IEEE 802.11
Wireless Access Method and Physical Layer Specifications

Title:

The Wireless Hybrid Asynchronous Time-bounded MAC Protocol

Authors:

Phil Belanger
Ken Biba

Xircom, Inc.
2105 Landings Drive
Mountain View, CA
94043

(415) 691-2500 Voice
(415) 691-1064 FAX

pbelange@xircom.com Email
kbiba@xircom.com Email

Abstract:

This submission is an extension of earlier MAC protocol contributions by Ken Biba. [1][2][3] A new Time bounded service is defined, the Asynchronous Service is described in more detail, and the terminology is updated to match current IEEE 802.11 definitions.

Introduction

The Wireless Hybrid Asynchronous and Time Bounded (WHAT) protocol is a Media Access Control (MAC) protocol with enhancements for improved reliability, support of hidden stations, and collocated wireless LANs. The WHAT MAC protocol provides two types of service to the higher layer protocols — an Asynchronous Service and a Time Bounded Service.

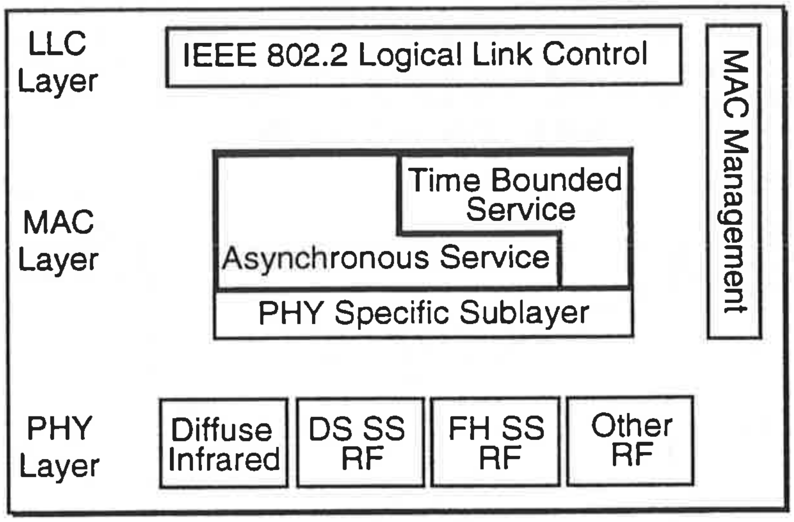


Figure 1 - WHAT Protocol Architecture

The Asynchronous Service is a Listen Before Talk (LBT) protocol that provides low-delay, asynchronous delivery of MAC Protocol Data Units (MPDUs) between any stations (STAs) within direct range of each other. It employs a Carrier Sense Multiple Access (CSMA) Distributed Coordination Function (CF). The Asynchronous Service supports ad hoc networks as well as those utilizing infrastructure

Time Bounded Service is used to transfer MPDUs between a wireless STA and an Access Point (AP) with a minimum delay variance. [4] The Time Bounded Service uses a Distributed CF *and* a Point CF. The Point CF is used to allocate bandwidth to time bounded connections. The Distributed CF is used to reserve the medium for transmission of specific MPDUs. A node that does not support Time Bounded traffic need only implement the Distributed CF for the Time Bounded Service. Time Bounded traffic and Asynchronous traffic can share the same medium.

Infrastructure configuration and Distribution System support is provided by the MAC Management module. It implements a protocol for Registration, Sign-On, Authentication, Privacy, and Power Management. MPDU relaying is built into the Asynchronous Service and the Time Bounded Service.

The WHAT protocol is PHY independent.

This submission focuses on the Asynchronous Service and Time Bounded Service of the WHAT MAC protocol.

Asynchronous MPDU Formats

The Asynchronous Service of the WHAT protocol has two MPDU formats: one for MPDUs addressed to a specific STA, directed MPDUs; and a second format for MPDUs addressed to more than one STA, multicast MPDUs. Each MPDU contains multiple frames.

Directed MPDUs

Directed MPDUs are sent to a specific STA. They consist of four frames: RTS, CTS, DATA, and ACK. RTS, Request-To-Send, is used to reserve the medium for transmission of the data frame. CTS is a Clear-To-Send frame that confirms that the medium is reserved. The DATA frame contains the payload of the MPDU, and the ACK frame is an immediate acknowledgment indicating that the DATA frame was received without error.

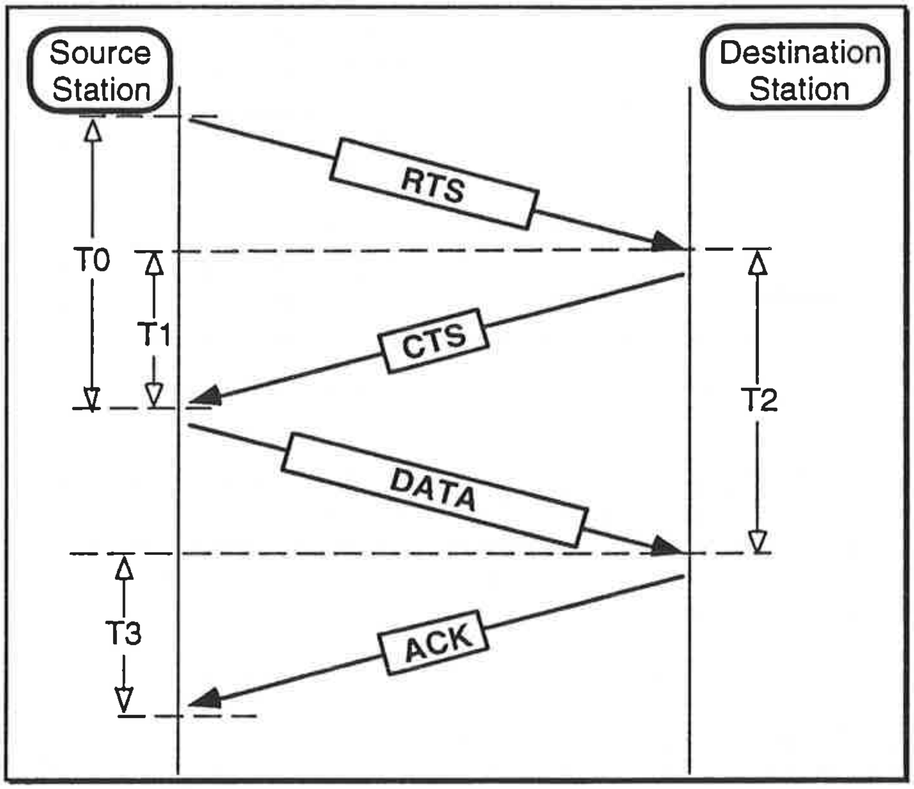


Figure 2 - Directed MPDU.

The minimum time between successive frames, the interframe gap (IFG), is defined by the PHY. It is desirable to keep the IFG as small as possible. Figure 2 illustrates the timing sequence for a directed MPDU. T0 is The slot time for the WHAT protocol. Slot time is the time it takes a STA to acquire the media. It is described further in the Asynchronous Protocol Operation section.

T1 is the CTS timeout, the maximum time that a sending STA will wait for a CTS frame after sending an RTS frame. T3 is the ACK timeout; the maximum time that a sending STA will wait for an ACK frame after transmitting the DATA frame. The maximum value of timeouts T1, T2, and T3 must be as small as possible.

