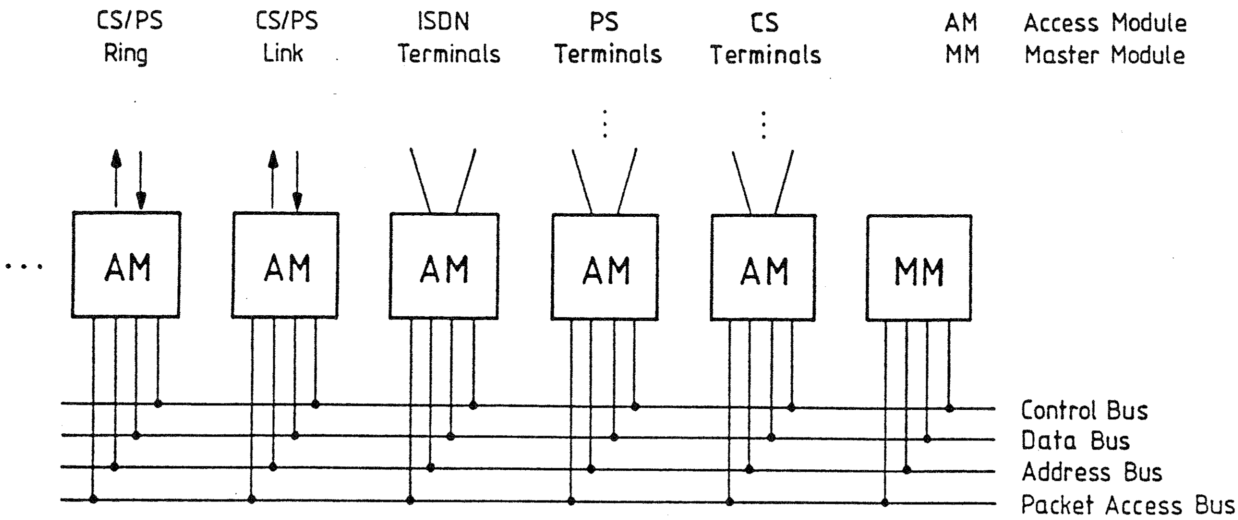


Figure 1: Structure of the CS/PS-PBX

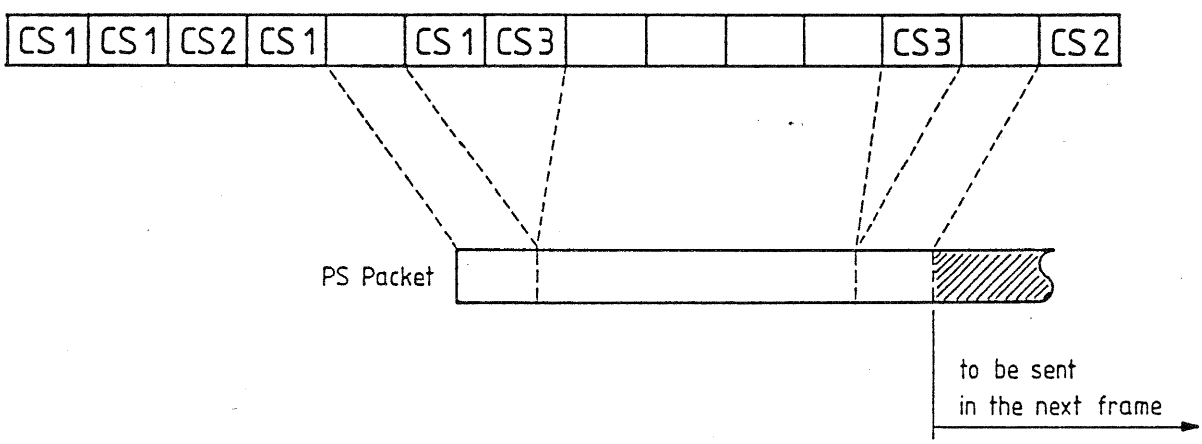


Figure 2: Integration of CS and PS by the Idle Slot Concatenation Principle

During a CS-call setup the master module assigns the time-slots necessary for the required bandwidth of that call to the access modules involved in that connection. All time-slots not being used by CS-connections are marked by the master module and then they are available for PS-traffic.

After arriving at an access module, the data packet waits for access to the PS-channel, which is controlled by the packet access protocol allowing fair access of all modules. Since the packet access protocol is implemented on a separate bus system (Packet Access Bus), it is possible that no switching overhead is necessary between two successive packets. Another advantage of this protocol and its implementation is that if a packet arrives during an idle period of the PS-channel, packet transmission can start within the next PS-time-slot.

