IEEE P802.11
Wireless Access Method and Physical Layer Specification

DATE: April 30, 1991 **REVISED:** May 2, 1991

TITLE: **COMMENTS ON "A SHORT TUTORIAL ON CSMA" (IEEE P802.11/91-44)**

AUTHOR: Chandos A. Rypinski,
Chief Technical Officer
LACE, Inc.
921 Transport Way
Petaluma, California 94954 USA
Telephone: 707 765 9627
Facsimile: 707 762 5328

SUMMARY

The analysis of the Author and of the authorities on whom reliance is placed is probably accurate for the premises and assumptions used. Issue is taken with the accuracy of those assumptions in the present circumstances as listed below:

1) The radio channel properties assumed, at best, are inexact for the cases of extensive reuse of operating frequencies, and for contiguous overlapping systems.
2) There is insufficient attention to the actual size distribution of packet traffic.
3) There is insufficient distinction between use of air-time and useful payload carried in some of analytical results presented.
4) The role of propagation time is not properly considered.
5) The system model for which the assumptions are true is too specialized.

This paper considers the above points in concluding that:

> notwithstanding the references presented, there are important aspects in which CSMA is not a usable access method for IEEE P802.11.

Further, there are some less apparent assumptions which might be invalid and which may motivate the positions taken:

6) Radio power drain is much greater when transmitting than receiving.
7) Design and implementation of a receiver signal level threshold circuit is simple and cheap.
8) Alternative positive control methods will raise cost, size and power drain relative to CSMA solutions for the sub-market or 7) above.

A discussion will be offered on points 6) and 7), also.

COMMENTS ON "A SHORT TUTORIAL ON CSMA" (IEEE P802.11/91-44)

Table of Contents Page

COMMENTS ON "A SHORT TUTORIAL ON CSMA" (IEEE P802.11/91-44)

SUMMARY

The analysis of the Author and of the authorities on whom reliance is placed is probably accurate for the premises and assumptions used. Issue is taken with the accuracy of those assumptions in the present circumstances as listed below:

1) The radio channel properties assumed, at best, are inexact for the cases of extensive reuse of operating frequencies, and for contiguous overlapping systems.

2) There is insufficient attention to the actual size distribution of packet traffic.

3) There is insufficient distinction between use of air-time and useful payload carried in some of analytical results presented.

4) The role of propagation time is not properly considered.

5) The system model for which the assumptions are true is too specialized.

This paper considers the above points in concluding that:

notwithstanding the references presented, there are important aspects in which CSMA is not a usable access method for IEEE P802.11.

Further, there are some less apparent <u>assumptions which might be invalid</u> and which may motivate the positions taken:

6) Radio power drain is much greater when transmitting than receiving.

7) Design and implementation of a receiver signal level threshold circuit is simple and cheap.

8) Alternative positive control methods will raise cost, size and power drain relative to CSMA solutions for the sub-market or 7) above.

A discussion will be offered on points 6) and 7), also.

BACKGROUND

A number of P802.11 participants (e.g. Biba 91-25 and Altmaier, Mathis 91-35) have expressed strong preference for the carrier-sensing CSMA access method asserting that it is proven, documented, workable and readily implemented. The strongest expressions have stated that this process is "good enough" for the needs of the portable computer market, and that there is a market window making time of the essence.

These proposals have lacked detail necessary for quantitative evaluation, however, a pending contribution (Altmaier[1]) reviews some prior analysis of this access method. The general tone of the figures is that 50% air-time utilization is practicable with good throughput, and that the CSMA method has been used on radio by the ARPA Radio Network from which it should be concluded that it can also be used on 802 wireless LAN.

There has been some resistance to CSMA in the form of proposals for alternate methods (Cheah[2] and Rypinski[3]) in which contention is allowed for an access request but not in communication space where data transfers take place.

To simplify discussion, the only question that will be taken up now will be <u>whether CSMA is properly characterized by the references given</u>. The question of what other protocol might be better is left to other contributions, but it is assumed that an alternative exists.

A disclaimer is entered that unintentional errors are possible in the interpretation of the information given, because of limited time availability for this topic.

SUBJECTIVE EVALUATION OF INFORMATION PRESENTED IN 91-44

The subject addressed is a well presented summary, and it is mainly based on attached Kleinrock 1975 papers (see p. 3). Only figures and highlights are shown so it is possible that additional insight and information would be contained in the presentation. The content is a widely accepted analysis which seems to have been extrapolated and applied beyond its range of validity.

