IEEE 802.11

802 LAN Access Method for Wireless Physical Medium

DATE: March 8, 1993

AUTHOR: Chandos A. Rypinski,
Chief Technical Officer
LACE, Inc.
655 Redwood Highway #340
Mill Valley, CA 94954 USA
Tel: +01 415 389 6659
Fax: +01 415 389 6746
E-m: rypinski@netcom.com

TITLE: **COMPARISON OF REGULAR AND ASYNCHRONOUS TIME DIVISION MULTIPLEXING OF THE WIRELESS PHY**

SUMMARY

Regular periodic time slots organized into frames and integer multiples of 125 μseconds have been proposed for time division multiplexing of the 802.11 wireless medium. Notwithstanding the widespread use of this philosophy in telecom, IEEE 802.9 and 802.6 and its further proposal for next generation 802.9, **this approach may result in serious negative technical and economic consequences when applied to the 802.11 medium.** In addition, there will be reduced compatibility with Asynchronous Transfer Mode cell relay technologies, and the synchronous timing is a material impediment to autonomous operation. The explanation for these assertions is the substance of this contribution.

The alternative is a purely asynchronous medium in which each new use may follow the previous use as soon as the channel is available. There is no timing structure that is a factor in when a station may transmit. Time can then be managed adaptively and in response to demand by an intelligent central controller where high capacity and utilization are required, and in a more casual way when capacity is not pressed.

Tables are shown which itemize and compare regularly slotted and asynchronous time division of the physical medium. In addition, the medium functions are tabulated for the common access method on both pair and wireless transmission mediums.

The asynchronous message-based access method has been described in previous contributions 91-19 and 91-95. In contribution 93/24 now submitted, the adaptation to direct use of 802.6 defined ATM cells is described; and in contribution 93-25, the radio PHY for this context is described.

COMPARISON OF REGULAR AND ASYNCHRONOUS TIME DIVISION MULTIPLEXING OF THE WIRELESS PHY

Table of Contents Page

Figures and Tables

COMPARISON OF REGULAR AND ASYNCHRONOUS TIME DIVISION MULTIPLEXING OF THE WIRELESS PHY

BACKGROUND

There have been thorough and important proposals to IEEE 802 for a regularly time-slotted physical medium to deliver both isochronous and packet services. They have used slots or groups of slots variously assigned to setup, data transfer and isochronous channels. Further refinements have led to time blocks for contention, inward and outward transfer modes; and adaptive boundaries between these blocks. To provide a perceived compatibility advantage with telecom, the proposed frames are integer multiple of 125 μseconds.

IEEE 802.6 and 802.9 both use time-slotted physical mediums with the 125 μsecond frames from isochronous PCM. A portion of each frame is designated for packet use. Both of these systems are examples of "PHY Layer" multiplexing of isochronous and packet services. Packets are mapped into the assigned "boxcars" of the PCM structured PHY layer.

Many of these examples of PHY layer multiplex depend on duplex transmission paths with separate facilities for inward and outward transmission.

For many proposed PHY Layer multiplexing plans, the only benefit is the saving of separate pairs (or alternative mediums) for the packet and isochronous types of service. Many do not provide the strong management functions in the IEEE 802.6 approved and IEEE 802.9 pending standards.

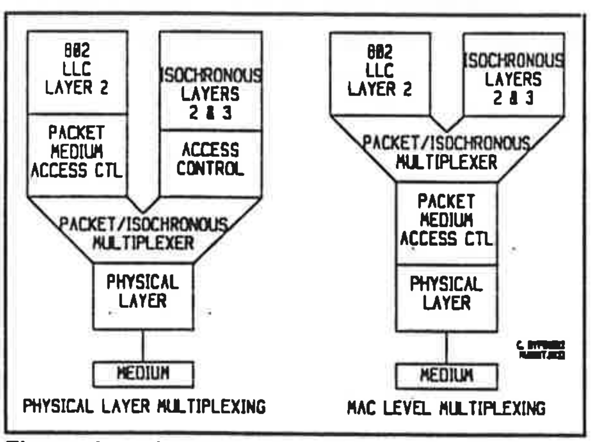

Figure 1 Layering for PHY and MAC Level Multiplexing of Isochronous / Packet Services.