Switching of data packets is similiar to *fast packet switching* [3]; flow control and error correction are done end to end. When a module gets access to the bus, the logical channel number in the header of the packet is translated into a new one and the packet is transmitted from the originating access module to the destination access module using PS-time-slots. This transfer may only be interrupted by CS-time-slots. All these necessary packet switching functions are implemented in hardware.

3 MODELLING

For proper performance evaluation of the hybrid switching node, it is necessary to develop a general queueing model, which is depicted in Figure 3.

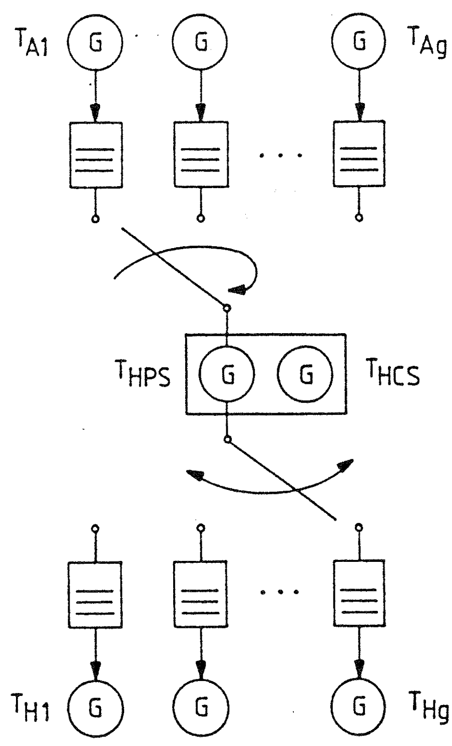


Figure 3: General Queueing Model of the CS/PS-PBX

The g modules connected to the bus system are represented as g transmitting queues with general independent arrival processes and interarrival times T_{Ai} and g receiving queues with general service times $T_{Hi}, i = 1, ..., g$. Transmitting and receiving buffers have finite capacities.

The bus system itself is modelled as a service unit with two phases representing PS-service time and CS-service time. The distribution of the data packet length is assumed to be general resulting in a generally distributed bus transmission time T_{HPS}. This service time is interrupted by single CS-time-slots or whole sequences of CS-time-slots, which may be arbitrarily distributed within the frame. The duration of the CS-interrupts is described by the discrete random variable T_{HCS}.

To achieve a fair performance for all data queues a simple cyclic bus access order with non-exhaustive service is used. Due to the implementation of the packet access protocol no overhead phase between two successive packets is necessary. The traffic model represents a polling system without switching overhead but with arbitrary service interrupts.

4 PERFORMANCE EVALUATION

4.1 General

In the early seventies research in performance evaluation for systems integrating CS- and PS-traffic began [4]. The base of these systems is also a synchronous pulse frame, but in contrary to our system, the frame has been subdivided in one CS- and one PS-part. The boundary between these two parts may be fixed (*fixed boundary*) or moveable (*moveable boundary*). For an integrated switching system with random distribution of CS-time-slots, but with additional switching overhead between two successive packets, an analytical approach was suggested in [5].

Performance evaluation for CS-traffic is done by considering a trunk group with full availability which is used by single-slot and multi-slot calls. The influence of the CS-time-slot distribution on the PS-traffic is studied for several CS-time-slot distributions:

packed distribution: all CS-time-slots are accumulated and packed within the first part of the pulse frame.
blocked distribution: equal-sized blocks of CS-time-slots are equidistantly distributed within the pulse frame.
equidistant distribution: CS-time-slots are equidistantly distributed within the whole pulse frame.
arbitrary distribution: CS-time-slots are arbitrarily distributed within the pulse frame.

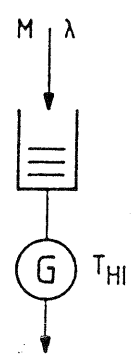
4.2 Simulation of the Hybrid Switching Node

By reasons of complexity for the performance evaluation of the hybrid switching node, a simulation tool based on the model in Figure 3 has been developed. Using this tool it is easy to obtain results in terms of waiting time in the transmission queue, total bus transmission time per packet and packet transfer delay as well as mean queue length, blocking probabilities for transmission and receiving queues for a hybrid switching node using the idle-slot-concatenation principle with the CS-time-slot distributions mentioned above.