T2 is The DATA timeout from the perspective of receiving STAs. It is described in the Enhanced Carrier Sense section of this document.

Multicast MPDUs

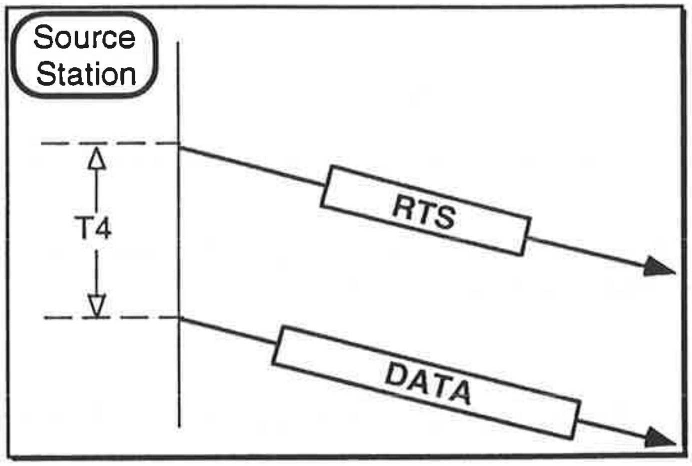


Figure 3 - Multicast MPDU

Multicast and broadcast MPDUs are addressed to a group of STAs from a specific STA. They consist of two frames: RTS and DATA.

Multicast and broadcast MPDUs do not have a specific destination STA to provide CTS or ACK frames. The removal of these frames decreases the delivery reliability of multicast and broadcast MPDUs. However, this is consistent with the level of service provided for multicast frames by other MAC protocols.

T4 should be the IFG time. The frame formats for multicast RTS and DATA frames are the same as the RTS and DATA frame formats for directed MPDUs.

When Access Points are present, the sending STA can control the scope of Multicast MPDUs. Multicasts are not forwarded by APs unless the HIERARCHICAL is set. (See CONTROL field definition below.) If the HIERARCHICAL bit is set, the Access Point will forward the packet to both the wired and wireless links. Multicasts stay within the BSS when HIERARCHICAL is clear and propagate through the ESS when it is set. Multicasts are not duplicated on the wireless side since STAs ignore any frames with the HIERARCHICAL bit set.

MPDU Header Fields

MPDUs are always preceded by a preamble and a Start Frame Delimiter (SFD). The length and format of the preamble and SFD are defined by the PHY layer.

A Cyclic Redundancy Check (CRC) field is appended to each frame. Control frames use an eight bit CRC-8 and the DATA frame has an IEEE 802.3 standard CRC-32.

Multiple byte fields are transmitted with the most significant byte first, and bit order is most significant bit first.

There are several MPDU header fields that are required to implement the WHAT protocol. A subset of the fields is transmitted with each frame.

TYPE

The TYPE field has two subfields; MPDU Type and Frame Type. All frames have a TYPE field.

Frame Types are:

RTS	= 0;	Request To Send
CTS	= 2;	Clear to Send
ACK	= 4;	Acknowledge
DATA	= 8;	Data
RTSI	= 1;	Request To Send Inbound, Time Bounded Service
CTSI	= 3;	Clear to Send Inbound, Time Bounded Service
RTSO	= 5;	Request To Send Outbound, Time Bounded Service
CTSO	= 7;	Clear to Send Outbound, Time Bounded Service
ANNOUNCE	= 6;	Announce MPDUs are one frame that says: "I'm here."

MPDU Type is actually a set of Boolean flags {SYSPKT, ENCRYPT, COMPRESS}

A SYSPKT of 1 means that this MPDU contains a WHAT Management Protocol message and should be processed by the MAC Management layer. When SYSPKT is 0 the MPDU data is passed to the higher layer protocols.

> ENCRYPT and COMPRESS are examples of other uses for this field.
>
> When ENCRYPT is 1, the data portion of the MPDU is encrypted. If ENCRYPT is 0, the data is not encrypted.
>
> When COMPRESS is 1, the data portion of the MPDU is compressed.

CONTROL

The CONTROL field is a set of Boolean flags that determine the processing of the frame by the receiver. CONTROL = {TB, HIERARCHICAL, AP, RETRY}. All frames have a CONTROL field.

A TB bit of 1 indicates that this MPDU is for Time Bounded service. It defines the format of the rest of the frame. If TB is zero, this is a frame for the Asynchronous Service.

If the HIERARCHICAL bit is 1, the MPDU will only be received by APs. Other STAs do not process MPDUs with the HIERARCHICAL bit set. APs always clear this bit before they relay an MPDU. The HIERARCHICAL bit is used to force Asynchronous traffic through an AP.

The AP bit is 1 when this frame was sent by an Access Point. This allows STAs to distinguish between an MPDU sent directly from another STA and one that has been forwarded by an Access Point.

The RETRY bit is 1 when the MPDU is a retransmission of an earlier MPDU.

DATA LENGTH

DATA LENGTH is the number of bytes in the payload of the DATA frame of the MPDU. The DESTINATION ADDRESS and SOURCE ADDRESS are considered separate fields, so DATA LENGTH describes the total length of the LLC MSDU minus the length of the SOURCE ADDRESS and DESTINATION ADDRESS fields. The maximum MAC MPDU length is TBS.

PHY SPECIFIC

The PHY SPECIFIC field contains information that is used to help control the PHY layer. The format and meaning of this field is defined by the PHY. This field is processed by the PHY specific sublayer in the MAC. For example, PHY SPECIFIC could contain the Transmitted Power Level or Frequency Hop Timing information. [5]

NETID

NETID is a 16 bit number that identifies the BSS of the sending STA. Stations will accept MPDUs with NETIDs matching their own. This field is used to differentiate between MPDUs from overlapping service areas.

MPDUID

MPDUID is a 16 bit number that is used to identify frames of a particular MPDU. MPDUID is transmitted in every Asynchronous frame and is used by receiving STAs to verify that each frame belongs to a particular MPDU. [6] [7] It is a simple sequence number that is incremented each time a new MPDU is sent. The MPDUID may also be used to help duplicate detection.

DESTINATION ADDRESS

DESTINATION ADDRESS is the 48 bit unique ID of the destination STA or a 48 bit multicast address. All addresses are standard IEEE 802.3 48 bit addresses.

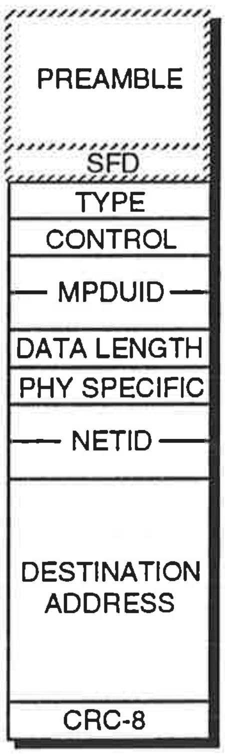
SOURCE ADDRESS

SOURCE ADDRESS is the unique 48 bit ID of source STA.

Asynchronous Frame Formats

RTS Frame

RTS, or Request-To-Send, is a transmission from source to destination initiating the MPDU. It contains general MAC header information plus the 48 bit ID of the destination STA. The RTS frame reserves the wireless medium in the vicinity of the source STA. RTS is a fixed length frame.



TYPE = 0x, where x is the MPDU Type.