Comments on Assumptions Used by Kleinrock Part I

[Square brackets are used below for contributor's comments.] Some of the key assumptions deserving closer examination are:

1) "*The radio channel ... is characterized as a wide-band channel with a propagation delay between any source-destination pair which is very small compared with packet transmission time.*" In the footnote on this sentence, the example dimensions given are for 10 miles, 100 Kb/s and 1000-bit packets.

   [44 p. 7 -- the example is unclear since no propagation time is given. For a packet length of 50 octets, including overhead, transmitted at 1 Mb/s the transmission time is 400 μseconds and for a round-trip propagation time of 10 μseconds or 1,500 meters range in air, (2/3rds in wire) a = .025--and also at 10 Mb/s and 150 meters range. In using the presented figures, a range of .005 to .05 for "a" seems more reasonable than the very small number in the example.]

2) "*The time required to detect the carrier due to packet transmissions is negligible (that is zero detection time is assumed).*"

   [This is an important inaccuracy to be further discussed under radio implementation. Whatever the delay is for the transition between absent-present states, its effect is similar to additional propagation time. Also delay on the release transition can be many times longer than the detection time. These time intervals add to the width of the contention time window.]

3) "*All packets are of constant length,*

   [the length distribution is now known to have two probability concentrations around 16 or 32 octets payload and around the maximum frame size allowed (if 1000 octets or less). The efficiency of handling short packets is a significant factor.]

4) *and are transmitted over a noiseless channel (i.e errors ... caused by random noise ... are neglected in comparison with errors made by overlap interference.)*"

   [If there were no loss from overlap, the channel errors again matter. This is particularly relevant if there is an ACK function, and resend is frequently used. There is also the loss from overlap of contiguous independent systems. Burst errors in the radio channel are not negligible except compared with an already unsatisfactory system.]

5) "*... the propagation delay (small compared to packet transmission time) to be identical for all source-destination pairs.*" In a footnote it is noted that the uniform assumption of the worst case propagation delay is on the pessimistic side.

   [An inefficiency of time-slotted systems is that the worst-case propagation delay is built-in, and this becomes increasingly significant at higher data rates. Other plans might allow effect from propagation delay proportional to the actual value which is far smaller than worst case.

By using the uniform packet-length, the analysis becomes minimally different from a time-slotted system.

The further elongation of apparent propagation delay by the carrier sensing function is inaccurately neglected by this approximation.]

6) *"Under steady-state conditions, S can ... be referred to as the channel throughput rate. ...if we were able to perfectly schedule the packets into the available channel space with absolutely no overlaps or gaps, we could achieve a maximum throughput = 1."*

[the throughput referred to is use of air-time, not payload carried. At high submitted load, a fair amount of the utilized air-time is spent on mutilated packet transmissions and resend which is counted as a credit by this measure. Also, this measure masks the difference in payload capacity between traffic predominantly short or longer packets. These are major inaccuracies in the analysis for high offered loads.]

7) *"Since conflicts can occur, some acknowledgment scheme is necessary to inform the transmitter or its success or failure."* In a footnote: *"The channel for acknowledgment is assumed to be separate from the channel we are studying (i.e. acknowledgments arrive reliably and at no cost."*

[Other analysis includes the air-time cost of the ACK, and the protocol to do it well. The time cost of this function is least negligible for short packets. Use of ACK causes resends of errored packets. Errored packets are more likely when they are long,]

8) *"Assumption 1: The average retransmission delay is X[overline] is large compared with T [duration of one message]."*

[the delay must be large and sometimes very large or the system could not work. If the delay were short, once contention occurred, it could not clear. For the LAN without collision detection, large is large compared with the longest permissible message, say 1500 octets requiring 12 milliseconds at 1 Mb/s.]

Main Difficulties

For the data presented the main cloud is on the S values for high levels of G, because:

S is not carried payload but utilized air-time, and because high values of G are associated with high values of errored packet frequency, of medium access delay and of resend attempts.

In the basic assumption (point 4 above), it is clear that the number of packets lost from contention is large, by definition, compared with the channel error rate; and the radio channel error rate is too large to meet 802 requirements without treatment.

With the distributions used, there may be a considerable difference between average delay and 1% probable peak delay.

It is possible for all of the air-time to be used by resends if the system is misproportioned, and good proportioning may be limited to one size of packet.

The peculiar representation of propagation time as a fraction of the packet length is of further concern. This works mathematically for a constant length message, but correspondence with measured data might be fortuitous depending on the details of the test, particularly the distribution of propagation time and packet lengths.