The next step in the progression should be a "MAC Layer" multiplex with asynchronously transmitted short packets. With this higher layer multiplexing, the lower layer medium access method and the physical medium are common; and they are shared by both types of service.

Careful design of the scheduling control can meet the needs of the time-bounded services. This is part of the function of the MAC design. User interface transfer rates above 10 Mbits/sec reduce the difficulty of achieving low access delay.

PCM multiplex was created for longer two-ended and contained transmission paths. This is a different problem than the uncontained medium and short transmission paths addressed by the 802.11 Committee.

IEEE 802.11 can only reach a satisfactory solution with a MAC layer multiplex as shown in Figure 1. When this is accomplished, it will be apparent that the contained medium is a subset. Loss in channel efficiency and affect on cost will be small, and much less than is believed by the advocates of subset service definitions.

PRIMARY CONSIDERATIONS FOR MAC OR PHY LEVEL MULTIPLEXING

Traditional telecom PCM time division multiplex is an efficient way of providing multiple, two-way, two-ended channels when:

1) there is a high duty cycle for information transfer in both directions at the same time (probably higher than voice activity in telephony).
2) a single bandwidth is required for most of the service provided.
3) the time required for the setup negotiation is small compared with the duration of the connection maintained.
4) when the cost of the transmission facility is high compared with the cost of the terminal equipment (Usually, this is the case when there is a long transmission distance).

Alternatively, a simplex, asynchronously used short packet medium may be considered.

The single channel transmission facility is used alternately for inward and outward transmission. Use is managed by a dynamic time scheduling function and by including all necessary handling information (or reference to it) in the header of each burst. This mode of operation is attractive for the inverse of the PCM conditions which are shown in **Table I** following, and which include the conditions below:

a) when a significant proportion of the use of the connection is directionally asymmetrical (e.g., video broadcast, interactive CAD).
b) when bandwidth-on-demand is a significant fraction of the service required (e.g., multimedia computer applications), and where that bandwidth may vary considerably during the virtual connection (e.g., compressed video and voice).
c) when much of the traffic is short length data messages (e.g., retail transactions, departmental LAN).
d) where an initial setup function of more than 5-10 milliseconds becomes part of access delay. First transmission setup delay greatly reduces the value of packet data communication.
e) Far better adaptation ease and compatibility with an ATM environment.

A summary of the differences in applications which cause a preference for one of these technologies is shown in **Table 1** following.

CONTRAST BETWEEN REGULAR TIME-SLOTTED AND ASYNCHRONOUS OPERATION

It is useful to compare methods of providing functions that all systems need. In some way channel time must be designated for the type of use that is permitted to a participating station. Also, there must be designated intervals for a station to request service (probably in a contention mode).

Examples of both time-slotted and asynchronous access mediums have been presented to IEEE 802.11. While telecom time-slotted systems are exclusively full duplex (separate and parallel inward and outward facilities), offered 802.11 plans assign directionality as a slot property enabling operation in a single channel of twice the bandwidth for equal capacity.

The following examples are simplifications of these taking up the points necessary only to illustrate principles.

UNIFORM PERIODIC FRAMES SUBALLOCATED

Frame periods of 125 to 40,000 μsec have been chosen for different proposals. The smaller number is used in the high rate mediums for 802.6 and 802.9. The larger numbers have been used in 802.11 derived from the period of a frequency hop or from the aggregate time required to provide all necessary defined capabilities at the given data rate.

All of these systems need a common time reference to which all participating stations are synchronized. In the simplest case this reference is broadcast by a master station (which might be any ordinary station in a special mode). A less obvious need is that **this rate may need to be synchronized to that of the telephone network to avoid loss or addition of octets from timing slip.** Special, but not impossible, measures are needed to abate this problem. The solution favors use of a designated common station as both the source of the timing reference and as a gateway to outside isochronous networks.

An Octet Per Frame in PCM

PCM practice is to use one octet per frame (the D-channel at 16 or 64 Kbits/sec) for setup of connections. Since a D-channel message is 20-40 octets. It requires that many frames to convey a change in configuration of the channel allotments within the frame.

Consideration will show the desirability of reconfiguring channel allotments within one frame. The slow rate of the telecom signaling channel is a primary reason for unsuitability for LAN applications.

