Date: September 14, 1992

Title: Alternatives to Issues Related to Time Bounded Services

Authors:
Paul Congdon — Hewlett Packard
Tony Parker — IBM
Jim Schuessler — National Semiconductor

The authors wish to thank K. S. Natarajan, IBM, for his suggestions.

15.1 What does Time-bounded mean? What are the bounds?

ALTERNATIVE:

Question 1: What does Time-bounded mean?

"Time-bounded" refers to a service class for which the data being carried is subject to specific human factors considerations with respect to the end-to-end delay and/or to the temporal presentation of the data to the end user. The bounds will be application specific for each class of application being considered.

More specifically, it refers to defined time bounds placed on end to end variance of arrival of successive MAC Service Data Units (MSDU). Time-bounded specifically bounds the delay variance in arrival time from one (MSDU) to another.

Question 2: What are the bounds?

In attempting to determine the bounds for "time-bounded" services, it is helpful to consider a list of potential applications which will have time-bounded characteristics:

- telephony/teleconferencing
- audio recording/playback
- telephone answering machine/voicemail
- shared still pictures with telephony
- shared still pictures allowing mouse and/or keyboard alteration with telephony
- motion video record/ playback
- motion video with audio record/playback
- motion videoconferencing
- motion videoconferencing with lip synchronized audio
- medical telemetry

Each of these applications has different user requirements with respect to the value of the bounds for END-TO-END delay and/or PERMISSIBLE VARIATION in END-TO-END delay and/or PERMISSIBLE VARIATION between END-TO-END delay for different components of the data (e.g. between video and audio). Hence it will be helpful if different applications have the capability to request different **Quality of Service** attributes when establishing a connection on the network. [For example, to the user, this may involve a decision on cost verses resolution or color/black and white pictures. This would impact the 802.11 standard by imposing a requirement for a mechanism of reserving multiple Time-bounded channels.]

Once the END-TO-END bounds are known for a particular application, there remains the problem of determining the budget that is allocated to the wireless LAN versus the budget that is to be allocated to the connection to the Wide Area Network (WAN) and to the WAN itself and versus the budget that is consumed by the end users station (by delays imposed by system software and hardware). These network and system budgets are outside the scope of 802.11 but have been examined in order to allocate appropriate budgets for time-bounded services over the 802.11 wireless network.

The process of this work has included:

1. An examination of the human factors characteristics for the set of applications listed for which we expect to use time-bounded services and development of the END-TO-END bounds. Much of this work has been done, especially for telephony applications and is codified in existing telephony standards.
2. An examination of budgets for existing network delays (again, this information already exists) and a reasonable estimation of internal system budgets for presentation of "time-bounded" data to the end user. (Existing systems are not always reasonable.)
3. An INFORMED allocation of budgets for bounds on "time-bounded" networks has then been made with the knowledge of what application types can be supported with which choice of bounds. To better understand the components of the delay, the subsequent diagram and table are offered.

Addressing Item #1: Information has been obtained from a variety of sources.

DECT specifies approx. 27mS. round trip between the mouth-to-handset-to-base-to-handset-to-ear interface (includes 5mS of overhead for the loopback).

GSM, like DECT, can be instructive as a model. There is a 37 ms delay for channel coding and TDMA frame interleaving. The speech coder has a 20 ms delay with an additional 5 ms for actual processing. There is a delay budget set aside for the rest of the system which includes a 16 Kbps link between the mobile switching center and a base station, A/D D/A conversions, EQ, channel encoding and switching within the network. This budget is around 28 ms bringing the grand total to around 90 ms just to gain one-way access to the fixed public network! This is link C+D in figure 1.

CCITT G.103 specifies several hypothetical reference connections, one of which gives a worst case international distance of 24000 km. This implies 120 ms worst case for an optical link (reference table below), not taking the delay through the switching components into account.

CCITT G.114 gives other planning values for delay through these various network components (switches, satellites, etc.).

