Dynamic Real-Time Channel Establishment in Multiple Access Bus Networks

Anita Mittal[1], G. Manimaran[2], and C. Siva Ram Murthy[1]

[1] Dept.of Computer Science and Engineering,
Indian Institute of Technology, Madras 600 036, INDIA
{naman@hpc., murthy@}iitm.ernet.in
[2] Dept. of Electrical and Computer Engineering,
Iowa State Univ., Ames, IA 50011, U.S.A.
gmani@shasta.ee.iastate.edu

Abstract. Real-time communication with performance guarantees is becoming very important to many applications, like computer integrated manufacturing, multimedia, and many embedded systems. Though several real-time protocols have been proposed for the multiple access bus networks, there is no guarantee based protocol which addresses the problem of integrated scheduling of dynamically arriving periodic and aperiodic messages. In this paper, we propose two guarantee based protocols, EDF and BUS protocols, which address this problem. In the simulation studies, the performance metrics, success ratio (measure of schedulability) and channel utilization are used to study the performance of the two protocols. It is observed that the EDF protocol offers higher schedulability as compared to the BUS protocol for periodic messages, while the BUS protocol offers better channel utilization for aperiodic messages as compared to the EDF protocol.

1 Introduction

Predictable inter-task communication is of prime importance in real-time systems, because unpredictable delays in the delivery of the messages can affect the completion time of the tasks participating in message communication. Also in multimedia applications, timely delivery of messages is required to guarantee the bounds on delay and delay jitter in delivering video/audio frames consistent with human perception. In order to support real-time communication with performance guarantees, real-time channels[1] are established with the specified traffic characteristics and Quality of Service (QoS) requirements.

Distributed real-time applications are implemented using either real-time multi-hop networks or real-time multiple access networks. While schemes for channel establishment and run-time scheduling algorithms have been proposed for real-time multi-hop networks [1, 2], not much work has been done for real-time multiple access networks. Multiple access bus networks are being increasingly used, for real-time applications, because they are simple, economical, and

[1] A real-time channel is a unidirectional virtual circuit with specified QoS guarantees.

their propagation delays are small. In this type of network, scheduling is the responsibility of Medium Access Control (MAC) protocol, which arbitrates access to the shared medium and determines which message is to be transmitted at any given time.

Several schemes for real-time communication on multiple access networks have been proposed [3, 4], but they generally belong to the best effort category. In this paper, we propose two guarantee based protocols for dynamic real-time channel establishment with support for scheduling of aperiodic messages in multiple access bus networks.

2 Background

2.1 Desirable Properties of Real-time MAC Protocol

A real-time MAC protocol should possess the following desirable properties.

1. Support Real-time Channel Establishment: A real-time channel is established by reserving the resources as per the specified traffic characteristics of the periodic message, if the admission of this connection does not jeopardize the guarantees given to the already admitted channels. The protocol should have an associated admission test which is to be performed while admitting a periodic message stream.

2. Provide Bounded Response Time: In many real-time applications such as air-traffic control and multimedia, periodic message requests for channel establishment arrive dynamically. The protocol should provide bounded response time to a channel establishment request (i.e, time taken to decide acceptance/rejection of a request based on admission test should be bounded) so that some alternate recovery action can be initiated in case the request is rejected.

3. Integrated Scheduling of Periodic and Aperiodic Messages: Besides these periodic messages which require hard real-time guarantees, the protocol should also support scheduling of aperiodic messages which may have soft deadlines or no deadlines at all.

4. Distributed in Nature: Since the channel is shared by a number of nodes in the multiple access networks, the nodes need to cooperate and coordinate with one another (i) for transmitting channel establishment requests, (ii) for admission test and resource reservation, and (iii) for transmission of messages on the real-time channels so that all the messages meet their deadlines. These functions can be performed either in centralized or in distributed manner. Centralized scheme is prone to single point failure and also incurs substantial overhead for giving channel access rights to the nodes. Therefore distributed scheme is preferred wherein every node participates in decision making in a decentralized manner.