CONTROL ={TB=0, HIERARCHICAL = x, AP = x, RETRY = x}

The DATA LENGTH field is not the length of this frame. It is the length of the payload of the DATA frame that will be sent if the source STA receives a CTS frame matching this RTS.

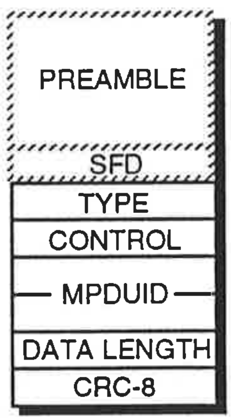
DESTINATION ADDRESS is the 48 bit unique ID of the destination STA or a 48 bit multicast ID.

CRC-8 is an 8 bit CRC computed on the entire frame.

CTS Frame

CTS, or Clear-To-Send, is a transmission from destination to source granting permission to transmit the DATA frame. CTS is not used for flow control in this context. CTS means that the destination STA can hear the source STA and is ready to receive the DATA frame. There is no implication about the availability of buffers in the receiver. The destination STA will not send the CTS if it perceives that the network will be busy during the transmission of rest of the MPDU. This will cause the source STA to retransmit the MPDU at a later time. [7]

CTS is an important part of the WHAT carrier sense mechanism since it conveys the length of the data frame to those wireless STAs that can hear the destination but not the source of the MPDU. The CTS frame reserves the wireless medium in the vicinity of the destination STA. Any STA that hears the CTS will defer its own transmission attempts for a period of time equal to the DATA and ACK frame duration. CTS is a fixed length frame.



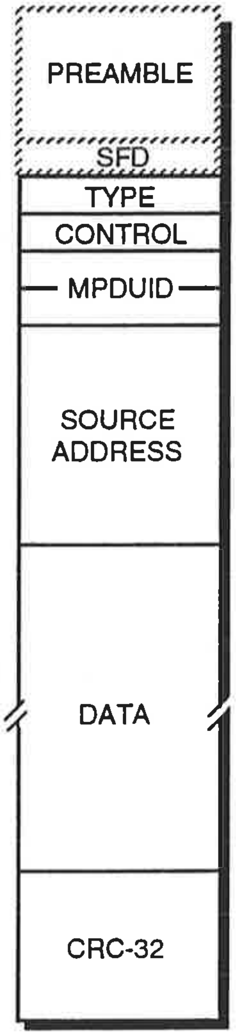
TYPE = 2x, where x is the MPDU Type.

CONTROL ={TB=0, HIERARCHICAL = x, AP = x, RETRY = x}

The DATA LENGTH field is not the length of this frame. It is the length of the payload of the DATA frame that will be sent if the source STA receives this CTS frame.

DATA Frame

DATA is a transmission from source to destination that contains the data payload of the MPDU. The payload (the LLC data) is 32 to 586 bytes in length as specified by the DATA LENGTH field.



TYPE = 8x, where x is the MPDU Type.

CONTROL ={TB=0, HIERARCHICAL = x, AP = x, RETRY = x}

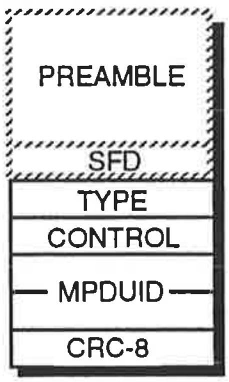
SOURCE ADDRESS = Unique 48 bit ID of the sending STA.

DATA is a variable length field specified in the DATA LENGTH field of the RTS and CTS frames. Minimum DATA LENGTH is 32 bytes; maximum is 586 bytes.

CRC 32 is a thirty-two bit IEEE 802.3 standard Cyclic Redundancy Check sequence on the entire DATA frame.

ACK Frame

ACK is a transmission from destination to source that acknowledges receipt of the MPDU's DATA frame with a correct MPDU check sequence. Negative acknowledgments are not used in this protocol. An ACK frame always indicates successful receipt of the DATA frame. The ACK is a fixed length frame.



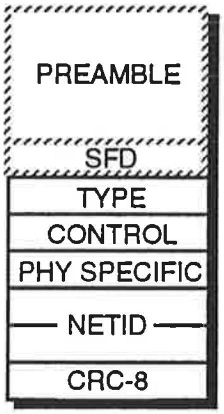
TYPE = 4x, where x is the MPDU Type.

CONTROL ={TB=0, HIERARCHICAL = x, AP = x, RETRY = x}

MPDUID = x. Same as MPDUID received in DATA frame.

Announce Frame

Announce is a single frame MPDU. It is used to indicate that there is someone using a particular NETID. It also conveys the PHY SPECIFIC information. A STA joining the network for the first time listens to determine what other STAs and APs belong to the same BSS. Announce frames are sent periodically when the network is idle, and this ensures that new stations can quickly find the other members of the BSS. ANNOUNCE is a fixed length frame.



TYPE = 61h, which is ANNOUNCE frame and SYSPKT.

CONTROL ={TB=0, HIERARCHICAL = 0, AP = x, RETRY = 0}

NETID = x, the NETID that identifies the BSS of the sender.

CRC- 8 Is an 8 bit CRC computed on the entire frame.

Asynchronous Protocol Operation

Enhanced Carrier Sense

Carrier sense in the WHAT MAC Protocol includes two factors: 1) the state of the channel as observed by the PHY layer; and 2) a Net Allocation Vector maintained by the MAC layer that indicates when the network is busy. The Net Allocation Vector describes when the network is idle or busy from the point of view of a particular STA. STAs maintain the Net Allocation Vector by listening to RTS and CTS frames.

Whenever a valid RTS or CTS frame is received, The Net Allocation Vector is updated to indicate that the network will be busy for the duration of the transmission of the DATA and ACK frames of the MPDU. Receivers can compute the total time the network will be busy because the DATA LENGTH field in the RTS or CTS specifies the length of the DATA frame and the other frames of the MPDU are fixed length.

Every receiving STA has a time out for receiving a DATA frame. This is described in figure 2 as T2. T2 is the DATA timeout. A STA that receives an RTS will mark the network as busy in the Net Allocation Vector. If a corresponding DATA frame is not received within the T2 time interval, the Net Allocation Vector is cleared. This limits the reservation time of an RTS that is never followed by the rest of the MPDU. This will occur when a STA attempts to transmit to another STA that is not on the network (or not within PHY range, or Sleeping, or in a null...).

Stations will attempt to transmit only when the Net Allocation Vector indicates that the network is idle.

> The Net Allocation Vector (NAV) can be implemented as an array of Boolean flags. Each flag (bit) represents a slot time (T0 in figure 2). If the network is idle the bit is True. If the network is busy the bit is False. When a STA hears an RTS or CTS it can mark the network as busy by setting the appropriate number of consecutive bits. This could be implemented as a ring buffer with a moving index that represents "now".
>
> It is also possible to implement the NAV as an ordered linked list. Each record in the list can represent the duration of a busy period in absolute time. Whenever an RTS or CTS is heard a new record is created. If the current time is later than the end time of a record or if a reservation is canceled, the record is removed from the list.

Channel Contention and Collisions

WHAT is a Carrier Sense Multiple Access (CSMA) protocol. Before sending, STAs check to see if the network is busy, and defer to any transmissions in progress. (This is classic Listen Before Talk, LBT.) It is possible that more that one STA can sense the channel, conclude that it is not busy, and begin transmitting at the same time. When this happens, a collision occurs.