Comments on Persistent Procedure

Kleinrock: *"The 1-persistent protocol is devised in order to (presumably) achieve* *acceptable throughput by never letting the channel go idle if some ready terminal is available. ... 2) if the channel is sensed busy, it waits until the channel goes idle ... and only then transmits a packet."*

"In the case of 1-persistent CSMA, we note that whenever two or more terminals become ready during a transmission period, they wait for the channel to come idle ... and then they all transmit with probability 1 [at once]." A finely slotted delay interval is proposed to spread ready terminal access attempts over time after the end of the current packet, but not as much time as *"large compared with packet length."*

The behavior modification from persistence indicated are qualitatively believable, however the time-cost of the backoff delay makes a substantial opportunity.

Hidden Terminal Problem

Kleinrock (I): *"Throughout this paper it was assumed that all terminals are within range and line-of-sight of each other. A common situation consists of a population of terminals all within range and communicating with a single "Station" (computer center, gate to a network, etc.) in line-of-sight of all terminals. Each terminal, however, may not be able to hear all other terminal's traffic. This gives rise to what is called the "hidden-terminals" problem ... [which] badly degrades the performance of CSMA ..."*

Comments on "*Hidden Terminal*" Problem and Busy Tone Solution--Kleinrock Part II

There is no need to argue with the conclusion of degradation for the model chosen. That is so. What may be arguable is this contributor's assertion that in typical systems most of the terminals will be hidden from each other because of details in the radio system design rather than walls between them. The "shared repeater" (in the jargon of land mobile radio) has far greater range than is available between stations. Moreover, the common Station (jargon used by Kleinrock above) or the Access-point (jargon by Rypinski) is essential to provide access to infrastructure data bases--the primary function on the University Campus.

The use of a separate busy-tone channel with enough bandwidth to be fast enough is thought to be costly, ineffective, spectrum inefficient and unnecessary; and it will not be considered further. It has been known for 35 years that "busy" information is unreliable--it is "available" information that is needed.

General Evaluation of the Kleinrock Papers

However brilliant these particular Kleinrock papers may have been at the time written (and they are considered classic), there are so many questionable assumptions that the results may only be applied to current 802 objectives with limited quantitative significance. They do not prove that CSMA is either efficient or stable.

The central assumption of CSMA is that: channel-busy may be detected reliably and in negligible time. As will now be taken up, this is untrue when the radio system design is fully considered.

For a modern distributed contention access method, refer to Dr. David Goodman and S. X. Wei[4]; and there the contention is separated from the data transfer capability.

COMMENT ON ARPA RADIO NETWORK

This contributor was shown an ARPA developmental packet radio system planned and designed by Stanford Research Institute about 1975. It used spread-spectrum and was operated at "pristine" government microwave frequencies. The radio throughput rate was 2400 b/s, and the range was the southern half of the San Francisco bay urban area. There was one base station and a handful of mobiles.

It would be impossible for me to see any connection between the pre-1980 government packet radio systems, and the present problems of P802.11. More likely, the methods used in government systems would be suspect for application in unrelated environments.

COMMENTS ON POSSIBLE RADIO LINK ASSUMPTIONS

A number of 802.11 participants are much concerned with portable computer applications, and more particularly the no-infrastructure peer-to-peer capability perceived as a very large market. For this case, there is concern that any requirement for periodic transmissions from Stations will significantly increase battery drain which the use of CSMA would avoid.

This group might also assume that Carrier Sensing circuits are easy and quick.

In the context of CSMA, these questions are addressed as interrelated system and radio design technology below.

The radio technology stated is the opinion of the contributor and LACE staff.

Transmitting Power Drain

The main irreducible radio power drain has been found to come from the receiver Local Oscillator at microwave frequencies. +3 to +10 dBm is required for diode mixers, and it is only a little lower for known dual gate FET mixers. This observation applies to a fixed frequency oscillator--one that is frequency-hopped would take more power for the synthesizer.

A separate transmitting oscillator cannot be turned on and off fast enough because of energy storage in the frequency determining element and other considerations. It is possible to use the receive LO for transmitting either by reversing the up-down conversion chain or by fast frequency shift using a VCO shifted by the IF frequency.