2.2 Prior Work

A guarantee based protocol for scheduling of aperiodic messages has been proposed in [5]. This protocol does not support scheduling of periodic messages. (ii)

A centralized guarantee based protocol for scheduling of periodic messages in multiple access bus networks has been proposed in [6]. It does not support integrated scheduling of periodic and aperiodic messages and also suffers from the disadvantages of a centralized protocol. (iii) A centralized scheduling algorithm for scheduling of periodic and aperiodic messages has also been proposed for the backplane bus of a workstation [7]. In the workstation, instant global knowledge of all the messages contending for the backplane bus is available through separate hard-wired control channels. Hence this scheduling algorithm cannot be used as such for multiple access networks. Thus we see that there is no protocol which addresses all the four requirements of a guarantee based real-time MAC protocol. This motivated us to propose guarantee based protocols for integrated scheduling of periodic and aperiodic messages in multiple access bus networks.

3 Proposed Protocols

3.1 Channel Model

- A population of N nodes share the channel.
- Channel access is slotted. Transmission can start only at the beginning of a slot. The clocks of all the nodes are synchronized.
- The slot time is equal to the maximum end-to-end round trip delay for a bit.
- Carrier sensing is instantaneous. Each node is capable of detecting an idle slot, a successful transmission, and a collision as in CSMA based protocols.
- All the nodes monitor the activity on the shared channel.

3.2 Proposed Protocols

We have proposed two protocols EDF protocol and BUS protocol. These protocols are distributed in nature with each node executing the same algorithm. The protocol alternates between two phases: reservation phase and transmission phase. The proposed protocols have the following components:

- Request Server (RS): This is a periodic server used for collecting periodic and aperiodic message requests and for giving bounded response time to channel establishment requests. During the activation of RS, in the reservation phase, message notifiers are transmitted. The notifier contains information such as, type of request - periodic/aperiodic, service time, period/deadline, and node number of the message source. One slot is required for transmission of a message notifier. Since all the nodes monitor the channel, and the channel is a broadcast medium, all the nodes get to know about the message traffic at the other nodes through these message notifiers. The preorder DCR protocol [8] is used for transmitting message notifiers. Preorder DCR protocol is an efficient protocol wherein when the load is light, say, only one message request is ready in the entire network, then only one slot of RS is used and the remaining slots can be used for transmission of messages. When the load is heavy, collision is resolved deterministically to bound the worst case channel access delay, giving every node a chance to broadcast at least one message notifier in one RS period.

– Aperiodic Server (AS): This is also a periodic server used for servicing aperiodic messages. The service could be either guaranteed service for which the guarantee based DCR protocol [5] can be used or could be a best effort service. In the best effort service category, the service policy can be either fairness oriented, where the AS bandwidth is fairly given to all the contending nodes, or aimed at maximizing throughput of aperiodic messages. The percentage of bandwidth reserved for aperiodic traffic is application dependent.
– Run-time Scheduling: The run-time scheduling algorithms, for the EDF and BUS protocols, are based on the Earliest Deadline First (EDF) algorithm [9] and backplane bus scheduling algorithm [7], respectively. These algorithms are used during the transmission phase of the protocols.
– Admission Tests: There is an associated admission test with each protocol for admission control of real-time channel establishment requests.

3.3 Protocols Description

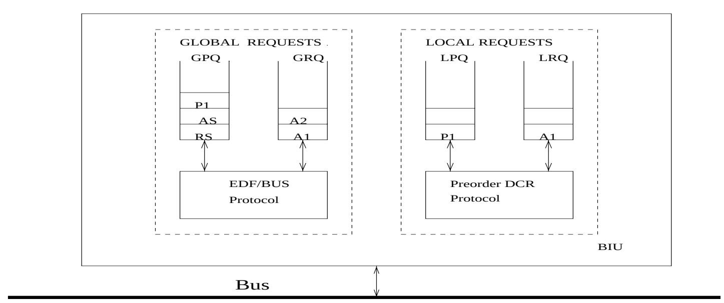


Fig. 1. A node in the multiple access network

Each node maintains two queues (local) for storing the local periodic (LPQ) and aperiodic message (LRQ) notifiers, as shown in Figure 1. A message notifier has to be broadcast before the transmission of the message. Each node also maintains two global queues, aperiodic queue (GRQ) for storing the aperiodic message notifiers which have been broadcast successfully on the bus and periodic queue (GPQ) for storing the notifiers of admitted real-time channels (periodic messages). The RS and AS notifiers, are put in the GPQ of each node. In the EDF/BUS protocol, channel access rights are given to the real-time channels for say, n slots, based on the EDF/BUS run-time scheduling algorithm. Each node executes the following protocol once the channel access is granted to a real-time channel.