CSMA/CD protocols (e.g., Ethernet) detect the collisions early in the frame using Listen While Talk (LWT) and reduce wasted bandwidth by aborting transmission as soon as a collision is detected. The WHAT protocol does not rely on a LWT PHY layer.

The WHAT protocol uses the RTS/CTS exchange to sort out channel contention. The RTS frame indicates a STA's intent to send a DATA frame, the destination address, and the length of the data transmission. The CTS frame is sent from the destination STA to the source STA, and it also includes the length of the DATA frame. After an RTS/CTS exchange, any STA that "heard" either frame will not attempt a transmission until the rest of the MPDU is finished. The channel is effectively reserved by a successful RTS/CTS frame exchange. This virtually eliminates collisions during the DATA frame transmission.

If collisions do occur, they will most likely happen during transmission of the RTS or CTS frames. If an RTS is damaged, no CTS will be sent; and the MPDU will not be completed. If the CTS is damaged, the DATA frame will not be sent. In either case, the source STA will know that it must attempt transmission of the entire MPDU at a later time. The RTS and CTS frames provide a virtual collision detect mechanism for the WHAT protocol. The effect of collisions is minimized because the MPDU transmission will be stopped after the RTS frame when a collision occurs (RTS is a short control frame); and other STAs that marked the network as busy will clear their Net Allocation Vectors after the DATA timeout.

Positive ACK Protocol

WHAT uses a positive ACK protocol. Low level acknowledgments are used to improve the reliability of the wireless medium. When a directed MPDU is sent, the source STA expects an ACK frame after each DATA frame. If an ACK is not received within a short period of time (t3 in figure 2), the ACK is assumed lost and the MPDU is scheduled for retransmission. This process is repeated N times. If no ACK is received after the Nth attempt (N is TBS), send failure is reported. Broadcast and multicast MPDUs do not expect ACKs and are never retransmitted.

Duplicate Detection

Whenever a protocol retransmits frames there is a possibility of duplicates. The WHAT protocol is designed to filter out duplicate MPDUs and conserve the channel bandwidth. In particular, the protocol guarantees that duplicate MPDUs will not be delivered to the client protocol layers.

Transmitting STAs never send a new MPDU to the same destination during retransmission attempts of a previous MPDU. In other words, there is only one MPDU outstanding per source STA/destination STA pair. The receiving STA may receive MPDUs from other STAs interleaved with retransmissions from one STA.

Each STA maintains a list of potential duplicate MPDUs. MPDUs are identified by their MPDUID. Sending STAs set the RETRY bit (in the CONTROL field of the frame header) whenever they transmit an MPDU more than once. If an RTS for an MPDU is received with the RETRY bit set, the destination STA will send the CTS and prepare to receive the DATA frame. The MPDUID list is scanned to see if there are any matching entries. If no match is found, the MPDU is not a duplicate, and reception proceeds as usual. If a match is found in the MPDUID list, the MPDU is a duplicate. In that case, the destination will send an ACK frame, but will discard the MPDU since it was already delivered to the client protocol software. Once the ACK frame is received, the source STA will stop retransmitting the MPDU.

Duplicate detection rules:

- MPDUIDs are stored in the duplicate list each time a directed MPDU is received with matching NETID and DESTINATION ADDRESS.
- The MPDUID list is scanned for potential duplicates only when an MPDU is received with matching NETID and DESTINATION ADDRESS, and the RETRY bit is set.

> It is possible to implement duplicate detection by a linked list of records containing the SOURCE ADDRESS and MPDUID. The list is ordered by SOURCE ADDRESS. Each time an MPDU is received, a new record is saved. If a record with a matching SOURCE ADDRESS exists, it is replaced with the new record.

Retransmission Algorithm

When a CTS frame is missed, the source STA assumes that a collision occurred and schedules a retransmission of the MPDU. When an ACK frame is not received, it indicates that the DATA frame was damaged or the ACK frame itself was damaged. In either case, the MPDU must be retransmitted. The same back off algorithm is used for ACK timeout retries and CTS timeout retries.

Each STA delays their retransmission for a random amount of time to reduce the probability of collisions. On the first transmission attempt for an MPDU, there is no back off time. The source STA defers to other traffic and transmits as soon as the channel is free. For subsequent transmission attempts of the same MPDU, the source STA waits a random back off time before sending the RTS frame.

The back off procedure uses a special timer that only counts down when the network is idle. The sending STA computes a random number, and uses it to set the initial value of the back off timer. Each "tick" of the timer represents a slot time (T0 is figure 2). The slot size is the time it takes a STA to reserve the channel for the transmission of a DATA frame. In other words, the length of an RTS frame, plus the length of a CTS frame, plus some overhead for IFG.

After the back off timer elapses, the sending STA senses the channel again and transmits if it is available. If the network is not idle the sending STA will repeat the back off procedure.

There are no retransmissions for broadcast or multicast packets.

The back off algorithm has three desirable properties.

1) Senders delay their transmissions for a longer time when the network is busy. This helps the network remain stable under heavy loads and happens automatically, because the back off timer will not elapse unless the network is idle.

2) The algorithm eliminates the "clustering" problem. Clustering occurs when multiple STAs are waiting to transmit and their back off timers elapse during the transmission of a long frame. Even though the timers may elapse at different times, the transmission attempts will be clustered at the same time, just after the end of the long frame. Since the WHAT back off timer does not count down during network busy periods, clustering will not happen

3) It works well with a small number of long ticks. The WHAT slot time is rather long, so it is important to limit the random back off to a small number of ticks — 8 or 16. When the network is busy a small backoff number will generate a long back off period.

Time Bounded Protocol Overview

The Time Bounded Service was designed primarily to support real time voice traffic. It transfers data over a point to point connection between an Access Point and a wireless station that are within range (defined by PHY) of each other. The protocol was designed to minimize the delay variance between successive frames sent over the connection.

APs control the Time Bounded traffic. STAs wishing to send Time Bounded messages negotiate with their AP to set up a connection using the Asynchronous Service. The exact format of the Call Setup messages is TBS. APs grant the connection request only if there is enough available bandwidth to support a new call. If the AP accepts the connection, it assigns a local identifier that is used in subsequent MPDU headers for that connection.

One of the parameters of the Call Setup is quality of service, which dictates the timing between Time Bounded MPDUs for that connection. Time Bounded data is transferred in fixed length MPDUs. After a call is established, the AP will send Time Bounded MPDUs to the STA at fixed intervals. The AP and STA exchange a series of modified RTS and CTS frames with data attached to transfer the Time Bounded data.

Underlying assumptions about Time Bounded Service:

- Time Bounded service will be used primarily for voice traffic
- voice traffic does not have to be as reliable as data traffic
- it is better to drop a voice MPDU than to delay it for a long time
- it is better to deliver a voice MPDU in error than to delay it for a long time
- voice traffic will always flow through the AP

Time Bounded MPDU Formats

There are two types of MPDUs for Time Bounded services: inbound MPDUs and outbound MPDUs. [4] All Time Bounded MPDUs are directed. There is no Time Bounded multicast. Each Time Bounded MPDU is a two frame exchange.

Inbound MPDUs

Inbound MPDUs cause data to flow from a STA to an AP - into the wired infrastructure. The AP sends an RTSI frame and the wireless STA responds with a CTSI frame that includes the Time Bounded data.