A Control Field Per Frame in LAN

If a control field contains a sufficient number of octets, the frame configuration may be changed effective with the immediately following suballocations in an open loop control logic. This adaptability is very useful, but it is implied that the frame needs to be many times longer than the control message. This has important numerical consequences.

I – COMPARISON OF TIME-SLOTTED AND ASYNCHRONOUS APPLICATIONS

FUNCTION	TIME-SLOTTED PLAN	ASYNCHRONOUS PLAN
Transmission Medium	Two-ended segments. One source and one destination.	Multi-drop segments. Always multiple sources and sometimes multiple destinations.
Predominant Service Provided	Module of transfer capability is fixed transfer rate during one connection, usually 64 Kbits/sec. Setup time not critical. Greater bandwidths accommodated by assembling parallel slots. Padding used for unfilled slots.	Bandwidth-on-demand and adaptive bandwidth. No padding. Large amount of short message traffic (e.g. sales transactions) present.
Traffic Directionality	High efficiency for full duplex symmetrical traffic. 50% only utilized for voice conversation and mainly one-way modem transfers.	No loss of efficiency for directionally asymmetric traffic like video and file transfers.
Bandwidth-on-Demand Adaptive Bandwidth	Allotted larger bandwidths cannot depend on availability of contiguous slots. Slot management complex and slow. Adaptive bandwidth service can only be provided by fixed allocation of peak bandwidth.	Inherent. Variation in bandwidth accommodated by variations in frequency or length of segments. Dimensions may be adaptive during a connection.
Set-up or Request Function	Setup message in allotted time slot. Telecom PCM transfer rate limited to 16 or 64 Kbits/sec for 30 octet setup of connections. Setup allots reserved fixed time slot for use which is long compared with setup time.	Header determined handling for setup of connections or packets transferred at medium rate. Virtual circuit identifiers used for following segments.
Allocated Relative Cost of Terminals and Transmission	High time efficiency for one type of connection favors long and valuable transmission mediums supporting higher cost channel banks and terminals.	Handshake protocol suits short distance transmission where cost of transmission and of terminals can be low at reduced medium utilization for connections and segmented packets.
Automatic Repetition of Failed Transfers	Possible for packets in some 802.11 proposals, but not provision for use on connections.	Provided by access method for packets and packet segments. NACK possible for virtual circuit segments.
Autonomous Mode	Only by requiring stations to monitor for unsupervised time slots	Only small differences from native mode.
ATM Cell Relay Compatibility	Requires buffering and segmentation for PCM slotting. Wide slot compatibility for transfer but VCI not fully utilized.	Full compatibility and use of 802.6 definitions for packet segmentation and telecom addressing.

SUBALLOCATION OF FRAMES

Some 802.11 proposals for LAN service use three types of time suballocation:

1) contention-type service request and/or transfers
2) outward non-contending transfers
3) inward non-contending transfers

Each of these category slots can then be further subdivided for multiple transfers, however this tends to make the frame very long and increases worst case access delay.

The boundaries between these may be fixed or adaptive in response to a control field. This function is better carried out when one station has a master function that is cognizant of traffic that is waiting.

Alternatively, a number of slots may follow the control field each of which can be used independently in these three ways as defined by the controller. One or more slots would be used for the transfer of an enabled station when larger bandwidth is needed. The option always exists that time use is fixed-assigned by definition, but it is unlikely that any fixed definition would be effective without an equally fixed traffic forecast.

The use of a physical medium that has time constraints on the availability of transmission time must somehow be integrated with the frame, field and suballocation process. This is a serious impediment to the use of frequency hopping mediums.

The preferred size for a suballocation is the amount of time required for one station to make one transfer. This is not a fixed parameter. If multiple transfers are made within one suballocation, then each active station needs a time specifier for it's position in sequence. This is a primary inducement for the use of slot suballocation rather than by sharable categories.

Slot Size Selection

Novell networks use a test pattern with half 48 octet messages. The payload of ATM cells is 48 octets (IEEE 802.6, B-ISDN and SMDS). It would be a commercial mistake to have a slot width that was too narrow for 48 octet payloads, and there are some known loads that would suggest up 288 octets for a necessary size.

The minimum size is subjective because it is the point where overhead from headers, handling, propagation time, clock acquisition becomes an objectionable fraction of the available channel time. The longer the slot size, the more efficiently long file transfers are handled.