These include:

Transmission Medium	One-way prop
Terrestrial Coax	4 us/km
Optical Fiber	5 us/km
Submarine Coax	6 us/km
Satellite 14000 km	110 ms
Satellite 36000 km	260 ms
PCM coder/decoder	0.3 ms
Digital Transit Exchange	0.45 ms
Digital Local Exchange (digital trunk to analog line)	0.975 ms
Digital Local Exchange (analog trunk to analog line)	1.5 ms
Echo Cancelers	1 ms

G.114 states the following limitations on the one-way delay:

0 - 150 ms	acceptable. Echo suppressers (G.161 type) may be used for delays over 50 ms.
150 - 400 ms	acceptable with increasing care between 300 and 400 ms, and provided echo cancelers or high performance echo suppressers are used.
>400ms	unacceptable. However, work is referenced in the standard that shows that the experiments used to obtain these results may have been flawed. New data shows that high quality cancellation allows delays of 500 to 600 ms with no significant difficulty to 84% of subscribers.

Part E (the public network) of figure 1 (below) should be broken-up into the national and international delay components. G.114 has formulas for calculating the national extension delays for purely analog systems, mixed analog/digital and purely digital. The analog and digital recommendations are as follows (a mixed scenario will fall somewhere in the middle):

Analog national network delays will probably not exceed 12 + (Dist_km * 0.004) ms. Assume the nearest international switching center is 4000 km away (half way between San Francisco and New York), then US analog delays would not exceed 28 ms. Perhaps a more realistic distance is 2000 km creating a 20 ms delay.

Digital networks between exchanges will probably not exceed 3 + (Dist_km * 0.004) ms. The factor of 3 ms covers one PCM codec and 5 digital exchanges. The digital delays will always be less than analog.

NSC's **goal** for end to end video conferencing delays is 150mS. total (from user to user through the public network) Beyond 300mS. becomes very uncomfortable with users talking over each other. This goal is admittedly not achievable with satellite hops. We believe a 15 to 20mS skew between voice and video channels is the maximum that should be allowed. Skews beyond this get noticeable.

Information on Delay C (ref. figure 1): Pan-European Project for Third Generation Wireless Communications lists the following delays incurred in low bit rate speech encoding.

GSM (RPE-LTP)	13 Kbps	20.0 ms
D-AMPS (VSELP)	8 Kbps	20.0 ms
DECT (ADPCM)	32 Kbps	0.125 ms

T1A1.1 PCS Guidelines

In document T1A1.1/92-032R1 issued April 2, 1992 this committee put forth several delay guidelines that are of use. The following is an excerpt from that document, section 6.2, Delays and Echo:

> The delays of many wireless access systems, being considered are very likely to be greater than the few milliseconds of delay for wireline access, due to low bit-rate speech processing and digital radio channel coding. Delay can have two effects on voice performance. It increases any echo impairment as perceived by users. Even when echo is controlled, large delays can interfere with the interactivity of voice conversation. In addition, delay can impair the performance of particular voiceband data applications, some applications being affected at even smaller amounts of delay than voice applications would be. CCITT Recommendations G.114 and G.131 give additional information regarding delay and echo, and ANSI/T1.508 gives useful delay guidelines for evolving digital networks.

Connection Type	Delay Guideline	Reference
Without Echo Cancelers (PSTN echo path loss assumed)	<5mS incremental (see text below)	ANSI/T1.508
With echo cancelers	80mS maximum (objective)	CCITT G.173

> A widely accepted guideline for "incremental" delay is that any system or network element that, by itself, adds more than 5mS of round trip delay should provide echo cancellation. Assuming more than 5mS of delay is introduced due to digital wireless technology, echo control will be needed in two places. The first need would be to control echo from the wireless terminal if the 45dB WAEPL requirement of ANSI/TIA-579 is not met. Second, echo would have to be canceled from any far-end reflection in the PSTN (Public Switched Telecommunications Network) using echo cancelers, total delay should still be limited because large delays can degrade many user applications. Therefore, delay introduced by new network com0onents and technologies should be minimized, taking into account the need to provide new network and service capabilities. The 80mS maximum round-trip delay given in G.173 for PLMN (Public Land Mobile Network) systems is recognized as an objective that may not be achieved by current systems due to technological and economic factors.

