

IEEE 802.11

Wireless Access Method and Physical Layer Specifications

TITLE: MORE COMMENTS ON CSMA**DATE:** July 8, 1991**AUTHOR:** Ad Kamerman

NCR Systems Engineering B.V.
Zadelstede 1-10
3431 JZ Nieuwegein
The Netherlands
Phone: +31 3402 76666
Fax: +31 3402 39125

SUMMARY

This contribution comments on performance aspects of various CSMA protocols. The presented information is based on literature, former contributions, analysis for product development and experience with measurements of products.

A number of CSMA techniques have been mentioned in former contributions. For the throughput performance of a CSMA protocol it is fundamental that the time between the moment a station has initiated a transmission, and the moment another station detects carrier activity, should be short compared to the packet transmission time. The propagation delay with regard to carrier detection does not correspond directly to the propagation delay in the data path.

The CSMA technique can be enhanced by slotting and by a backoff method according to p-persistence or collision avoidance. An adaptation in the p-persistence or collision avoidance approach which depends on the estimated load, can improve the throughput performance.

Network load as generated in common LAN performance benchmarks differs in general from the one in theoretical analysis. The performance benchmarks lead to a request-response situation, with requests from a number of heavy loaded workstations. The request-response interaction gives a buffered load and a controlled process for the 'effective' arrival rate of packets. Most performance benchmarks result in throughput figures different from the ones in theoretical articles based on a large number of lightly loaded stations.

1. INTRODUCTION

Former contribution [3] mentioned advantages and disadvantages of using CSMA in general. Former contribution [4] gave a quantitative observation. Figures presented in [3], [4] (and [5], [6], [7]) illustrate the dependency on the normalized propagation delay. This contribution will look to this normalized propagation delay parameter in more detail. Further this contribution discusses enhancements for CSMA like slotting, p-persistence or collision avoidance and adaptation in p-persistence or collision avoidance scheme. Next this contribution compares the model for offered load as applied in common theoretical analysis from literature with the effective load in common LAN performance benchmarks. Finally the error recovery aspect is mentioned.

2. THROUGHPUT PERFORMANCE AND NORMALIZED PROPAGATION DELAY

The normalized propagation delay is generally presented as parameter a defined by the following equation:

$$a = t_d / t_p ; \text{ where } t_d \text{ is the propagation delay and } t_p \text{ is the packet transmission time.}$$

The normalized propagation delay is very relevant for the network throughput performance. This behavior is illustrated in well known figures (Figs. 3 - 5). At the low end side where $a = 1$, the value of the maximum normalized throughput TP_{max} goes to 0.18 (like pure Aloha) and 0.36 (slotted Aloha). At the high end side where $a \rightarrow 0$, the value of TP_{max} goes to 1.00.

The term normalized propagation delay is confusing. For network throughput analysis on the MAC layer or more precisely at the interface between MAC and PHY, the time required to recognize a carrier signal is very important because of the collision risk. The duration of the interval related to this risk is the time between the moment a station has initiated a transmission, and the moment other stations have recognized that transmission. So the moment a station initiates a transmission (point of no return), and the moment all other stations have made a reliable carrier detection, give the edges of the interval in question. This interval can be regarded as the period during which collisions can be generated. This interval gives the abovementioned 'propagation' delay t_d . It should be small compared to the packet transmission time.

The 'propagation' delay reflects an end-to-end delay between MAC-PHY Interface at the transmitting station side (say LAN controller activates RTS) and MAC-PHY Interface at the receiving station side (say LAN controller gets activated CRS). In a radio LAN the 'propagation' delay has contributions from the transmitter side, the through-the-air delay and the receiver side. The major contribution comes from the receiver side. For a minimum 'propagation' delay the contribution from the receiver has to be small; so fast initial gain adjustment, fast and reliable carrier detection is needed.

The contribution of the through-the-air propagation will be small (say it will be less than 1 μ sec - 300 m). The end-to-end delay in the data path is less relevant, because the carrier detection can already be made, when a fixed preamble signal is received. This preamble signal has a fixed pattern and is ready-for-transmission before data is transferred from MAC (LAN controller) to PHY (transceiver). The period reserved for preamble and postamble symbols influences the total packet transmission time. For the effective throughput only the part of the packet reserved for data is relevant. So the effective (normalized) throughput can be found using a proportional correction for the (normalized) throughput related to the 'propagation' delay and total packet transmission.