Slots Too Wide and Too Narrow

The lengths of actual messages are not uniform. For data, retail transactions may be taken as example where the message lengths are quite independent of any choice of slot width that might be made.

Slot width is traditionally uniform for a connection at 64 Kbits/sec or at any other transfer rate. The better video and voice compression algorithms take advantage of the fact that less bandwidth is required to send silence on a voice channel or absence of motion on a video channel. The uniform bandwidth concept is potentially grossly inefficient in this context.

Accepting that there are regular slots dimensioned to fit an assumed traffic pattern, the consequences of misfit messages must be considered. Many slots will not be fully occupied, except when the traffic is predominantly transfer of long data files. The effect of low slot utilization can be diluted by tests with a large proportion of long message transfers, and this is often reported.

ADDRESSING PRACTICE

In telecom, the setup process associates source and destination in the tandem links composing a connection. At setup a path or channel number is temporarily assigned to that connection. After setup, the time position of the data in a multiplexed data stream is a secondary temporary identifier maintained in dynamically managed data bases at network nodes. Any change in slot position(s) used for a connection may impact other connections in other slots. Any change takes time to propagate and verify.

In LAN, source and destination long addresses are contained in the header of every packet. The time of occurrence or the facility of appearance is not used to determine handling and routing at network nodes.

The time required to deal with the secondary addressing system that is part of regular time-slotted multiplex is not required in asynchronous packet systems. The setup and slot allocation makes efficient use of the medium for long connections. It also makes a plan that is inflexible for bandwidth-on-demand and unusable for LAN type packets or adaptive bandwidth connections.

SPECIAL CONSIDERATIONS FOR MULTI-DROP TELEPHONE PAIR MEDIUMS

An expanded function subscriber loop (telecom jargon) is defined to have the following capabilities:

1) Transfer rate of 16-24 Mbits/sec
2) Reach from active repeater of 150 meters for single station Using 24 gauge DIW (PVC insulated) 25-pair cable horizontal type. Much greater reach with plenum grade or heavier conductors.
3) DC power transfer limited to 24 vdc @ .04 amperes per pair.
4) Multi-drop operation limited only be aggregate traffic and power division factors.

Provided that this loop emanates from a suitable electronic hub, it would be capable of furnishing both LAN and isochronous services up to PRI simultaneously.

As contrasted with 802.9, this expanded loop would provide multi-drop service. For example, one office could have one outlet on each of four walls daisy-chained on a single loop to the wiring closet. IEEE 802.9 cannot support this configuration for two reasons: 1) Each station and AU transmitter is ON continuously, and 2) the use of the line appearance to identify the source station is allowed. Each outlet should be usable independently (not as a party line} not exclusively.

For multi-drop loops, transmitters can be ON only when transferring traffic. Since the 802.9 loop reach is limited primarily by like-signal crosstalk rather than by impulse noise, much greater reach is possible when transmitters are quiescently OFF.

IEEE 802.9 has no provision for line-powered stations, which are desirable for the following functions:

1) local power can be OFF without impairing the ability to receive incoming calls and messages, and
2) so that a minimum telephone voice capability can be maintained during failure of normal AC mains power.
3) so that loopback testing of the physical medium and station presence can be performed from a central location.

The concentration function is equally important for both wire and wireless when high rate/capacity user ports are provided. Moreover, the pair medium will be essential to link access-points to hub controllers.

In **Table II** following the functions of the wired and wireless mediums are compared when both use the suggested access method.

SPECIAL CONSIDERATIONS FOR WIRELESS MEDIUMS

In a radio system it is absolutely necessary to transfer information in bundles of some kind. Before the first octet of information can be transferred, not less than 12 octets of time will be consumed by overhead functions, radio transmit-receive switchover and propagation time. The minimum size data bundle for the predominant traffic modes might be dictated by compatibility with B-ISDN ATM cells of 48 octets. The best possible utilization of air time would then be 75%.

In wired systems, the source address is usually assumed from the identity of the wire pair on which the signal appears. This cannot be simulated in a concentrated transmission medium. **Each station must identify itself within the content of every transmission that it makes.**

If there is to be a resource allocation, the addressed manager must have a description of the nature and size of all pending transfers. It must be impossible for one long transfer to block access to the manager to queue further requests.

