IEEE 802.11
Wireless LAN Medium Access Control and Physical Layer Specifications

A Distributed Time Bounded Service

10 January 1994

Philip Rakity, Larry Taylor
Apple Computer Europe, Inc.
Center de Recherche et Developpement, Le Toronto
54 route de Sartrouville
78230 Le Pecq
France
Tel: (33-1) 34 80 30 20
Fax: (33-1) 39 76 35 55

Kerry Lynn
Apple Computer, Inc.
One Infinite Loop, MS: 301-4J
Cupertino, CA 95014
(408) 974-2394

Abstract

The 802.11 PAR requires that the standard support a time bounded MSDU delivery service. The current foundation provides this service optionally, in certain configurations, through a point coordination function. We show that a PCF is not required for this purpose and is in fact undesirable for a variety of reasons.

Introduction

Time Bounded Service (TBS) has been made an integral (i.e. non-optional) feature of Hiperlan. The extent to which a protocol architecture has the ability to support traffic which has desired time constrained relationships between the generation and delivery of its packet flux determines its applicability to the transmission of audio (e.g. coded speech or music) or moving images (e.g. video). In addition, certain data will have value for limited time and hence must be delivered to its destination before expiry (e.g. alarm or signal information)[1]. We feel that there are compelling reasons to provide TBS to any 802.11 station and describe a mechanism by which this can be accomplished.

[1]There are at least two classes of TBS data sources. Fixed Rate sources generate data at a given [constant] rate, Variable Rate sources do not. It should be noted that many TBS data sources, including fixed rate sources, contain scalable information e.g. a video codec generates high and low frequncy information. The significance of the data may be exploited in adapting to varying communications conditions.

There have been several papers presented which discuss aspects of this subject and these papers generally fall into one of two classes:

1) TBS via synchronization and centralized channel access scheduler [2,6]
2) TBS via channel access prioritization [1,6,7]

The components contributing to delay and an evaluation of delay variance have been described in [5].

Time Bounded Service Constraints

The architecture of any TBS system must address the following problems:

1) Management of channel access such that the constraints on TBS data are respected. TBS constraints are specified in terms of Quality of Service (QoS) parameters Transit Delay (time from submission to delivery) and Delay Variance (time jitter on Transit Delay)[2]

2) Controlling channel access such that overload conditions are avoided (minimized). For TBS this means, in general, measuring the TBS load against an estimate of the TBS capacity. Channel access rights are controlled by this measure.

In the remainder of this document, these general TBS issues are investigated in terms of centralized and distributed approaches. As far as possible, the specific radio communications issues have been isolated from the time bounded transfer issues. Section 1 describes TBS in terms of Time Division Multiplexing (TDM) or scheduling via bandwidth reservation, Section 2 describes TBS in terms of Decentralized Priority channel access. Observations on each technique are made to assist comparison .

1. Time Division Multiplexing

An approach to TBS has been described in [2] in some detail. In brief, TBS and Async Data Service (ADS) are multiplexed on the same channel via an access prioritization and synchronization mechanism. TBS connections are negotiated via a centralized scheduler which must not only manage TBS capacity but also schedule TBS channel access.

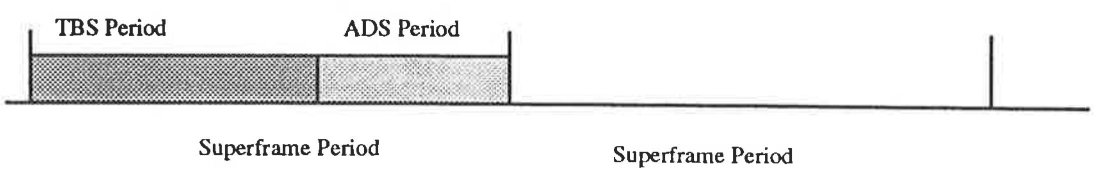

Figure 1. Multiplexed TBS / ADS Superframe Structure

[2]In a radio communications environment, owing to fading and interference, these characteristics cannot be guaranteed.

Channel access is managed during a periodic superframe time which consists of a TBS period and an ADS period. The ADS period is long enough to guarantee access for at least one non-TBS source in each superframe. Access during the TBS Period is managed by a scheduler. The following description applies to the TBS Period part of the repetitive superframe periods.