With today's technology a cost effective 'propagation' delay of 10 - 20 μ sec is possible. Fig. 1 illustrates the individual contributions of the 'propagation' delay; 'transmit carrier' delay, 'medium propagation' delay and 'carrier detect time'. The 'transmit carrier' delay is technology dependent, the 'carrier detect time' is symbol time dependent.

For a raw data rate of 4 Mbps and packets of 512 octets, the total packet transmission time could be 1.1 msec (1 msec for data symbols and 0.1 msec for preamble, postamble, interframe spacing symbols). This brings a in the range 0.009 - 0.018 and gives for non-persistent CSMA by Fig. 5 a TP_{max} score of 82 % at $a = 0.009$ and 78 % at $a = 0.018$. Since the total packet transmission time is 1.1 msec and 1 out 1.1 msec is used for data, a correction factor of 0.91 has to be applied, which gives respectively an effective TP_{max} score of 75 % and 70 % .

2. SLOTTED CSMA

Slotted CSMA has a higher throughput than non-slotted CSMA, because slotted CSMA has a smaller collision window.

The slot time duration has to be based on the 'propagation' delay. When some station has initiated a transmission, all other stations have to be able to detect carrier signal at the end of the slot time interval. In slotted CSMA the 'collision window' corresponds to a slot time interval. There will be minor effects from timing offset between individual stations and from the through-the-air delay, which depends on the distance. These effects have to be covered by the slot time.

The slotting could be based on resynchronization during a transmission. Then a station will still try to walk away slowly because of some local oscillator offset. Only after a long period without transmissions a significant misalignment may occur. At a certain network load the synchronization is reestablished frequently and so an insignificant misalignment occurs.

Fig. 5 gives for slotted non-persistent CSMA a TP_{max} score of 86 % at $a = 0.009$ and 83 % at $a = 0.018$. The correction factor of 0.91 gives respectively an effective TP_{max} score of 79 % and 76 % .