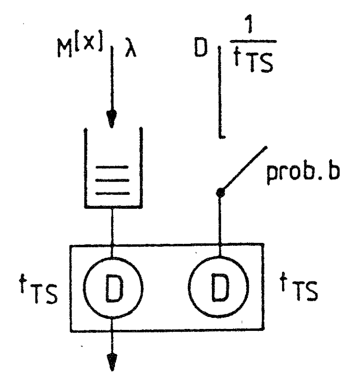
4.3 Inflated Service Time Approximation

For the determination of the mean waiting time per data packet, we can represent the polling system without switching overhead as a single server queueing system.

The traffic model for the *inflated service time approximation* with Poisson arrivals, which is used for the packed, blocked and equidistant CS-time-slot distributions with nailed up CS-connections is depicted in Figure 4a.



a)



b)

Figure 4: Simplified Queueing Models

For this analysis the following notations are used:

T_{HI} random variable of the inflated packet service time
h mean value of T_{HI}
λ total data packet arrival rate
t_{TS} duration of one time-slot
n_c number of successive CS-time-slots
n_p number of successive PS-time-slots
n average number of time-slots per packet (packet length)
c_h coefficient of variation of the packet length

First, the mean waiting time w_0 for packets meeting no other packets in the system on their arrival is calculated:

$$w_0 = \left(\frac{n_c}{n_c + n_p} \cdot (\frac{n_c}{2} + 1) + \frac{n_p}{2(n_c + n_p)}\right) \cdot t_{TS} \tag{1}$$

Then, the inflated service time per packet h is determined:

$$h = \frac{n_c + n_p}{n_p} \cdot n \cdot t_{TS} \tag{2}$$

Assuming a Markovian arrival process and infinite queue size, equations (1), (2) and the well known Pollaczek-Khintchine formula are used to calculate the mean waiting time w per data packet:

$$w = h \cdot \frac{\lambda \cdot h \cdot (1 + c_h^2)}{2 \cdot (1 - \lambda h)} + w_0 \tag{3}$$

4.4 Random Service Interrupt Analysis

Using the assumption that the CS-time-slots are distributed arbitrarily within the pulse frame, the calculation of the mean waiting time for packets with Poisson arrival in a queue of infinite size is based on the *random service interrupt analysis*. The appropriate traffic model is shown in Figure 4b.

For this analysis we consider an incoming packet of size n time-slots as a batch arrival with n packet-units (fraction of a packet fitting in exactly one time-slot). Therefore, the service time per packet-unit is always one time-slot. It is assumed that each time-slot is occupied by CS-traffic independently with probability b; this is equivalent to a PS-service interrupt with probability b. Based on an embedded Markov chain approach, we calculate the mean queue length in terms of packet-units from which the mean waiting time per packet is obtained.

The notations and symbols used in this analysis are:

t_{TS}	duration of one time-slot
λ	total data packet arrival rate
b	probability that a time-slot is occupied by CS-traffic
p_i	probability that i packet-units are within the system at the end of a time-slot
q_j	probability that j packet-units arrive during one time-slot
$G(z)$	generating function of the state probabilities at the end of a time-slot
$H(z)$	generating function of the transition probabilities
$H'(z)$	derivation of $H(z)$ with respect to z
$F(z)$	generating function of the batch size

Let the generating functions $G(z)$ and $H(z)$ be defined by

$$G(z) = \sum_{i=0}^{\infty} p_i z^i \tag{4}$$

$$H(z) = \sum_{j=0}^{\infty} q_j z^j \tag{5}$$

First let us consider the transition diagram with two sucessive regeneration points t_m^- and t_{m+1}^- (Figure 5). These two points are just before the end of time-slot m and its successive time-slot $m+1$. The left side of this figure represents the transitions from state 0 (0 packet-units within the system) at t_m^- to all other possible states at t_{m+1}^-. On the right side the possible transitions for all other states k ($k \geq 1$) are depicted.