Protocol($RTchannel$,n) /* $RTchannel$ - the channel which has access rights for n slots */

1. If $RTchannel$ = RS,
   Repeat steps (a)-(c) until a slot goes idle or RS service time is exhausted.

   (a) Transmit message notifiers from the local queues with priority to LPQ over LRQ as per the preorder DCR protocol.
   (b) If notifier transmission is successful and the notifier is for an aperiodic message then insert the notifier in the GRQ.
   (c) else if notifier transmission is successful and the notifier is for a periodic message then
      i. $result$ = Admit(notifier) /* Perform admission test using the EDF/BUS protocol admission test */
      ii. If $result = success$ then
         A. Insert the notifier in GPQ.
         B. Reserve resources for this real-time channel and update system utilization information.
      iii. else reject the request.
2. If $RTchannel$ = AS,
   (a) Grant channel access rights to messages in GRQ based on predetermined aperiodic message service policy/protocol.
3. If $RTchannel = P_i$, /* real-time channel i */
   (a) Grant channel access rights to the node which is the source of the real-time channel P_i.

When a node wishes to tear down the real-time channel, the node transmits channel termination request as the last message on this channel. All the nodes then update their GPQ and channel utilization information accordingly.

When none of the periodic messages are using the channel (no message is ready in the GPQ), then aperiodic messages (if any) are serviced. In the event of GRQ also being empty, (i.e, when the system load is light) the RS is activated, thus resulting in very less average response time to channel establishment requests compared to RS period. It can be noted that, when the system load is light, these protocols behave like a contention based protocol, since RS follows a CSMA/CD based DCR protocol. When the system load is heavy, the channel is granted to the admitted requests without any collisions based on the EDF/BUS policy. Thus the protocols adapt well to the system load.

3.4 EDF Protocol

The nodes follow the EDF algorithm of serving the earliest deadline first message in the global periodic queue. EDF [9] is a dynamic priority scheduling algorithm that always schedules the message with the earliest deadline, preempting any currently scheduled message if required. Preemption of messages takes place at packet boundaries.

Admission Test

A periodic message is admitted if it satisfies the following condition:

$$util_{curr} + util_{new} \leq 1$$

where $util_{curr}$ is the current utilization of the channel and $util_{new}$ is the channel utilization requirement of the new message, given by the expression:

$$util_{new} = \frac{\text{maximum message size/slot size}}{\text{period of the message/slot time}(\tau)}$$

3.5 BUS Protocol

The BUS protocol is based on backplane bus scheduling algorithm [7] and uses RS for request collection and AS for servicing aperiodic messages. In a multiple access bus network, the periodic traffic arrives at each node independently unlike in backplane bus where it is centralized. Moreover, there is no central Bus Arbiter (BA) which performs access arbitration based on channel reservation of the admitted requests. Also there are no control lines to convey bus request and bus grant information. In the BUS protocol, the nodes follow a periodic service policy similar to the one executed by the BA with S slots and cycle time (period) $S\tau$. The service times and periods of the RS and AS are converted to the number of slots required per service cycle to R_s and A_s, respectively.

Admission Test

A periodic message stream with message size M_i and period T_i is admitted if the following condition holds:

$$c_i + Q \leq S - \alpha$$

where c_i is the smallest integer that satisfies

$$M_i \leq c_i(\lceil T_i/(S * \tau)\rceil - 2)$$

Q is the number of time slots in a cycle reserved for periodic messages, and α denotes the number of slots dedicated for random (aperiodic) messages. α is a system parameter that should be changed only when a change in the fraction of allowable periodic traffic is desired.