Medical Telemetry: There is a standards group called the AAMI that has created an ANSI document with delay requirements for a synchronized defibrillator. The total delay acceptable between the EKG waveform and the defib is 60 ms. They allow 35 ms for the defib machine, so there is 25 ms for a digitized EKG and over the air transmission.

This information can be found in the ANSI/AAMI document DF2/1989 paragraph 3.2.1.23. It should be noted that there is no public network connection, only the inbuilding connection.

The information above was used to set the Absolute One Way End to End delay (ref. figure 1.): One satellite hop in one direction (about 260mS) produces acceptable telephony, but a satellite hop in both directions produces noticeably degraded useability for two-way conversations. According to CCITT G.111 recommendation for one-way voice communication path delay, delays above 1000mS. are unacceptable. Delays above 300mS. are of limited use, and the typical delay target is less than or equal to 200 ms. Obviously end-to-end delay for one-way traffic (i.e. playback or record) is much less important than for two-way interactive traffic (i.e. audio or video conferencing).

Addressing Item #2:

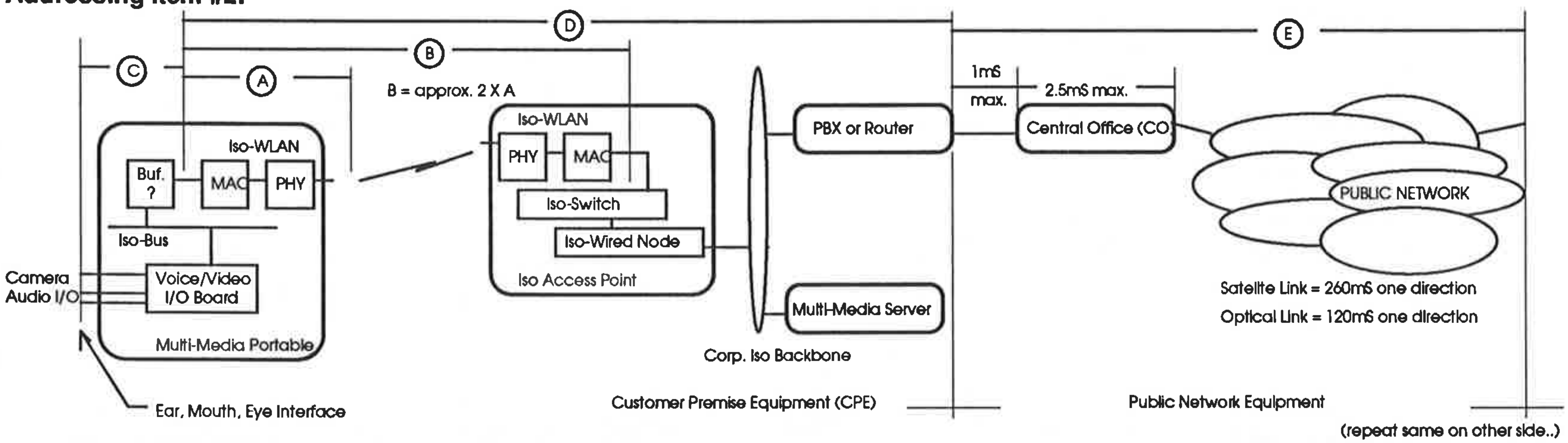

Figure 1.