For the TBS Period, consider the follow example:

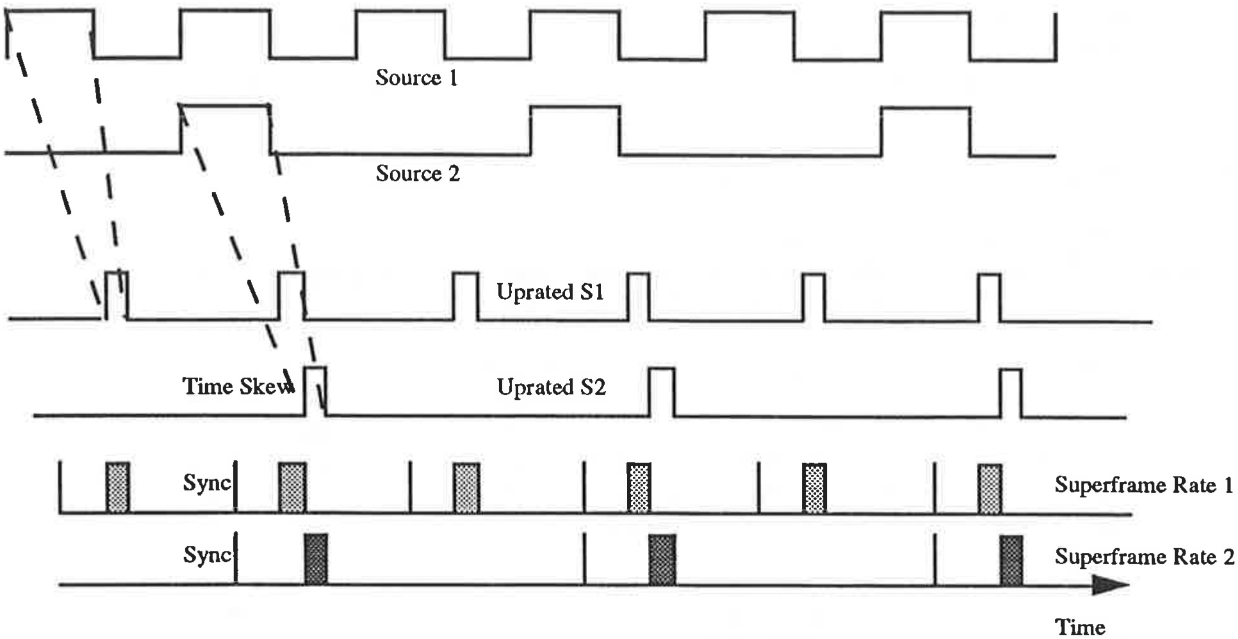

Figure 2. Time Division Multiplexed Scheduler

Figure 2 illustrates a simple case of TDM scheduler complexity. The problem is represented by two coincident data sources of different rates. In this simplified example, the data rates are integer multiples of each other - but in general they could have any values of either packet size (including variable packet size) and repetition rate. The source data is uprated to the transport data rate (depicted as compressed in time).

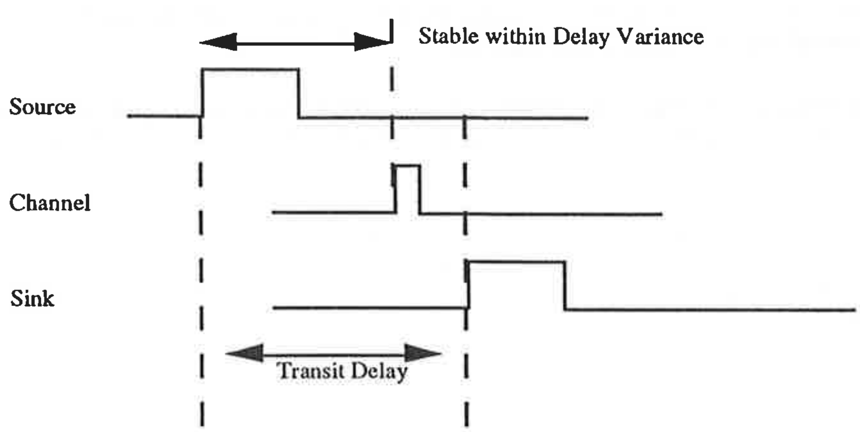

Figure 3. Scheduler Constraints

The scheduler's job is to find repetitive relative positions in time such that each TBS data source can always be delivered within its constraint of Delay Variance. Figure 3 shows the timing relationship between the scheduled and delivered data.

The scheduler must allocate positions to all occurrences of TBS sources for the time interval of one superframe at the slowest superframe repetition rate i.e. Superframe Rate 2 in Fig. 2. The example shows that within one superframe at the slowest superframe rate, one slowest rate data packet must be scheduled together with two faster rate packets.

A problem occurs when the data streams are originally coincident. The scheduler must skew all but one coincident source so that they can be allocated positions in time. This implies a negotiated start time for any transfer on the skewed streams or the constraints of at least the first packet may not be respected.