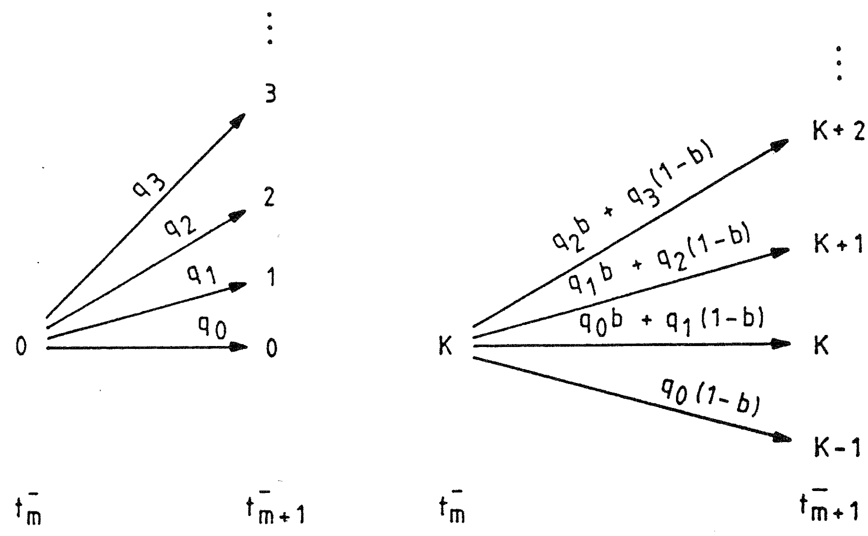


Figure 5: Transition Diagram

Using the state probabilities p_i, the transition probabilities q_j and the probability b that no packet service is possible, we can calculate the generating function $G(z)$:

$$G(z) = H(z) \cdot (p_0 + p_1 bz + p_1(1-b) + p_2 bz^2 + p_2(1-b)z + \ldots) \tag{6}$$

After some transformations we get

$$G(z) = H(z) \cdot \left(bG(z) + \frac{1-b}{z} G(z) - bp_0 - \frac{1-b}{z} p_0 + p_0 \right) \tag{7}$$

and with the property that $G(1) = 1$, used for eliminating p_0 we obtain

$$G(z) = H(z) \cdot \left(1 - \frac{H'(1)}{1-b}\right) \cdot \left(\frac{(1-b)\cdot(1-\frac{1}{z})}{1 - H(z)\cdot(b - \frac{b}{z} + \frac{1}{z})} \right) \tag{8}$$

The generating function $H(z)$ for the transition probabilities takes the batch arrival process into account:

$$H(z) = e^{-\lambda t_{TS}\cdot(1-F(z))} \tag{9}$$

This result leads to the generating function $G^*(z)$ of the state probabilities at an arbitrary point within the time-slot by the following integration

$$G^*(z) = \frac{1}{t_{TS}} \int_0^{t_{TS}} G(z) \cdot H^*(z)\, dt \tag{10}$$

and with

$$H^*(z) = e^{-\lambda t\cdot(1-F(z))} \tag{11}$$

we get

$$G^*(z) = G(z) \cdot \frac{H(z) - 1}{\lambda t_{TS}\cdot(F(z) - 1)} \tag{12}$$

By the derivation of $G^*(z)$ and with $z = 1$ we obtain the mean number of packet-units in the system at an arbitrary instant within a time-slot. Now we are able to calculate the mean waiting time per packet.

For packets with the constant length of n time-slots, we get

$$F(z) = z^n \tag{13}$$

and with equ. (8, 9, 13) and the derivation of equ. (12) the mean number of packet-units in the system is calculated by

$$G^{*'}(1) = \frac{\rho}{2} \cdot \left(\frac{1 - \rho + n}{1 - \rho - b} + 1 \right) \tag{14}$$

$\rho = H'(1) = \lambda n t_{TS}$ is the carried packet load.

Now the mean number of packet-units Ω in the queue is determined by

$$\Omega = G^{*'}(1) - \rho \tag{15}$$

Then, with equ. (14) the mean waiting time w per packet, which is equivalent to the interval between batch arrival and start of service of the first packet-unit of the batch, is given by

$$w = \Omega \cdot \frac{t_{TS}}{1 - b} + t_r \tag{16}$$

t_r is the residual service time of the time-slot. Due to the constant time-slot duration we obtain the following result for the mean waiting time w:

$$w = \frac{t_{TS}}{2} \cdot \left(\frac{\rho \cdot (n + b)}{(1 - b) \cdot (1 - \rho - b)} + 1 \right) \tag{17}$$