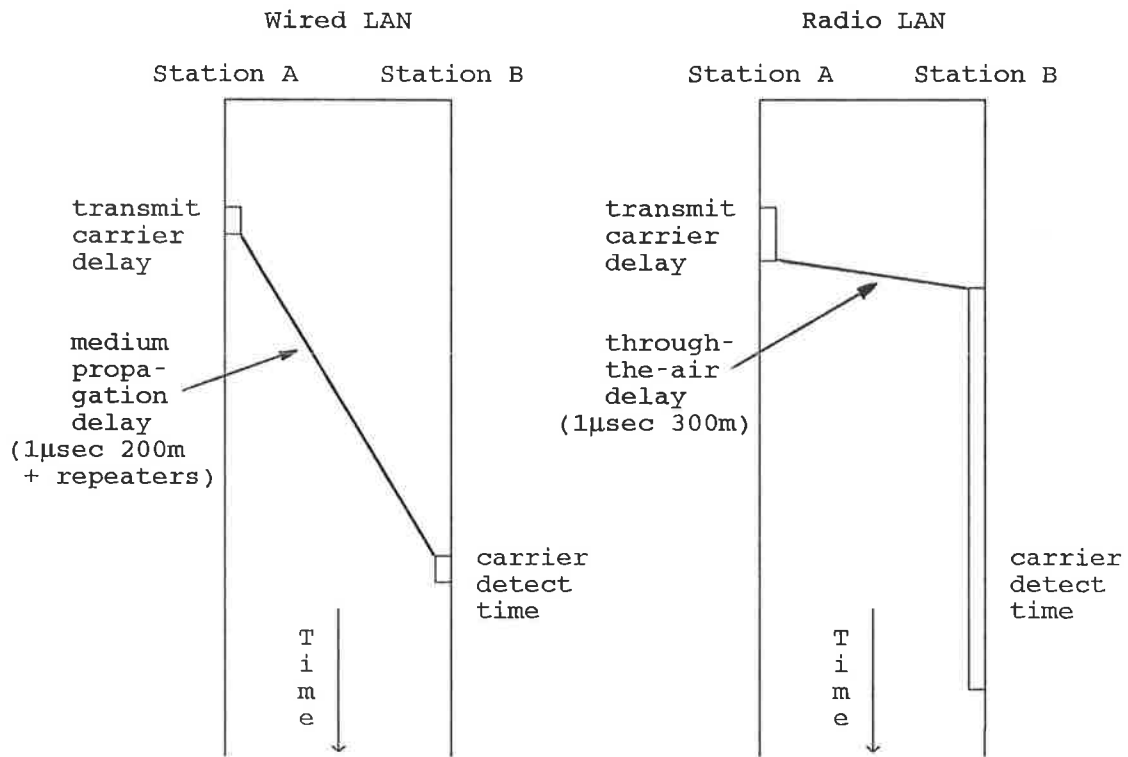


Fig. 1: 'Propagation' delay for wired and radio LAN

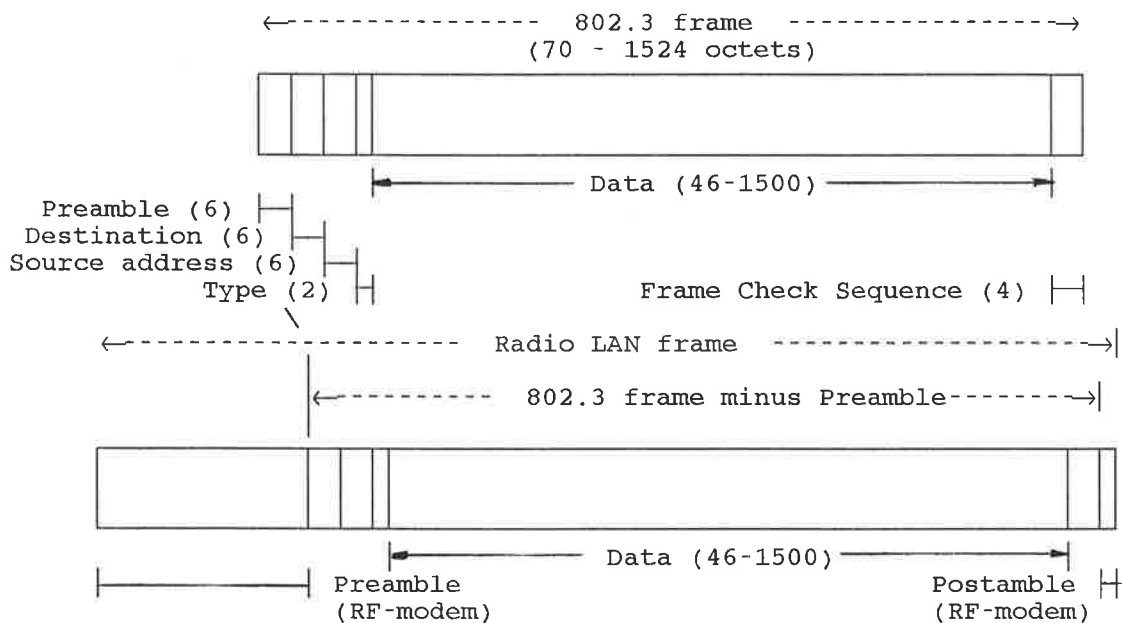


Fig. 2: Data encapsulation in 802.3 and radio LAN packet