Wireless by radio and to a lesser degree by optics is an unreliable medium that can be interrupted or disturbed by foreign and local interference. One consequence is that handshake procedures are valuable and probably essential. Also, excessive investment in channel time for one data transfer should be avoided.

One of the big problems for high speed transmission on telephone pairs is susceptibility to infrequent but exceptionally high levels of impulse noise. A radio access method which accepts the inevitability of some transmission loss would perform much better than traditional telecom methods that hope for but do not always get a near perfect transmission medium.

The steps described below can be part of either a time-slotted or an asynchronous system though implementation details would differ significantly.

II – COMPARISON OF FUNCTIONS AND REQUIREMENTS FOR PAIR WIRED AND RADIO WIRELESS INTEGRATED SERVICES LOCAL DISTRIBUTION

FEATURE OR FUNCTION	TELEPHONE PAIR WIRING WITH STAR TOPOLOGY	RADIO WIRELESS WITH ASYNCHRONOUS ACCESS
Medium Transfer Rate	16-24 Mbits/sec for 200-130 meters.	16-24 Mbits/sec for 50-30 meters.
Local Powered Stations	Via loop possible and necessary for power system independence.	Inherent in battery-powered stations.
Access-point Power	Same as local concentrator in telecom.	Via telephone pairs limited to 24 vdc @ 0.04 Ampere.
Multiple-users Per Port	Multi-drop for virtually simultaneous users within power division constraints.	Inherently multiple users per Access-point.
Direct Peer-to-peer	Possible on multi-drop acting as an up-down bus.	Provided by access method for single channel simplex operation.
Continuous Transmit Modems	802.9 characteristic. Not allowed.	Not used and not allowed.
Clock Acquisition and Synchronization	Upon system start-up, as in 802.6 and 802.9 and others, not allowed.	Synchronization at start of each received burst.
Remote Loopback	Required. Sets minimum function with power OFF.	Necessary to the extent possible.

STATION INITIATED FIVE STEP HANDSHAKE DATA TRANSFERS

For communication in an unreliable medium, the data transfer should not be attempted unless there is assurance that the channel can be used exclusively and that the necessary path exists. For this purpose, a station originated data transfer should consist of the following sequence of messages:

	Manager Originate	Station Originate
1)	Transmit enable	
2)		Request to transfer
3)	Grant to transfer	
4)		Data transfer
5)	Acknowledgement	

CHANNEL MANAGER ORIGINATED DATA TRANSFERS

The Channel Manager is presumed to have knowledge of which stations are reachable and in-service and when the channel may be used for any purpose. When the transfer is initiated by the Channel Manager, the transmit enable is not required. If the transfer is very long, the Channel Manager might use the request-grant steps. Otherwise they should also be omitted.

AUTONOMOUS PEER-TO-PEER DATA TRANSFERS

When there is no Channel Manager, the same steps are still useful though not necessarily required. It is better to have an originating station verify that the destination is capable of receiving before investing channel time in a long data transfer. This criteria involves numerical proportions.

The enable function can be omitted, or it can be broadcast and simulated by the first station in the group to become active. If enable is omitted, interference with in-progress data transfers or other requests is more probable. Even so this probability can be small for small numbers of stations and high data transfer rates.

LENGTH OF DATA TRANSFERS

The probability of error and of need for repetition increases with the length of the message.

A message should be of any needed length within a fixed or adaptively defined maximum. Segmentation of long packets will occur at a layer above the MAC enabling local communication with the Channel Manager. The maximum value must be set by an algorithm or a system configuration function.

TECHNICAL CHALLENGES

The described procedures are dependent upon frequent use of very short burst transmissions for the handshake messages. The medium transducer (radio) must switch from receive to send and back again very rapidly. There is also a need for fast clock acquisition. These requirements will be very difficult except for designers with considerable insight into the operation of radio and signal processing circuits.

The transition from transmit to receive should take place in less than one bit and the acquisition of bit clock should not require more than one octet of preamble though two would be defined.

EXTENDED SERVICE AREA SYSTEM ARCHITECTURE

While some designers still hope that a high megabit radio can be designed that will completely replace wiring, this is believed to be a futile effort. To provide necessary capacity and security, radios must be quite limited in range. Area coverage must be obtained from many short-reach radio access-points made as simply as possible.