Application	Absolute one way end to end delay (worst case)	Delay "A" one node delay. (x2)	Delay "B" MAC to MAC delay (x2)	Delay "C" human i/f to mac delay (x1)	Delay "D" total CPE delay (x2)	Delay "E" total Public Net. delay (x1)	Video/Voice Synch.
• telephony/teleconferencing	400mS	½ B	20-30mS (assumes 10-20mS in Corp. Backbone)	20mS (Bellcore PCS doc FA-NWT-001013)	40mS	"280mS - International w/ Satellite. 28mS - Domestic."	
• audio recording/playback	don't care	½ B		-	don't care	same as above	
• telephone answering machine/voicemail	>400mS	½ B		-	-	same as above	
• shared still pictures with telephony	400mS	½ B	20-30mS	20mS	40mS	same as above	
• shared still pictures allowing mouse and/or keyboard alteration with telephony	300mS	½ B	20-30mS	20mS	40mS	same as above	
• motion video record/ playback	don't care (if not editing)	½ B	-	-	-		
• motion video with audio record/playback	don't care (if not editing)	½ B	-	-	-		15mS
• motion videoconferencing	300mS	½ B	20-30mS	20mS	40mS	280mS - International w/ Satellite. 28mS - Domestic."	15mS
• motion videoconferencing with lip synchronized audio	300mS	½ B	20-30mS	20mS	40mS	same as above	15mS
• medical telemetry	60mS - EKG to Defibrillator	½ B	25mS	none	60mS (35mS for Defib. machine	no public access	N/A

The architecture shown in figure 1 above would accommodate all the applications listed in this table. Note the above diagram is symmetric except for part "C" where a delay is incurred in the encoding, but decoding is considered instantaneous.

Addressing Item #3:

Using 20 ms analog delays as the worst case national public network component and a single satellite as the worst case international component we see that 300 ms of the 400 ms budget for one-way propagation are used by the public network! This leaves 100 ms for components C+D+D if we stick with the original G.114 recommendation of 400 ms. However, if we assume only a domestic link (48 states), or international using terrestrial fiber, the numbers work out much better. (28mS domestic delay leaving 372mS budget for C+D+D, or in the later case, 28mS + 120mS for a total of 148mS one way delay yielding a budget of 252mS.) The table above uses the satellite based path.

Delay component C in figure 1 will primarily be consumed by A/D conversion and digital compression techniques. Taking voice as an example, a VSELP coder waits 20 ms to accumulate and compress. We now have 80 ms for D+D. Less is known about delays in emerging premises switching and LAN equipment. Although it is fairly arbitrary, based on delays in public switching equipment, we are allocating a budget of 10-20 ms for the "corporate backbone" and premises switching equipment. **Therefore, we have about 20 to 30 ms for component B (MAC-MAC delay).**

> **Motion to Close Issue 15.1:**
> <insert name> moves to close issue 15.1 by accepting the definition stated in answer to Question 1 "What does Time-bounded mean?" in document 92/xx, and by accepting the bound of 20 to 30 milliseconds as the definition of a working limit placed on MAC to MAC Time-bounded MSDU delivery.

15.2 What does "coexist with a Basic Service Set (BSS)" mean for both types of services: Asynchronous and Time-bounded.

ALTERNATIVE:

A station that uses either asynchronous or time-bounded service should be functionally oblivious to the presence of both of the following:

a) The number of co-resident stations in the basic service area, and
b) The types of data delivery service they utilize.

Being functionally oblivious means that the station will follow the same protocol regardless of the presence of zero or more stations co-resident in the basic service set. The only impact will be on the performance of the transfers with the station.

No motion to close is made, pending discussion.

15.3 What protocols above the MAC would drive the Time-bounded services? (see 12.7)

ALTERNATIVE: There are two protocols necessary above the MAC to support Time-bounded service. 1.) Data or in-band protocol and 2.) Call Control or out-of-band protocol.