Throughput

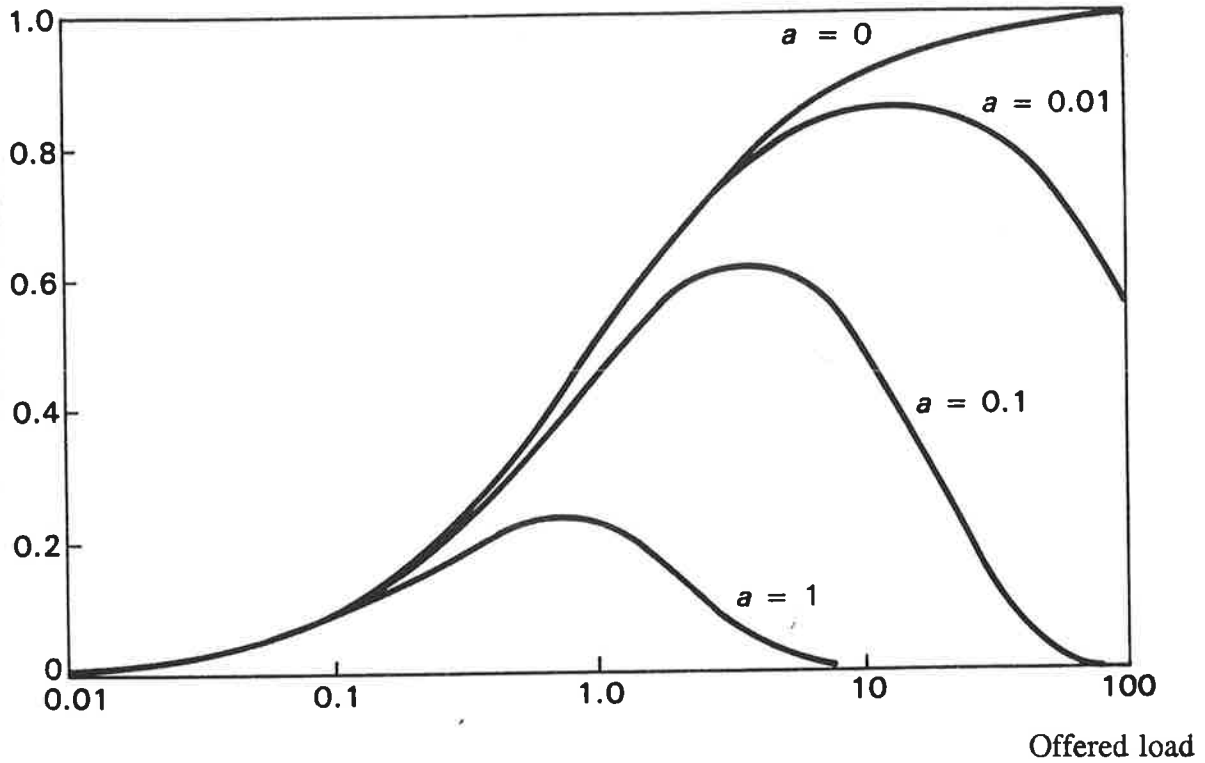


Fig. 3: Throughput versus offered load for slotted nonpersistent CSMA (from [5], Fig. 9.20)