The algorithm executed by each node is:

Algorithm BUS(R_s,A_s,S,Q) /* Initially n and q counters are set equal to S and Q, respectively */

1. If $n = S$, service RS for R_s slots or until a slot goes idle whichever is earlier.
2. If $n > q$ and aperiodic queue is not empty, service AS for $(n - q)$ slots or until aperiodic queue becomes empty whichever is earlier.
3. If $q > 0$ and $n > q$, service periodic requests and decrement q accordingly.
4. If $n > 0$ and $q = 0$, service AS if aperiodic queue is not empty otherwise service RS till a request arrives.
5. if $n = 0$, set n and q to S and Q, respectively.

3.6 Protocols Comparison

Schedulability of Periodic Messages: The EDF protocol offers higher schedulability than the BUS protocol. This is because the admission test of BUS protocol slightly over allocates bandwidth to periodic messages and as a result of this over allocation lesser number of real-time channels can be established. However, at run time this over allocated channel bandwidth is made available for servicing aperiodic messages.

Response Time to Aperiodic Messages: The BUS scheduling scheme gives fast response to aperiodic messages without jeopardizing the guarantees given to admitted periodic messages. Aperiodic messages (if any) are serviced when there are no periodic messages to be transmitted, by activating the AS for idle channel duration. In the BUS protocol, the idle slots are uniformly distributed since the service cycle time is much less than the deadline of aperiodic messages, thus resulting in better channel utilization for aperiodic messages. In the EDF protocol, besides lack of over allocated bandwidth being made available for aperiodic messages, occurrence of idle channel instances, when there are no periodic messages to be transmitted, is not uniform (the idle channel bandwidth is made available in bursts) and hence aperiodic messages cannot utilize the idle channel bandwidth efficiently.

Granularity of AS: In EDF protocol, though the idle channel time cannot be given uniformly, the AS time can be made available uniformly by changing the granularity of the AS, i.e., the period and service time would be divided by the same constant factor. For example, service time of 4 units and period 10 units can be transformed into a server with service time 2 units and period 5 units. Similarly, the granularity of RS can be reduced to improve response to message requests.

The differences mentioned above between the two protocols can be clearly observed in the simulation results.

4 Simulation Study

The simulation parameters are given in Table 1. The slot time is assumed to be 5 micro seconds and slot size 64 bytes.The bandwidth reserved for AS/RS can be generalized as $(C_s/m_{as/rs})/(P_s/m_{as/rs})$, where C_s and P_s gives the service time and period of the AS/RS, respectively and $m_{as/rs}$ parameter (m is an integer, $m \geq 1$) decides the granularity. As $m_{as/rs}$ increases, granularity decreases which means decrease in period and service time of the AS. When $m = 1$, granularity is maximum.

parameter	explanation	distribution	value/range
N	number of nodes		16
RS_BW	bandwidth reserved for RS		1%
AS_BW	bandwidth reserved for AS		10%
PerMesgSize	periodic message size (bytes)	uniform	4500-6400
Period	period of messages	uniform	4000-6000
AperMesgSize	aperiodic message size (bytes)	uniform	1000-5000
AperLaxity	laxity of aperiodic messages	uniform	20-100
Instances	number of instances of periodic messages	uniform	30-50
PLoad	periodic load	Poisson	0.1-3
ALoad	aperiodic load	Poisson	0.5-3

Table 1. Simulation parameters

4.1 Performance Metrics

1. Effective Channel Utilization, ECU, defined as
   $ECU = \frac{\text{total time units channel used for transmitting messages}}{\text{total time units simulated}}$
   This metric when applied to periodic, aperiodic messages and RS individually gives the metrics PCU, ACU, and RCU respectively. RCU captures the overhead for collecting requests and this should be small.
2. Success Ratio, SR, defined as
   $\text{SR} = \frac{\text{number of messages serviced}}{\text{total number of messages generated}}$
   This metric when applied to periodic and aperiodic messages individually gives the metrics PSR and ASR, respectively.