1.) Many diverse protocols may interoperate with an Time-bounded-MAC, as long as that MAC will accept and generate standard 8KHz. framing. Most all voice codecs operate on a multiple of 8KHz (see Alternative to 15.1), in addition the H.261 (P*64) video compression algorithm uses H.221 framing and synchronization at multiples of these rates. Due to protocol efficiency reasons, some multiple number of samples may need to be gathered (n * 125OS) before a frame is issued or received. This would introduce additional delay, but not effect synchronization.

2.) We propose that the modified Q.931 as proposed to the 802.9 committee be used for the out of band call control signaling required to set up and tear down connection oriented services. (See Sanjay Popli's submission to 802.9 and 802.11, doc 92/81) We wish to establish a common signaling set to promote interoperability among a wide variety of voice, video and data equipment.

Motion to Close Issue 15.3:
<insert name> moves to close issue 15.3 by accepting the constraint of a provision for 8KHz. framing at the MAC to LLC interface. Furthermore, we endevor to use the call control protocol as ultimately defined by P802.9 IVD LAN. If P802.9 fails to produce, or declines to work on a call control protocol, that work will be undertaken by P802.11.

15.6 What is the algorithm for managing the partitioning of capacity between Time-bounded and Asynchronous services?

ALTERNATIVE 1: Give priority to Time-bounded up to a lower bound of asynchronous traffic. (i.e. 80% Time-bounded, 20% asynchronous minimum)

Asynchronous Load (of payload capacity)	Time Bounded Load (of payload capacity)	Asynchronous Grant	Time Bounded Grant
100%	0%	100%	N/A
0%	100%	20%	80%

Admittedly tradeoffs are application specific. We may want to leave this "unstandardized" and let the implementor decide. ..which leads to..

ALTERNATIVE 2: There is none. It is implementation specific.

PRO 1: More deterministic than #2. Besides, in systems that have no Time-bounded users, 100% of the BW would be asynchronous anyway.

PRO 2: Recognizes the obvious widely varying needs of the user base. Not specifying this parameter should have no interoperability problem with a "foreign" station, it would only change the amount of blocking the user would experience for a given available BW.

CON 1:

CON 2: Some systems could choose to allocate 100% of capacity to Time-bounded services. (Actually such a system could not work, unless call control signaling were handled through a previously established Time-bounded channel.)

No motion to close is made, pending discussion.

15.7 What is the common service: Asynchronous or Time-bounded? (common means default)

The meaning of this question is taken to be, "Given a common coordination function for all stations (due to interoperability considerations), what service is used to establish further communication?"

ALTERNATIVE:

In the case of packet data services, these same packet data services will be used to accomplish any necessary overhead associated with initialization and termination of the transaction. In the case of time bounded services (packet voice, video or medical telemetry), an asynchronous packet data service will be used for similar purposes. (e.g. The Q.931 call control protocol requires small amounts of messaging to set up and tear down a connection. This messaging would use the asynchronous packet data services.) In this context, the asynchronous packet data services can be considered common or "default".

Motion to Close Issue 15.7:
<insert name> moves to close issue 15.7 by accepting the text in document 92/xx in answer to this question.

15.8 Do all stations and infrastructures support the Time-bounded service?

It kind of depends on what is meant by "support"!

ALTERNATIVE: All stations, and all 802.11 infrastructures support Time-bounded services. However, "support" in this context does not mean all 802.11 STA convey Time-bounded channels to various user interfaces, or that 802.11 AP convey Time-bounded channels to wireline Isochronous networks. These choices are left up to the market to decide. What "support" does mean is that, given an environment in which Time-bounded services are being used between STA, AP and DS, a STA, AP or DS not choosing to convey this type of traffic will INTEROPERATE with those services it was designed to support (namely, asynchronous packet data).

PRO: Without all 802.11 stations and infrastructures supporting Time-bounded service, the usefulness of these services would be degraded to a point where they would not be used. The market would look outside of 802 for a solution and interoperability of the resulting systems would be very difficult.

CON: One may argue cost and complexity may be added to systems not desiring these services. (However, in the limit, the additional cost would be negligible.)

No motion to close is made, pending discussion.