A transmit amplifier to a level of more than 0.1 watts would draw significant power. If the equipment is normally used with infrastructure, the fixed receiver-antenna function can be much more elaborate than on the portable computer. It is then beneficial to have an <u>asymmetric transmission plan with low power at the portable</u> and higher power at the fixed point. The minimum receiver-antenna advantage at the fixed access-point is 6 dB, and it is likely to be 10 dB or more. If 0.1 watts is required between omni-directional antennas, the required power at the portable computer is 10 to 25 milliwatts.

At this level, direct group communication will be satisfactory if no walls or trees are in the way.

Given the low power and use of the same oscillator for transmit and receive, <u>the increase in battery drain during transmission will be either zero or not much</u>.

Also, the minimization of transmitter power output of the portable transmitter is vital to battery drain, to dynamic range requirements of the portable receiver and to minimization of inter-system interference.

Carrier Sensing Circuit

The limit on speed is the rise and fall time of the receiver passband IF filter(s). More selectivity is worse. One μsecond rise-time to 95% of final value might be associated with two-tuned-circuits bandpass with 2 Mhz between the 3 dB down points. There is also delay between input and output. Doubling the bandwidth would halve the rise-time. Increasing the number of poles (tuned-circuits) will increase delay and rise-time.

There is a matter of threshold or decision level. The obvious first choice is slightly above the sensitivity of the receiver, but at this level considerable false operation from noise bursts is observed. In exceptionally noisy environments, false busy will be a permanent state. If the threshold is set far above sensitivity, then usable signals may not be detected.

The experience of the designers is that <u>the absolute value of the radio signal level is unusable as a measure of signal present-absent</u>. Without special and atypical care, this type of circuit tends to be unbuildable in production and unstable in use.

The universally used <u>criteria is output signal-to-noise ratio</u> which in FM receivers is inferred from limiter-discriminator quieting. In linear receivers this is not so easy. Linear receivers for signals with a varying amplitude envelope commonly use automatic gain control accomplished with feedback from an output signal level measurement. To make the speed of this circuit within an order of magnitude of the rise/fall-time of the IF selectivity requires a very high order of design capability and additional circuit refinement like balanced voltage-controlled-attenuators or multiple cascaded small range stages. The level detector may have to be special to interpret signal level from an input waveform that is level coded.

If receive levels are measured by ordinary diode dectors, they are peak-reading, fast-charge, slow-discharge circuits which would be unsuitable. There are better circuits, but they are not always used.

Circuits for this function are notorious for false detect and slow undetect. False operation is often traced to causes that appear at the antenna and not circuit faults. These circuits are not normally smart enough to distinguish true signal from impulse noise.

A circuit solution to this function is not offered here. The point is that a fast circuit to detect absence of signal is not obvious or simple . The setting of the threshold for this circuit is not obvious or simple either.

Dynamic Range Function

All radios have a certain dynamic range from the weakest to strongest signal that can be properly processed. When overloaded by signals that are too strong (e.g. two portables a few feet apart), the recovery to a proper receiving capability make take a long time unless the design takes this need into account.

A portable which has gotten service from the wrong access point creates a dynamic range problem.

This subject can only be mentioned in this context. A full treatment of implications is reserved for another contribution. The radio system plan must take into account a need for minimization of receiver dynamic range.

Radio System Plan Considerations

The range at which a transmitter signal is detectable might be ten times the reliable service range.

Any system with contiguous like systems all on the same channel will be subject to a high degree of false signal-present indications unless some higher level of coding is used that distinguishes systems. Once coding must be read, it is pointless to sense carrier.

CONCLUSION

The CSMA access method would be satisfactory only for isolated small groups of autonomous users with a small traffic potential. It would not be usable for either the large University or Shopping Mall models. The references cited in 91-44 uses assumptions which preclude its relevance to the general problem of P802.11.

There is a fundamental choice of whether the radio range is "horizon-limited" or "interference-limited" as in cellular telephone or in any spectrum efficient radio system. CSMA is inoperative in the interference-limited environment that must eventually be prevalent.

REFERENCES

1. "A Short Tutorial on CSMA," P. Altmaier, May 1991, P802.11/91-44 (Advance copy received April 27)

2. "A Proposed Architecture and Access Protocol Outline for the IEEE 802.11 Radio LAN Standards--Part II," Jonathon Y. C. Cheah, March 91, P802.11/HH/8 (paper distributed at the meeting but not presented)

3. "Access Protocol for IVD Wireless LAN," C. A. Rypinski, Feb 21 91, P802.11/91-19

4. "Efficiency of Packet Reservation Multiple Access," David J. Goodman and Sherry X. Wei, IEEE Transactions on Vehicular Technology, V40-1, Feb 91