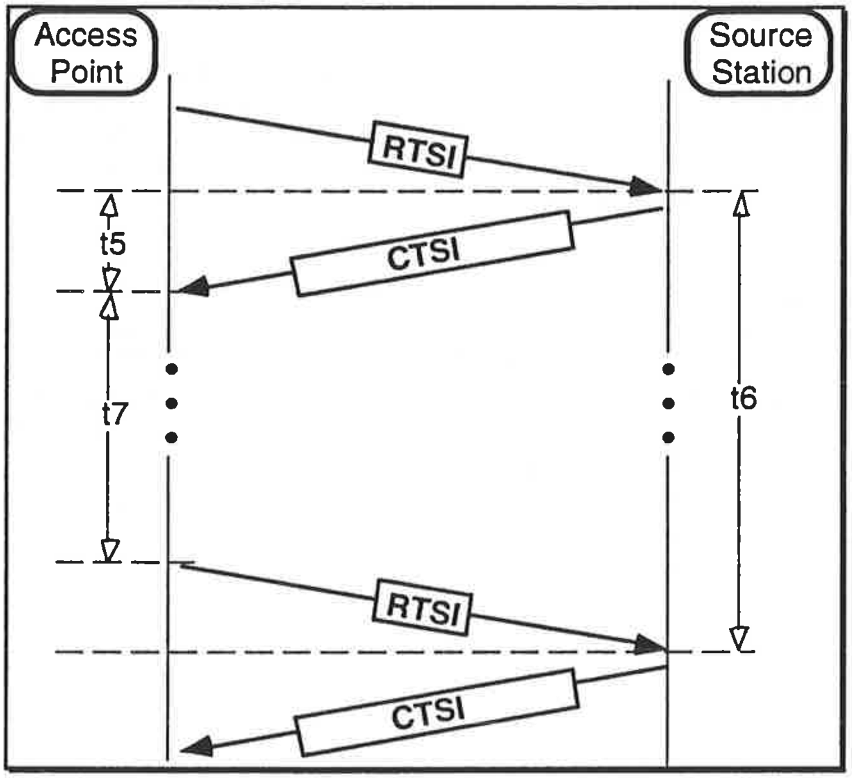


Figure 4 — Inbound MPDU exchange

The AP sends RTSI frames at fixed intervals, t6, for the duration of the connection. Each time the source STA receives an RTSI, it responds with an immediate CTSI frame. There are no ACKs and no retransmissions. If one of the frames is lost, it is ignored. The space between the Inbound MPDUs, t7, is available for Asynchronous traffic or other Time Bounded MPDUs.

Outbound MPDUs

Outbound MPDUs cause data to flow from an AP to a STA - out from the wired infrastructure. The AP sends an RTSO frame which includes the data and the wireless STA responds with a CTSO frame. RTSO frames are sent at a constant interval, t6, in the figure below.

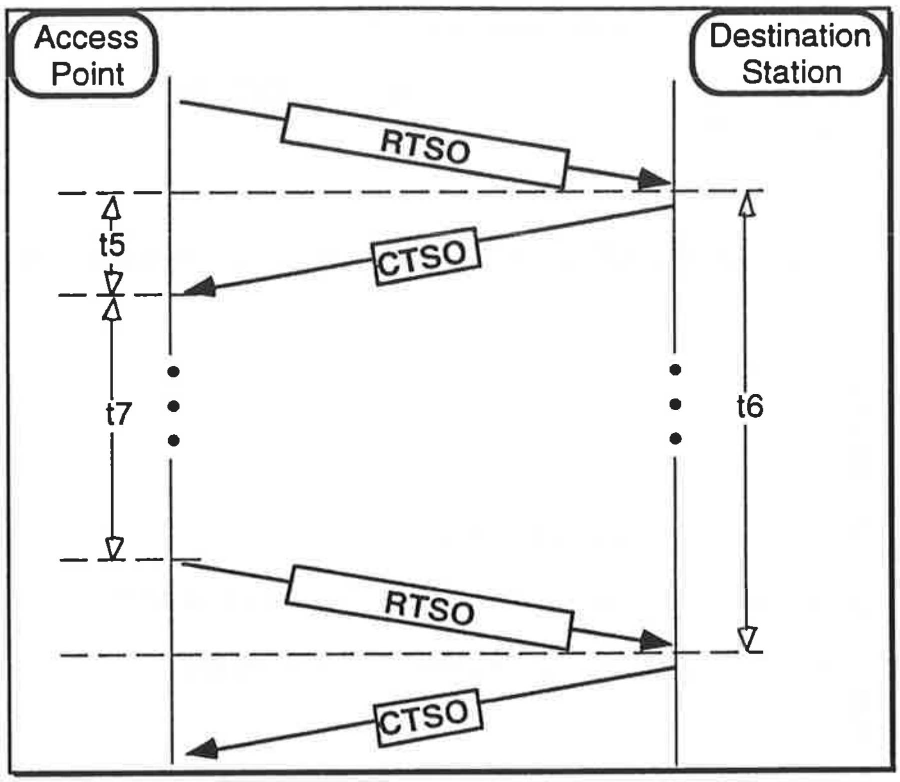


Figure 5 — Outbound MPDU exchange

Time Bounded Frame Formats

Time Bounded Header Fields

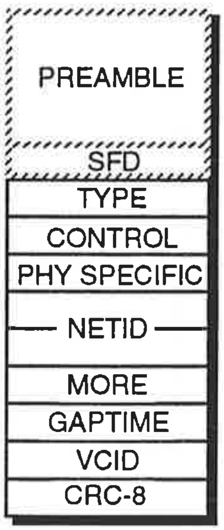
VCID — A Virtual Circuit IDentifier is assigned by the AP each time a Time Bounded connection is established. VCID is unique within a BSS, and the NETID, VCID combination is unique within a domain. VCID is an eight bit number.

MORE — The number of times in the future to reserve the medium for MPDUs from this connection. Four is a reasonable number for this.

GAPTIME — Indicates the time between successive MPDUs from this connection. This is measured in slot times.

RTSI Frame

RTSI frames are sent from the AP to the STA, and are used to maintain the timing between MPDUs of the connection.



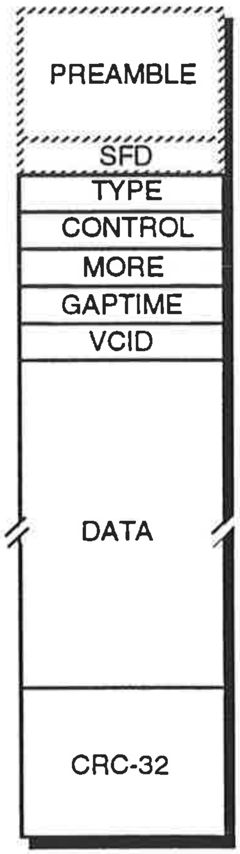
TYPE = 1x where x is the MPDU Type.

CONTROL ={TB=1, HIERARCHICAL = 0, AP = 1, RETRY = 0}

CRC 8 Is an 8 bit CRC computed on the entire frame.

CTSI Frame

CTSI frames are sent from the STA to the AP in response to an RTSI frame. They contain the data payload for the MPDU.



TYPE = 3x where x is the MPDU Type.