This example allows us to make the following observations:

- The complexity of the scheduler's task is proportional to the number of sources, the number of different source repetition rates and the accuracy with which the superframe repetition rate can be scheduled.

- Scheduler complexity for multiple hop paths will be additive.

- The number of TBS sources which can be supported is dependent on their total reserved load irrespective of their actual generated load. (*This scheme is efficient for fixed rate TBS sources.*) If the actual load is less than the reserved load, no new TBS sources can be serviced as the scheduler cannot be adaptive.

- TBS load is controlled by reserving bandwidth for the collection of TBS sources. New requests for TBS cause the scheduler to search for a time relative position given the source data rate (maximum packet size and repetition rate). If the scheduler cannot find suitable free positions in the relevant periodic superframes, the request is refused.

- If a new TBS reservation request is at an existing repetition rate then the schedule can be examined for free positions in order to decide if this request can be supported. If a new repetition rate is requested, the scheduler must re-calculate the schedule for the common (sub-)multiple interval before a decision can be made. Repetition rates that have no common factors will therefore extend the schedule geometrically.

- Equitable access according to Resource Sharing rules cannot be applied to this scheme as the resource used to calculate TBS schedules would not be constant.

This last observation is an important one. Since the scheduler calculates reservation periods with respect to conflict free channel access, two TDM control points may not share the medium (in the sense defined in [4]). Consequently, TDM control points must be separated by at least twice the interference range (or different channels). (*The physical placement of TDM management devices must be planned.*)

Some additional observations may be made about TDM approaches:

- Unused, but reserved, capacity may be offered only to non-TBS traffic. If, as is the case in [2], TBS traffic is permitted to be variable length, then subsequent sources may acquire channel access rights sooner than their scheduled time position. The overall length of the TBS Period will reduce thus allowing more time in the ADS Period.

- If variable length traffic is supported, then it is implicitly assumed that TBS sources always have data ready to be transmitted, even before their scheduled transmit time. Hence, TBS traffic is delivered with Transit Delay related to the scheduled repetition rate and must be controlled with the bounds of Delay Variance.

- There is no (efficient) support in bandwidth reserved schemes for TBS events i.e. non-periodic time bounded data.

2. DTBS

DTBS is described in terms of priority access (queue ordering). TBS traffic is given a higher priority than ADS traffic such that transmission requests are priority ordered [6]. If a channel access mechanism is used that supports access priority, then DTBS traffic will have such access priority. If no such support is available, each station's traffic must contend for the channel irrespective of traffic type, with priority being applied simply to the internal per-station queues.

In the approach described in section 1, QoS is used to negotiate TBS service. Negotiation precedes initiation of service. In DTBS, the QoS information is transported with the TBS data. This is similar to how a datagram carries its addressing information with its data.

2.1 DTBS Traffic Management

DTBS must control TBS load such that the QoS constraints of the TBS sources can be respected. In order to control the TBS load, each station estimates both its own TBS load and the total channel load. Before adding further TBS load, it must be determined if the QoS of queued TBS requests would be violated by such additional load - and if so the request for additional load will be refused. Other agreed constraints such as TBS / ADS ratio must also be respected.

A TBS request is made in terms of:

- Bandwidth
- QoS
  - Transit Delay
  - Delay Variance

In DTBS, each transmission request carries its QoS parameters. Bandwidth is determined implicitly by the data length of the request.

The following example shows a number of unsynchronized TBS sources within a station:

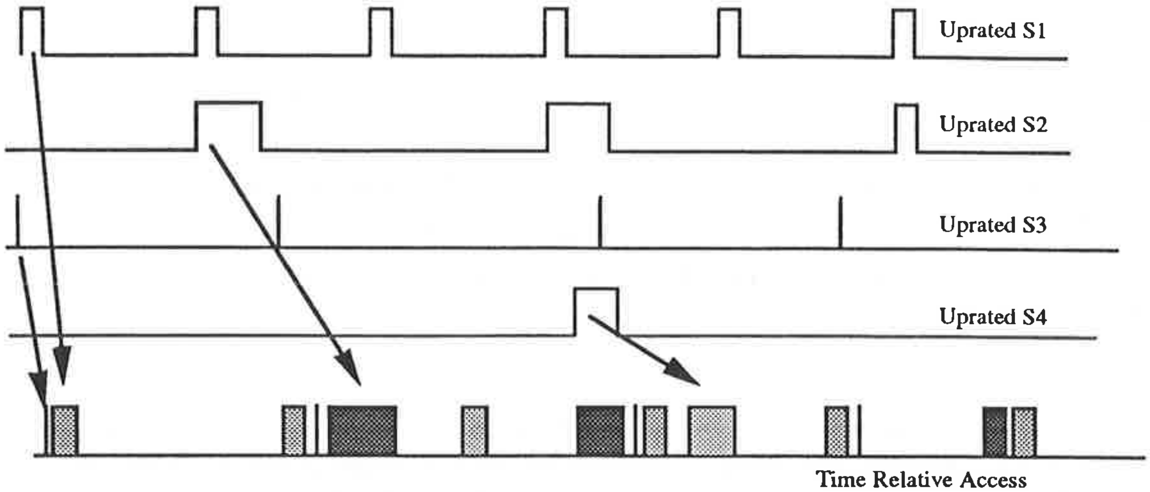

Figure 4. Queued Priority Access

The relative timing of the various TBS sources is not necessarily maintained in their access to the channel. In one instance S1 accesses the channel before S2 (S1 & S2 are coincident), in another S2 accesses the channel before S1. There is no knowledge of any underlying repetition rate nor any requirement for skewing of start time. Data is presented to the queuing mechanism in generated order and contends for the channel with other TBS queued transmitters.

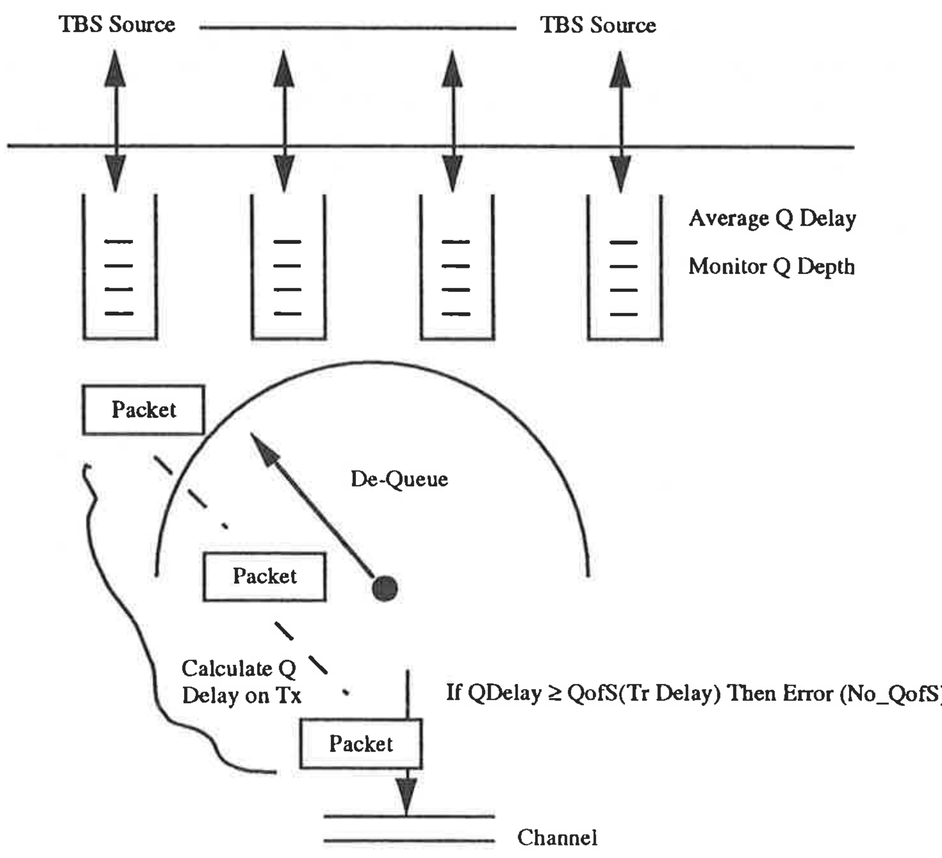

Figure 5. DTBS Queue Delay Measurement

DTBS needs to have a measure of the average delay experienced by a TBS MSDU within the MAC i.e. the portion of the total MSDU delay that occurs before transmission. In DTBS this measure can be made for each request by timestamping the enqueue and dequeue operations. The average queuing delay, and its change as a function of time, yields useful information about the performance of the TBS.

At transmission time, the measured queue delay must be subtracted from the Transit Delay to give the Residual Transit Delay (RTD) i.e. the time left before the frame becomes out of date. RTD may be used in subsequent handling of the frame. If RTD should become less than or equal to zero, the frame should in all cases be discarded.