Throughput

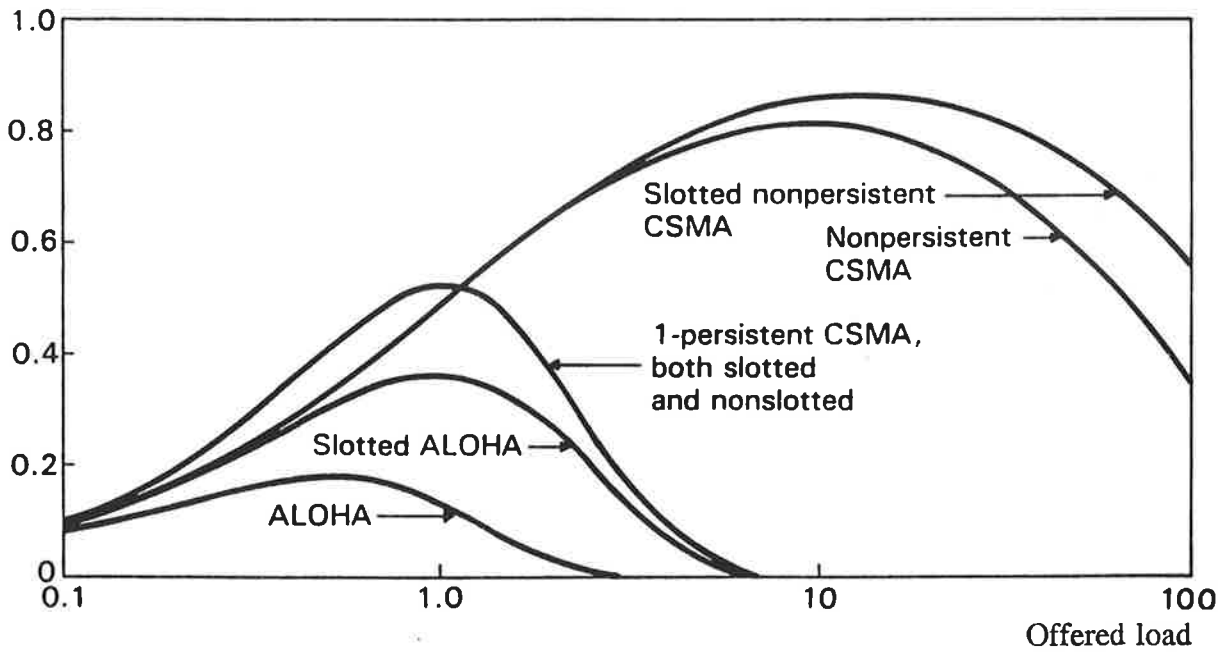
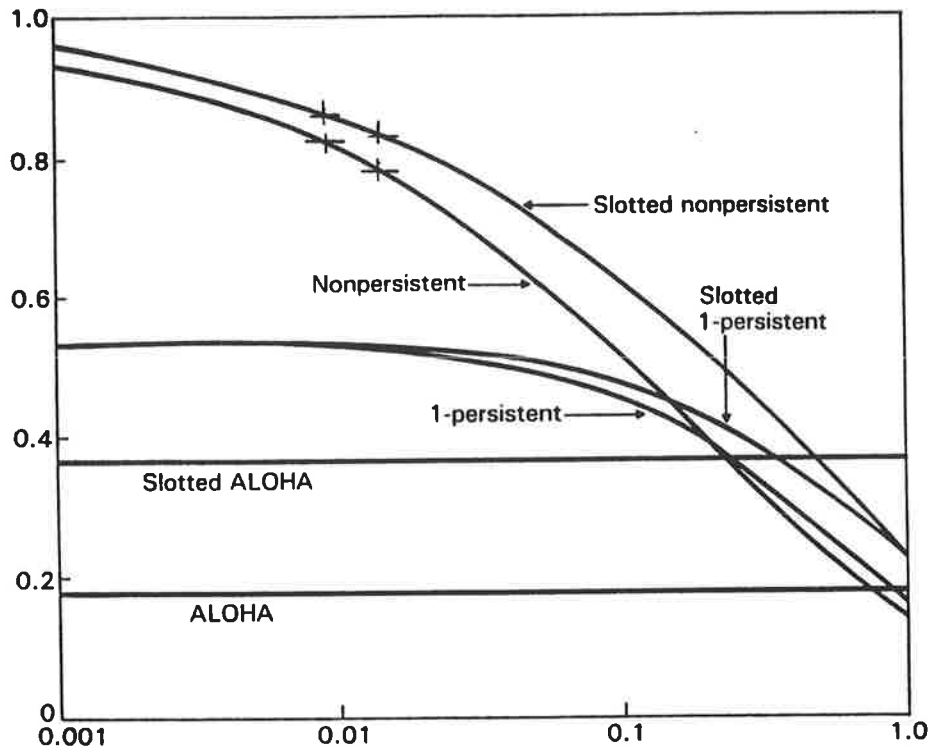


Fig. 4: Comparison of various CSMA techniques (from [5], Fig. 9.21)

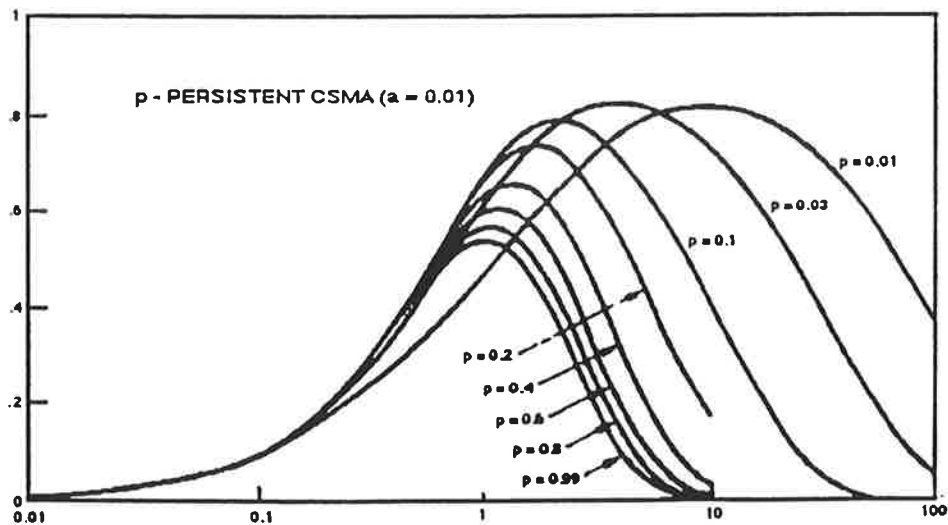
Maximum throughput (TP_{max})