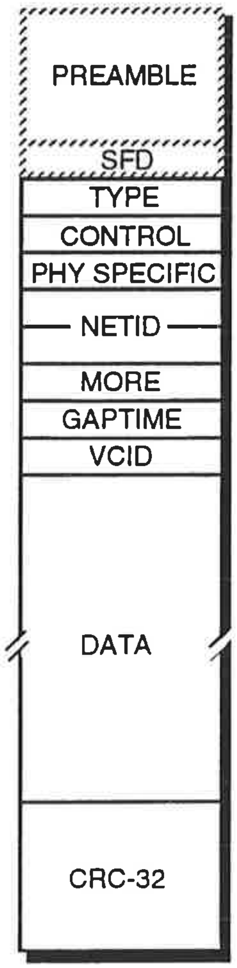
CONTROL = {TB=1, HIERARCHICAL = 1, AP = 0, RETRY = 0}

DATA — The payload of the inbound MPDU.

CRC 32 — Is an 32 bit CRC computed on the entire frame.

RTSO Frame

RTSO frames carry the data for outbound connections. RTSO frames are sent from AP to STA.



TYPE = 5x where x is the MPDU Type.

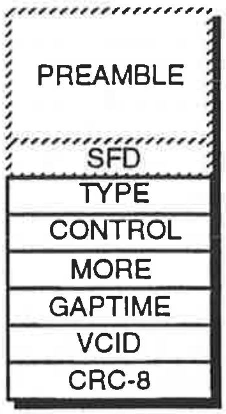
CONTROL ={TB=1, HIERARCHICAL = 0, AP = 1, RETRY = 0}

DATA — The payload of the outbound MPDU.

CRC 32 — Is an 32 bit CRC computed on the entire frame.

CTSO Frame

CTSO frames are sent from the STA to the AP, and are used to propagate the timing information for an outbound connection.



TYPE = 7x where x is the MPDU Type.

CONTROL ={TB=1, HIERARCHICAL = 1, AP = 0, RETRY = 0}.

CRC 8 Is an 8 bit CRC computed on the entire frame.

> The payload length for Time Bounded MPDUs is TBS. It is a fixed length.
>
> A likely digital Distribution System for Time Bounded connections will be based on ATM cell switching. One option for 802.11 is to specify the payload of Time Bounded MPDUs as an ATM cell or multiple ATM cells. If we can make this assumption about the payload we could also simplify the frame check sequence for Time Bounded MPDUs.

Time Bounded Protocol Mechanism

Point or Distributed Coordination Function?

The Time Bounded service should minimize the MPDU delay variance for it's clients. The simplest way to achieve minimum delay variance is to provide guaranteed bandwidth for Time Bounded connections through a central Point Coordination Function. The APs are a natural place to put this centralized CF.

The Time Bounded service operates in conjunction with the Asynchronous service which uses a distributed, CSMA protocol. STAs operating in Asynchronous mode do not have to check with a Point Coordination Function before transmitting. If a STA perceives that the network is idle, it may initiate an Asynchronous transmission. This is a problem for the Time Bounded service. Once a call is established, Time Bounded connections should not contend for the media for each MPDU during a connection. Because of the minimum delay variance requirement, Time Bounded MPDUs can not defer to other traffic once the connection has been started.

Bandwidth Allocation

The solution to this dilemma is a two step process that is a compromise between distributed and centralized control. Call set up and bandwidth allocation are managed by a Point Coordination Function in the WHAT protocol. All Time Bounded connections are set up through the AP. APs are aware of traffic level for their service area. They know how many calls are in progress and the number of Asynchronous nodes in their area; and they manage the bandwidth accordingly.

The APs also control the timing of the connection. They initiate all Time Bounded MPDUs. Once a connection is granted, the AP sends the first MPDU at a time when the media is idle and coordinated with other connections that are ongoing. After the first MPDU, subsequent MPDUs for that connection are sent at fixed intervals.

The AP will manage the bandwidth according to a policy that is TBS. Some examples are shown in figure 6 below. The AP could group all of the Time Bounded connections together in order to keep maximum size gaps available for asynchronous traffic or it could space the connections evenly in an attempt to minimize latency for the Asynchronous Service. The operation of the MAC is the same regardless of the policy the Access Point chooses.

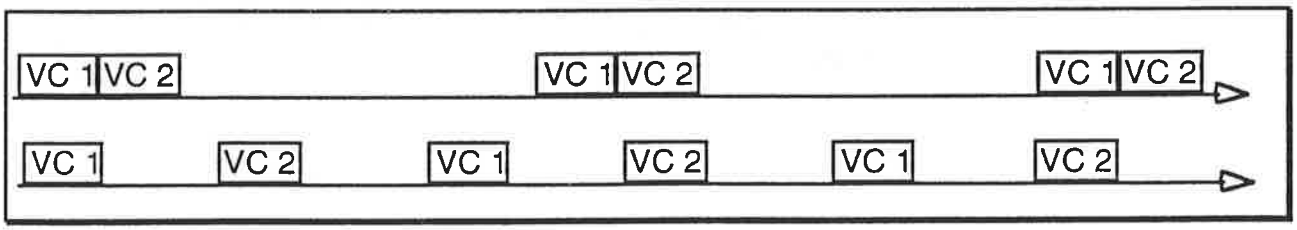


Figure 6— Bandwidth allocation scenarios

Bandwidth Reservation

The second part of the Time Bounded protocol mechanism is a Distributed Coordination Function that is used to reserve the bandwidth for Time Bounded MPDUs and inform all of the STAs in the BSS of the reservation to prevent contention with Time Bounded connections. This is done by an extension to the carrier sense method that is already used by the Asynchronous service.

Each Time Bounded MPDU is a fixed length. Any node that receives a Time Bounded frame can determine the length of time the network will be busy for that MPDU exchange. There is a GAPTIME field in all Time Bounded frames that specifies when the next MPDU for this connection will be sent. It is used to reserve bandwidth for subsequent frames of the same connection. Any node that "hears" a Time Bounded frame will assume that the network will be busy for the length of a timed based MPDU after GAPTIME time has passed. This "reserve ahead" mechanism prevents other nodes within range of either end of the connection from contending for the network during the transmission of Time Bounded frames for this connection.

There is an additional field in the header called MORE that indicates the number of intervals the reservation covers. With MORE and GAPTIME, it is possible to "reserve ahead" up to four transmission opportunities. This makes the mechanism more reliable since each connection will actually reserve a specific MPDU transmit opportunity up to four times.

Once an AP and STA have successfully exchanged the first MPDU of a Time-bounded connection, the medium will be reserved at both ends of the connection. However, the APs initiate the timing of a connection without knowing the state of the Net Allocation Vector of the STA. To resolve this potential problem, the response to the Call Setup message and the first MPDU of the connection are timed as if they were successive MPDUs from an established connection. The STA will respond with a confirmation message if its Network Allocation Vector indicates the network will be available for the first MPDU transfer. Otherwise, the STA will not respond and the AP will attempt to initiate the connection at a different time.

Impact on Asynchronous Service

Most of the complexity of the Time Bounded Service is in the Point CF that is implemented in the AP. However, there are enhancements to the Distributed CF in every STA that are required to support Time Bounded Service. In particular, the carrier sense mechanism has to be expanded. A STA that only implements the Asynchronous Service needs to know about the current transmission in order to implement carrier sense. When the Time Bounded Service is added, each STA must keep a map of what will happen in the future. The Net Allocation Vector must be able to record network availability for some period of time into the future.