2.2 TBS Forwarder Behavior

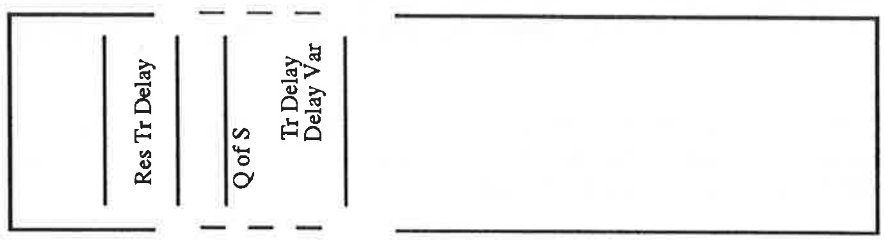

Figure 6. MAC DTBS Control Fields

In the current foundation, the point coordination function also provides a relay (forwarding) service. In fact, this service is separate from TBS and could be provided by any station. As indicated in Figure 6, DTBS frames carry RTD and QoS with the TBS data. The RTD field must be available for manipulation by any intermediate station on the transmission path to the destination. DTBS MPDUs are ordered for transmission by RTD within a forwarder and the RTD field is updated at re-transmission time. (The initial value of the RTD field is the QoS Transit Delay. RTD is reduced for each queuing operation.) RTD therefore acts as a lifetime parameter for the frame.

3. Excess TBS load

When the QoS that can be achieved falls below that which was requested by the TBS user, the DTBS service must indicate this to the TBS user so that (future) adaptive algorithms may react in a suitable manner. This indication is made at the enqueue request time i.e. MSDUs accepted for transmission may or may not be successfully sent. However, via feedback from the DTBS service, a DTBS user will be informed when requesting transmission whether the QoS cannot be achieved.

Observations that can be made on the DTBS approach:

- DTBS complexity is limited to the measurement of the queuing delay and treatment of the RTD parameter at the transmitter. Since there is no concept of repetition rate there is no inter-frame scheduling. In the receiver, a DTBS provider will be required to buffer frames for delivery within the constraints of QoS.

- TBS data may be immediately offered for transmission. There is no requirement for buffering at the transmitter except residence in the transmit queues. Therefore the total number of TBS sources that can be handled is dependent on the measured TBS channel load (with respect to its bounds).

- TBS offered load and the number TBS sources may vary, with peak load exceeding the average TBS for short periods. Events or aperiodic TBS sources are treated with the same efficiency as periodic sources.
- DTBS load is controlled by the Transit Delay parameter of TBS requests. The average queue delay is used as a measure of service quality together with the queue depth. If a request for DTBS transmission (simplified case) is made with a Transit Delay value shorter than the average queue delay it will be refused.
- Equitable access according to Resource Sharing rules can be applied to DTBS. In addition, DTBS frames may pass via forwarders with no additional complexity.
- Support for statistics of the DTBS operation is readily available i.e. average queue delay and TBS queue length. Statistics will be required for the efficient utilization of TBS by non-fixed rate sources.
- DTBS QoS information is embedded in the MAC header added to the PDU and is carried with the frame on its path to its destination.
- Any PDU that has a zero or negative RTD is discarded i.e. RTD is the maximum time the PDU is allowed to continue in the network.

4. Conclusions

A distributed method for Time Bounded Service management has been suggested which controls the TBS load offered by a given 802.11 station based upon that station's measure of the total load offered to the channel. The method will support different bandwidth, transit delay and delay variance values for each TBS request and will permit multiple hop TBS within the bounds of total transit delay of the QoS. Each TBS request carries its QoS parameters such that the residual transit delay can be calculated at each hop and used to control the maximum lifetime of the frame. Aperiodic TBS traffic, or TBS Events, may also be supported with no additional complexity. Since resources are not centrally administered, there are no limitations on overlapping DTBS implementations. Therefore spacial or channel planning is not required in the deployment of DTBS devices.

Bandwidth efficiency, simplicity of implementation and ease of deployment are gained at the cost of guaranteed access. However, this cost is measured with respect to the nature of any radio communications environment susceptible to both interference and fading.

References

[1] Priority in Hiperlan, INRIA, RES10SAG93/28, May '93
[2] DFWMAC, Diepstraten et al., IEEE 802.11-93/190, Nov '93
[3] Systems Architecture Document V0.8, PT41, RES10SAG93/61, September '93
[4] Hiperlan Services and Facilities, IEEE P802.11-93/43, February 1993
[5] Time Bounded Data Delay, PT41, RES10SAG93/38, July '93
[6] A Survey about Time Bounded Services, INRIA, RES10SAG93/25, May '93
[7] Prioritization in Hiperlan : Solutions at MAC sub-layer, INRIA, RES10SAG93/28, May '93
[8] Resource Sharing, Apple, RES10SAG93/48, October '93