4.2 Simulation Results

Variation of system load: Figure 2 shows the channel utilization by periodic and aperiodic messages and RS for varying system load for both the BUS and EDF protocols.
Effect on RCU: As the load increases, RCU decreases. This is because the RS is active when the channel is idle and hence when load is low RCU is high. When system becomes overloaded, idle time decreases and hence RCU decreases, and approaches the reserved bandwidth utilization bound.
Effect on PCU: The PCU for the EDF protocol is slightly higher than that of BUS protocol. Reason is the same as mentioned earlier. It may be noted that at low loads PCU is same for both the protocols, since channel bandwidth exceeds the total bandwidth demand of all the message requests.
Effect on ACU: ACU increases with increasing load up to a point and then starts decreasing. Aperiodic messages use the bandwidth reserved for them and also use the channel when it is idle. When the periodic load increases, PCU increases and hence idle time decreases. Therefore aperiodic message utilization decreases. ACU for the BUS protocol is higher than that of the EDF protocol. Reason is the same as mentioned in Sect. 3.6.

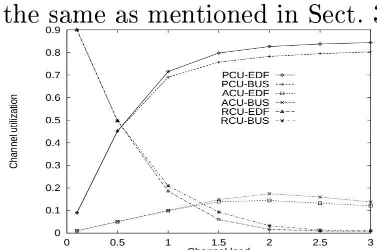


Fig 2. Effect of system load

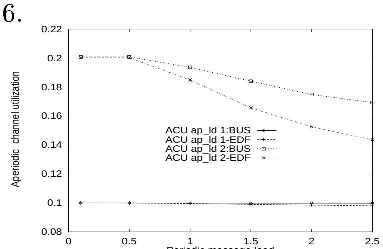
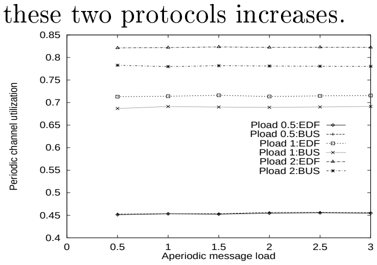


Fig 3. Effect of periodic load

Variation of periodic load: Figure 3 shows the channel utilization by aperiodic messages for varying periodic message load. ACU has been plotted for fixed aperiodic loads of 1 and 2. As the periodic load increases, ASR remains same as long as the aperiodic load does not exceed the reserved AS capacity. When aperiodic load is twice the AS capacity, then ACU decreases with increasing periodic load. When periodic load is negligible (0.1) all the aperiodic messages get serviced. This is because the entire channel bandwidth is available for RS and AS. As periodic load increases, the idle time decreases and hence ACU decreases.

Variation of aperiodic load: Figure 4 shows the PCU for varying aperiodic message load plotted for fixed periodic loads of 0.5, 1 and 2. We see that PCU is independent of aperiodic load. This is expected since the presence or absence of aperiodic requests does not affect the bandwidth reserved for periodic messages and hence insensitivity of metric PCU to aperiodic load. When the periodic load is 0.5 both the EDF and BUS protocols have the same PCU, but as the load increases, the EDF protocol performs better and the difference in PCU between these two protocols increases.



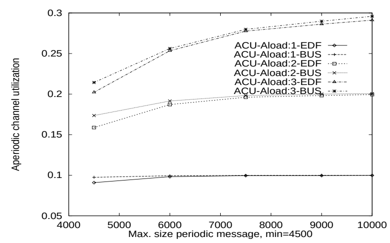


Fig 4. Effect of aperiodic load Fig 5. Effect of stochastic periodic message size

Effect of stochastic nature of periodic message size: Figure 5 shows the effect on ACU by variation in the range (difference between the maximum and minimum size of messages of a periodic message stream) of periodic request message size. It is to be noted that the admission test is based on the maximum message size of the periodic message request, and hence some of the reserved bandwidth is not used when instances of that periodic message stream have shorter message sizes. This bandwidth is made available for servicing aperiodic messages and hence the aperiodic channel utilization increases. The implication of this result is that the bandwidth reservation for aperiodic messages needs to be less, if the periodic message sizes are stochastic in nature.

Effect of server granularity: Here, we examine the effect on ASR, by varying the AS granularity for the EDF protocol.