The transmit procedure for Asynchronous MPDUs is slightly more complex. A sending STA must find an idle period that is long enough for transmission of the entire MPDU (rather that the first idle period...) before contending for the media.

These changes are not an unreasonable burden on the MAC implementation. The WHAT protocol requires that every STA implement the enhanced Distributed Coordination Function whether or not it provides Time Bounded Service.

Behavior with Overlapping Service Areas

The Time Bounded service works well with collocated or overlapping service areas. All nodes maintain their own Net Allocation Vector based on the frames they "hear" on the network. This includes frames with matching NETID and those frames from different service areas with different NETIDs. Network reservations for Time Bounded connections are automatically propagated only to those STA that are within range of one of the ends of a connection.

In figure 7 below, node c can hear Access Point 1 (AP1) and AP2. Node c's net allocation vector will include reservations for all of the Time Bounded traffic for AP1 and AP2. Node d can only hear AP1 and node c. If node c is registered with AP1 and has established a Time Bounded connection, node d's net allocation vector simply includes any asynchronous traffic and all of the Time Bounded traffic from AP1. Traffic from AP2 and nodes a and b will not affect node d.

If node c is registered with AP2 instead, node d's net allocation will include everything described above plus the reservations for any Time Bounded traffic between AP2 and node c. As soon as the first Time Bounded MPDU is sent from AP2 to c, d will learn about the relevant Time Bounded traffic from the other service area.

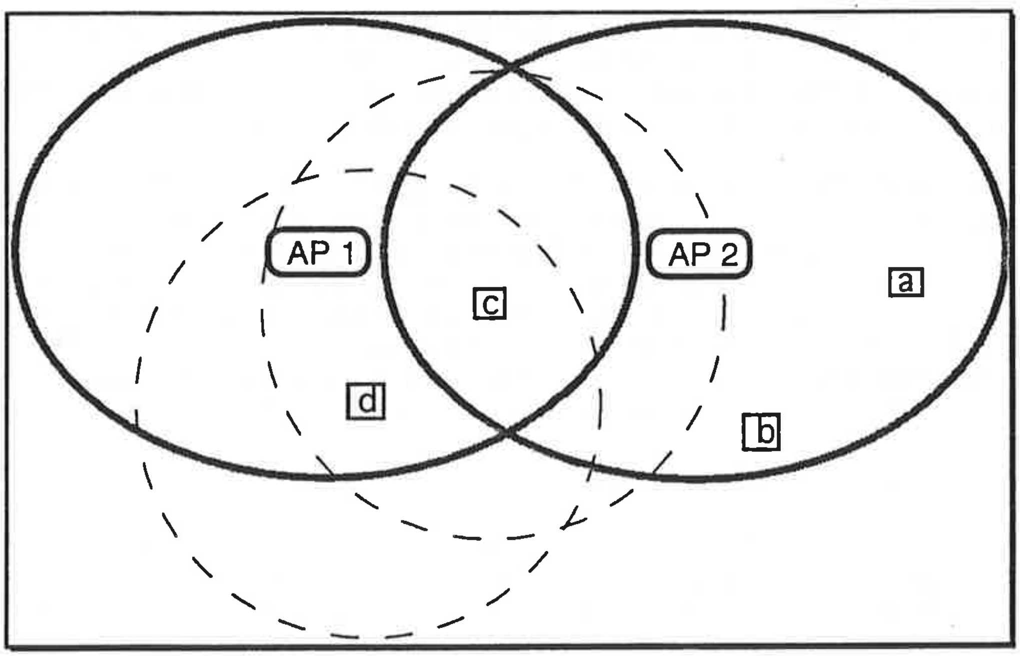


Figure 7 — Overlapping BSS with Time Bounded traffic

In the real world, the boundaries of service areas are not as well defined as in figure 6. Consider the case when node b is moving towards AP1. When node b initiated a Time Bounded connection with AP2, there were no conflicting reservations in its Network Allocation Vector. As it crosses the boundary into the AP1 service area, node b may encounter new traffic with reservations that conflict with its connection with AP2. If the conflict is for a single MPDU, node b could choose to ignore the conflict and remain silent during the conflicting period. The result would be a dropped MPDU, but the connection would continue unhindered after the first successful new MPDU exchange. If the conflict is with another Time Bounded connection, node b must abort its connection. It does so by sending a CTSO or CTSI with GAPTIME and MORE = 0. At that point it can initiate a new connection with AP2 or associate with AP1 and establish a connection there. The exact procedure and timing of a BSS transition during a Time Bounded connection is TBS.

Network Types and Communication Styles

Ad-hoc Networks

The WHAT protocol is well suited for ad-hoc networks since the coordination function is distributed among all of the participating nodes. A single STA must start the network, but once the network is established there is no master STA. STAs may join and leave the network in any order and at any time (spontaneously); and the network will continue to operate correctly. Ad-hoc networks support the Asynchronous Service only. the Time bounded Service requires APs and some preexisting infrastructure.

A station starts an ad-hoc network by periodically broadcasting Announce frames. Announce frames convey the NETID, frequency, timing, and other PHY specific information. Any other STA within PHY range will be able to hear the Announce frames and may join the network by exchanging MPDUs with other STAs on the network. As other stations join the network, responsibility for sending the Announce is spread among all STAs in the BSS. Announce frames are only sent when the network has been idle for a period of time (TBS, two maximum length MPDUs perhaps.). Every MPDU carries the information required for other stations to join a network. With a sufficient traffic level, Announce frames are not required at all.

Networks with Infrastructure

Access points connected to a distribution system are used to build an infrastructure that extends the geographical coverage of the wireless LAN. Each AP in an administration domain belongs to a separate BSS with a unique NETID. When APs are present they alone send Announce frames for their BSS.

In networks with infrastructure, the WHAT protocol supports both peer to peer communication within a BSS and hierarchical communication through the APs. This is controlled by the HIERARCHICAL bit in the CONTROL field of the MPDU header, and can be changed on an MPDU by MPDU basis. HIERARCHICAL indicates that the MPDU should only be processed by an Access Point. The following rules apply to the HIERARCHICAL bit:

- Whenever an AP hears an RTS with proper NETID and the HIERARCHICAL bit set, it sends a CTS frame and accepts the MPDU regardless of the DESTINATION ADDRESS.
- STAs other than APs will not respond to an RTS frame if the HIERARCHICAL bit is set.
- APs always clear HIERARCHICAL before forwarding an MPDU.

When a wireless STA is Associated with an Access Point, the default value of HIERARCHICAL is 1. This means that all traffic from an Associated STA will flow through the AP. MPDUs sent to another STA in the same BSS will be transmitted twice — once from source STA to the AP and then from the AP to the destination STA. When direct peer to peer communication is desired, the source STA can optimize throughput and bandwidth utilization by turning off HIERARCHICAL.

MPDU Relaying

Although an AP will accept any MPDU with a matching NETID and the HIERARCHICAL bit set, it will not automatically relay the MPDU. An AP will relay an MPDU if the DESTINATION ADDRESS does not match the AP's 48 bit ID and the SOURCE ADDRESS matches one of the STAs that is Associated with (formerly called registered) this Access Point. APs maintain a list of the addresses of all of the STAs that Sign-On; and only provide the MPDU relaying service to those STAs that are currently signed on (Associated).