For these access points to work like a distributed antenna system covering an extended area, the existing telephone pair distribution system must be used. The access-point is then a transducer from radio frequency to NRZ baseband enabling the telephone pairs to provide the interconnection function. The access-point is supplied power by these pairs avoiding a further cost for power wiring.

The value of compatibilities between telephone pair and wireless transmission becomes much greater when the cost and feasibility of the extended service area is considered.

ARQ FOR CONNECTION-TYPE SERVICES

For repetition of failed connection-type transfers, there must be a fixed tolerance allowed for access delay. This constant will include the time required for retry adaptively used. When provided, this is a unique error correction method for connections.

ASYNCHRONOUS TIME ALLOTMENT

Asynchronous time allotment has a simple meaning: A new transfer may begin when the preceding transfer is finished. With fixed-length time slots, new uses always begin at the start of the next time slot even though the previous use may have ended before the end of its slot. It is also assumed that there is no fixed length for a message, but there is a length limit.

Time-bounded transfers are noticed in the Channel Manager. It contains the algorithms that consider the isochronous 125 μsec frame period. When the instant is right for a station to transmit, the Channel Manager sends the enabling or grant message to start a transfer. The station is only required to respond in the defined manner to such messages, but has no responsibility for the timing of its transmissions.

The main function of the Channel Manager is to serialize the uses of the channel rotating successively to the different types of use. The capacity not used by one category is available to all others. Priority must be given to requests so that the Channel Manager is at all times aware of the backlog of pending traffic and the size and priority of its components. It is possible for the message succession to result in near continuous use of the channel without loss from "pad" or unused slot time.

COMPARISON COMMENTS

As shown in **Table III** following, most of the system functions are similar though accomplished by different methods whether the plan uses regular time-slots or not. All the necessary functions can also be achieved with asynchronous methods. Also, any function done by time measurement can also be done by using the managing entity to send a defining message.

Time slots based on the 125 μsec period frame used in telecom are unsuitable for the context now proposed which includes wireless and wired unconfined multi-drop mediums.

III -- COMPARISON OF TIME-SLOTTED AND ASYNCHRONOUS ACCESS METHOD

FUNCTION	TIME-SLOTTED PLAN	ASYNCHRONOUS PLAN
Timing Distribution	Broadcast of frame synch field. Each station locks and counts time from this reference, Time offset is slot address	None. No frame definitions used in autonomous mode.
Frame time use definition	Broadcast control field is one per frame of one per sub-frame. Format valid for next frame or sub-frame. Frame may be multiple transfers of like category.	Per user enabling messages and response. Enabling messages for requests and transfers from access-point. No enabling for access-point originate.
Request-grant protocol	Request message uses separate contention defined time slot(s). Grant is allocation of slot(s) for transfer or connection thereafter reserved.	Request message follows invitation message where contention is possible but of reduced probability. Grant is for immediate transfer. Connections and long packets get auto-Grant for successive segments.
Acknowledgment function for packets	In following separate time slot for packets.	Immediately follows receipt of message for packet segments.
Isochronous function	Setup of fixed periodic slot allocation. Frame space reserved after setup. No ACK for segments.	Periodic automatic Grant messages and NACK message for failed segments.
Slot width	Fixed and uniform	No slots. Time used as needed.
Frame length	One frequency hop, several slot widths	Not defined. One data transfer is-limited but adaptive length

Should it prove desirable or necessary to obtain a greater efficiency in interworking with ATM cell wide area networks, short packet payloads of four **fixed lengths could be assigned adaptively for payload sizes of 0.5, 1, 2, and 4 times 48 octets..** The longer size would be used for large file transfer, system load permitting; and the shortest size for compressed voice where minimum transfer delay is required. This modification might usefully increase utilization in the LAN without material disadvantage.

CONCLUSION

Wireless should be recognized as an important alternative within the general communication function. Unless it is considered in this way, it will not be usable within large scale networks, but would be limited to an archipelago of unrelated small user groups.

The increased versatility in service function and flexibility of transmission mediums has become a sufficient advantage to now consider the asynchronous methods as a more general solution to local distribution using one or more of the described transmission medium for a full range of services operated at transfer rates of 16 Mbits/sec or higher:

The implications of these assertions reach wireless and wired local distribution of all types including metropolitan, office, factory and residential areas and including service subsets such as LAN, PBX and service provider PCS.