(i)Effect of RS granularity: Figure 6 studies the influence on ASR by variation in AS granularity for RS granularity m_{rs} values equal to 1, 12, and 24. The curves are plotted for periodic and aperiodic message load of 1, aperiodic message size of 1000 bytes, and laxity value of 20. It can be observed that RS granularity significantly effects ASR. This is because as the granularity of RS decreases (m_{rs} increases), responsiveness to newly arrived requests increases. Since the period of RS becomes small, the number of new requests in that period decreases, decreasing the probability of collisions and as a result the efficiency of RS increases. ASR is maximum when the RS granularity is least i.e, $m_{rs} = 24$. The ASR increases with decrease in granularity, till $m_{as} = 12$ and then decreases for further decrease in granularity ($m_{as} = 24$). This is because the number of slots required for transmission of an aperiodic message equals the slots serviced per AS period when $m_{as} = 12$, and on further increase in m_{as} no aperiodic message is serviced in entirety in one period of AS and hence the decrease in ASR. It is to be observed that ASR is maximum for $m_{rs} = 24$ and $m_{as} = 3$.

Appropriate choice of these two parameters needs to be made depending on the characteristics of aperiodic messages.

(ii)**Effect of laxity**: Figure 7 shows the effect on ASR for variation in AS granularity for different laxities of aperiodic messages. It can be seen that the granularity effect is almost negligible for higher laxities. This is because with increase in laxity of aperiodic messages, the deadline of messages becomes larger than the highest granularity of AS with $m_{as} = 1$. The RS granularity effect can still be observed though it is moderate for higher laxities.

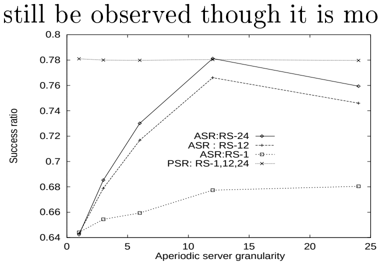


Fig 6. Effect of request server granularity

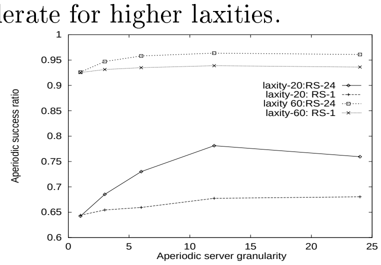


Fig 7. Effect of laxity

5 Conclusion

We proposed the EDF and BUS protocols, for integrated scheduling of dynamically arriving periodic and aperiodic messages in multiple access bus networks. The proposed protocols can be extended to exploit the variable QoS requirements of messages to improve schedulability and fault tolerance of the system.

References

1. D.Ferrari,and D.C.Verma,"A Scheme for real-time channel establishment in wide area networks",*IEEE J.on Selected Areas in Commun.*, vol.8, no.4, 368-379, 1990.
2. D.D.Kandlur, K.G.Shin, and D.Ferrari, "Real-time communication in multi-hop networks", *IEEE Trans. on Parallel and Distributed Systems*, vol.5, no.10, 1044-1055, 1994.
3. J.F.Kurose, M.Schwartz, and Y.Yemini, "Multiple- access protocols and time constrained communication", *ACM Computing Surveys*, vol.16, no.1,43-70, 1984.
4. N.Malcom and W.Zhao, "Hard real-time communication in mutiple-access networks", *Real-Time Systems*, no.8, 35-77, 1995.
5. S.Norden, G.Manimaran, and Siva Ram Murthy, "Dynamic Planning based protocols for real-time communication in multiple access networks", *Technical Report, Dept.of Computer Science and Engg., Indian Institute of Technology Madras*, 1997.
6. C.C.Chou, and K.G.Shin,"Statistical real-time channels on multi-access bus networks",*IEEE Trans. on Parallel and Distributed Systems*,vol.8,no.8,769-780, 1997.
7. S.H.Khayat, and B.D.Bovopoulos, "A simple and efficient bus management scheme that supports continuous streams", *ACM Trans. on Computer Systems*, vol.13, no.2,122-140,1995.
8. V.Upadhya, "Design and analysis of real-time LAN protocols", *M.S.Thesis, Dept. of Computer Science and Engg., Indian Institute of Technology, Madras*, 1996.
9. C.L.Liu, and J.W.Layland, "Scheduling algorithms for multiprogramming in a hard-real-time environment", *Journal of the ACM*, vol.20, no.1,45-61, 1973.