The AP must then decide where to relay the MPDU. It does so as follows:

```
IF (DESTINATION ADDRESS is NOT in the list of Associated STAs) THEN
    relay MPDU to Distribution System
ELSE
    relay MPDU to wireless BSS.
```

Relaying from the DS to the Wireless LAN operates in a similar manner. The AP is capable of receiving traffic for multiple addresses. When a STA signs on to an AP, it's address is placed in a list of addresses that the AP will accept. The STAs may also add multicast addresses to this list by sending a MAC Management message to the AP. The AP will filter traffic coming from the DS and will relay only those packets with addresses in the list.

If security is desired, it is imposed during the Sign-On process. The AP authenticates STAs that attempt to sign on using a protocol similar to that proposed in [9]. If authentication is successful, the AP can optionally pass the STA a session key that is used to encrypt subsequent MPDU payloads.

Issues Addressed [9]

4.1 Will the standard support ad-hoc networks?

Yes. The standard should support ad-hoc networks. The WHAT protocol describes one way to support ad-hoc networks. Ad-hoc networks are a natural mode of operation for the WHAT protocol because of its Distributed Coordination Function.

4.2 Will the standard support infrastructure networks?

Yes. The standard should support infrastructure networks. MPDU relaying is built into the WHAT protocol, and there is a protocol for managing and configuring APs that is part of the MAC Management layer.

4.4 Does the 802.11 standard support geographic coexistence of multiple overlapping 802.11 networks?

Yes, it must. The WHAT protocol operates effectively even when there is no channel isolation for overlapping or adjacent BSAs. When traffic from different BSS is present on the same channel, STAs in the overlapping area behave as if their network is the union of the overlapping BSS. The result is that stations in overlapping areas perceive that their network is more congested than those in a single BSA. Of course this congestion can be reduced or eliminated if the PHY layer can provide channel isolation of adjacent BSAs.

4.5 Can a STA be a member of an ad-hoc network and a non ad-hoc network (an infrastructure network) at the same time?

No. At any point in time a STA is a member of one, and only one, BSS. A STA may be within range of both types of networks, but will participate in one or the other.

5.4 Is the interface to the Distribution System performed at the MAC Layer, the PHY Layer, or both?

The MAC layer. There is no relation between the wireless PHY and the Distribution System.

5.6 What is the direction for the Association Service transaction?

The Association Service flows from STA to AP. There is no need for a bi-directional service. If the AP causes a Disassociation, the STA can sign on with a different AP and cause a new Association. Only the STA knows which AP is the best one to choose for the new Association, so it does not make sense for an AP to cause an association on behalf of a STA. If we require the APs to know about the real time signal strength of every Associated STA in relation to every AP; and communicate this information through the DS in a timely manner, then we are making too many assumptions about the performance of the DS. Distribution Systems are built out of existing systems. We can not redefine their performance to suit our requirements.

5.9 How do STAs determine that Access Points are present?

The WHAT protocol handles this in two ways: 1) Each MPDU that is transmitted by an AP is marked with a bit that indicates it was transmitted or relayed by an AP. A STA observing a BSS that includes an AP will very quickly learn that the AP is present; and can attempt to sign on using a broadcast with the appropriate NETID. 2) When the network is idle, APs send out periodic Announce frames. Announce frames are also marked with the AP bit, so a receiving STA can distinguish an ad-hoc BSS from one that includes an AP.

10.1 What Coordination Function will be specified in the standard?

15.7 What is the common service? Asynchronous or Time-bounded?

Asynchronous Service should be the common service. Time-bounded Service should be an optional enhancement. A Distributed Coordination Function should be specified as the default mode of operation. A DCF is simple to implement, sufficient for Asynchronous Service, and well suited to ad-hoc networks. A Point Coordination Function should be added as an optional extension when Time-bounded service is required. The WHAT protocol is an example of this approach.

12.3 What is the intelligence level at the MAC/PHY interface?

Keep it simple. A few generic primitives with parameters to control specific PHYs.

12.8 Does a PHY independence layer need to be specified in the MAC? Yes.

24.3 How will multiple PHY support for the MAC be specified?

The intelligence should be in the MAC layer. There should be a PHY specific sublayer in the MAC to accommodate different wireless PHYs. One way to parameterize the interface is to provide a field in the MAC header that is used to pass PHY specific information across the MAC/PHY interface, and from MAC to MAC. The WHAT protocol follows this approach.

17.3 What is the extent of Multicast? (BSS or ESS?)

Both ESS and BSS multicasts should be supported. A STA should be able to explicitly control the scope of multicasts. The WHAT protocol provides this capability with the HIERARCHICAL bit.

17.5 What is meant by addressing? is IEEE 802 addressing OK?

IEEE 802 addressing is required. Wireless STAs should be identified by 48 bit unique IDs that are compatible with other IEEE 802 standards. All Asynchronous Service MPDUs carry the full 48 bit address in the WHAT protocol. Time-bounded MPDUs use a short local identifier. However, the Call Setup message for Time-bounded connections contains the full 48 bit addresses of the source and destination.

19.2 Will the IEEE 802.11 MAC look like the other IEEE 802 MACs regarding delivery reliability?

Yes, it must provide a comparable level of service to client software.

19.5 What kind of error recovery mechanisms are to be incorporated into the MAC?

The 802.11 MAC should include a positive acknowledgment protocol with low level retries. This mechanism helps the MAC present approximately the same level of MSDU delivery reliability as other IEEE 802 protocols.

25.1 Will the standard provide a procedure to reserve medium channel capacity?

Yes, the standard should provide the ability to reserve the medium. The WHAT protocol uses this technique to allow Time-bounded MPDUs to have higher priority media access than Asynchronous MPDUs.

References

[1] Ken Biba, A Modest Proposal for an Asynchronous, Data Intensive, Wireless Local Area Network, IEEE P802.11- 91/25, March 1991.

[2] Ken Biba, A Hybrid Wireless MAC Protocol Supporting Asynchronous and Synchronous MSDU Delivery Services, IEEE P802.11- 91/92, September 1991.

[3] Ken Biba, An Update to the Hybrid Wireless MAC Protocol Supporting Asynchronous and Synchronous MSDU Delivery Services, IEEE P802.11- 92/14, January 1992.

[4] K. S. Natarajan, C. C. Huang, D. F. Bantz, Medium Access Control for Wireless LANs (An Update), IEEE P802.11- 92/39, March 1992.

[5] Phil Belanger, MAC interaction with a Frequency Hopping PHY, IEEE P802.11- 93/13, January 1993.

[6] Mike Smith, Further Simulations of the Hybrid MAC Protocol, IEEE P802.11- 92/37 March 1992.

[7] Wim Diepstraten, A Wireless MAC Protocol Comparison, IEEE P802.11-92/51, May 1992.

[8] ed. David Bagby, Functional Requirements IEEE Project 802.11 Version 1.0, IEEE P802.11-92/91, July 1992.

[9] ed. François Simon, IEEE P802.11 Issues Document, IEEE P802.11-92/64, March 1993.

[10] Whitfield Diffie and Ashar Aziz, Wireless Network Security IEEE P802.11-93/08, January 1993.

Private conversations with Bill Baugh of Xircom helped clarify the use of MPDUID and other WHAT protocol concepts including the scope of RTS/CTS media reservation.