5 RESULTS

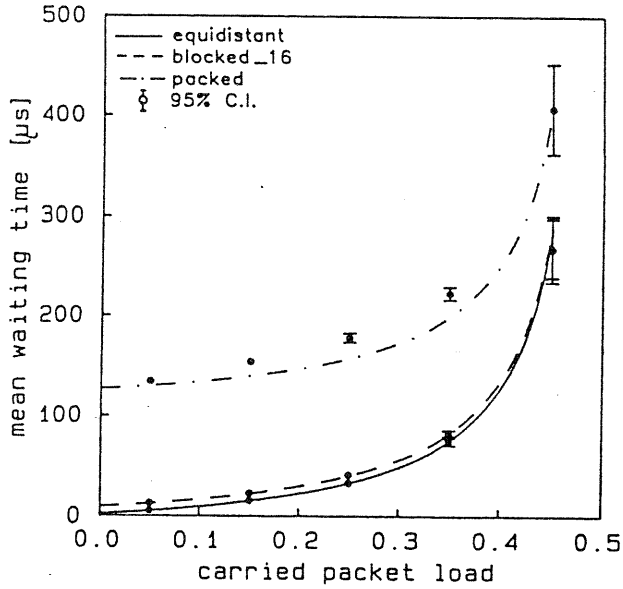
We have studied a CS/PS-node with the total bandwidth of 32 Mbps. The pulse frame of 1 ms is subdivided into 512 time-slots of 64 bits each. It has been assumed that the packet arrival process is Poisson, the data packet length is constant 1024 bits (16 time-slots).

Figure 6a depicts the mean waiting time w versus the packet load for the CS-time-slot distributions packed, equidistant and blocked with block-size of 16 time-slots for a system with 50% CS-load. The packet load is normalized by the total bandwidth of the switch. These results are obtained by the inflated service time approximation and are validated by simulations. The equidistant distribution of the CS-time-slots shows the best performance, because the waiting time w_0 is the smallest.

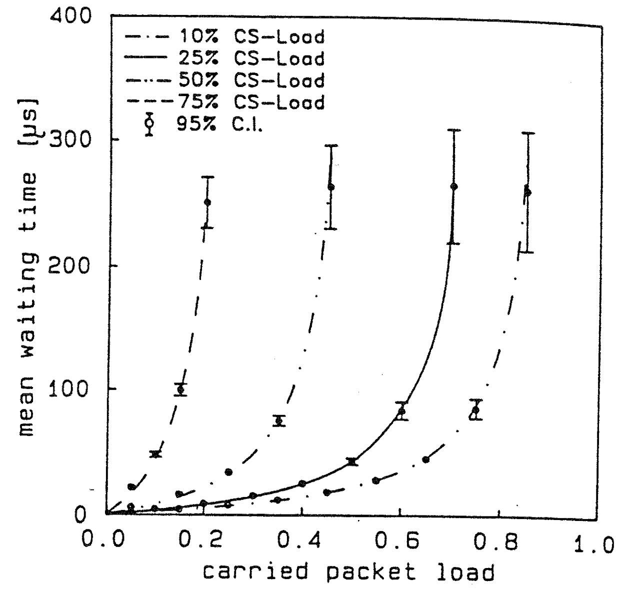
Figure 6b shows the mean waiting time for different CS-loads for this system using the arbitrary CS-time-slot distribution. These results are obtained by the random service interrupt analysis and yield good accuracy compared to the simulation results.

6 CONCLUSION

In this paper, the architecture of a hybrid switching node and its appropriate queueing model were presented. For the integration of CS and PS the idle slot concatenation principle was used. Performance evaluation for the general case was done by simulation. For some regular CS-time-slot distributions the mean waiting time for a packet was estimated by the inflated service time approximation, whereas for the arbitrary CS-time-slot distribution the random service interrupt analysis was developed. The results, obtained by these analytical approaches, were validated by means of simulations.



a)



b)

Figure 6: Mean Waiting Time versus Carried Packet Load

ACKNOWLEDGEMENTS

The authors would like to thank W. Fischer for helpful discussions and G. Roessler for his programming efforts.

This work is partly supported by the German Research Council (DFG).

REFERENCES

[1] E.-H. Goeldner, M.N. Huber, P.J. Kuehn: Architecture of a New Inhouse Communication System Providing Integrated Circuit and Packet Switching. Communication in Distributed Systems, Aachen, Feb. 1987, pp. 338-350

[2] E.-H. Goeldner: An Integrated Circuit/Packet Switching Local Area Network - Performance Analysis and Comparision of Strategies. Computer Networks and ISDN Systems, Vol. 10, Oct. 1985, pp. 211-219

[3] J.S. Turner: Design of an Integrated Services Packet Network. Ninth Data Communications Symposium, ACM Sigcom computer communication review, vol. 15, no. 4, Sept. 1985, pp. 124-133

[4] K. Kuemmerle: Multiplexor Performance for Integrated Line- and Packet-Switched Traffic. Proceedings ICCC '74, Stockholm, pp. 507-515

[5] D.R. Manfield: Transport Delays in a Hybrid Switching System with Polling. ICC '85, Chicago, June 1985