Normalized propagation delay a

Fig. 5: Maximum throughput versus normalized propagation delay for various CSMA techniques

Throughput



Offered load

Fig. 6: Throughput for unslotted p-persistent CSMA for various values of p (from [3], page 9)

3. ADAPTATION IN P-PERSISTENCE AND COLLISION AVOIDANCE

In a p-persistent scheme a station which has sensed carrier activity, defers initially. After this station has noticed end of transmission, it will transmit with a probability p . With a probability $(1 - p)$ it will wait for a slot time interval.

In a collision avoidance scheme a station which has sensed carrier activity, also starts with deferring. After this station has noticed end of transmission, it will wait during a random number of backoff slot time intervals. This random number is chosen from $\{1, 2, \dots, R\}$.

Fig. 5 shows for p-persistent CSMA the dependency of the throughput on the value for p . When there are N stations which have deferred their transmissions, the risk of a collision for respectively p-persistent and collision avoidance will depend on $(N \cdot p)$ and (N / R) .

To guarantee a limited risk on collisions the values of p and $(1 / R)$ could be adapted.

The measured carrier activity during the last time gives an indication for the desired values for p or $(1/R)$. Considering in Fig. 6 the curves for $p = 0.01, 0.03, 0.1$ etc., an adaptation can result in a maximum-combiner curve. Such adaptation will result in better throughput and more stability.

4. LOAD MODEL

Most literature (current references and articles) gives theoretical analysis based on a load process with a very large number of stations, where each station generates the same very small load contribution according to a Poisson arrival process. However, commercially available wired and wireless LAN products are generally evaluated by performance benchmark measurements for a network configuration existing of a server and number (say 6) of workstations. Such benchmark represents a peak load situation for a LAN with many stations.

Contribution [1] listed the most common LAN applications; file sharing, device sharing, software sharing, gateway access, E-mail, client-server database. That contribution described also the principal underlying access method in most of the mentioned applications. This common access method is based on datagram request/reponse.

When a server receives a request from an individual workstation, the server will return with a response before a next request can be issued by that workstation. With one server and one workstation there are no collisions possible, this situation compares to a half-duplex connection. With one server and a number of workstations a server-transmission can collide with a workstation-transmission, and two workstation-transmissions can collide.

In a throughput performance benchmark there exists a buffered load situation. Results will differ from what is found by theoretical analysis for many stations and Poisson arrival processes. If there would be a throughput performance benchmark where the request and response packets have the same length and contain only data, then this benchmark would give a better result than found by theoretical analysis. However, actually the request packets are relatively short (say 80 octets), the response packets are longer (say 512 octets) and in general a benchmark measures only transferred (higher layer) data encapsulated in the response packets. Therefore conclusions with regard to throughput figures based on theoretical analysis, might be not applicable for throughput performance

benchmark.

In real life the activity on the LAN will correspond to a variable number of workstations which are busy with request-response interaction to a server. For such a situation and a certain load a model could be made using a transition rate diagram. This transition rate diagram can reflect the probability figure for the number of workstations, with buffered load, actively involved in request response interaction (not waiting for human interaction). The load in terms of job arrival rate and job size, and the throughput per station for 1, 2, 3 and more workstations need to be available. Fig. 7 illustrates an example of such a transition rate diagram. This approach would give a relevant upper bound for the number of workstations which can actually be active parallel in time, while the performance meets the requirements. A future issue could bring light on this.

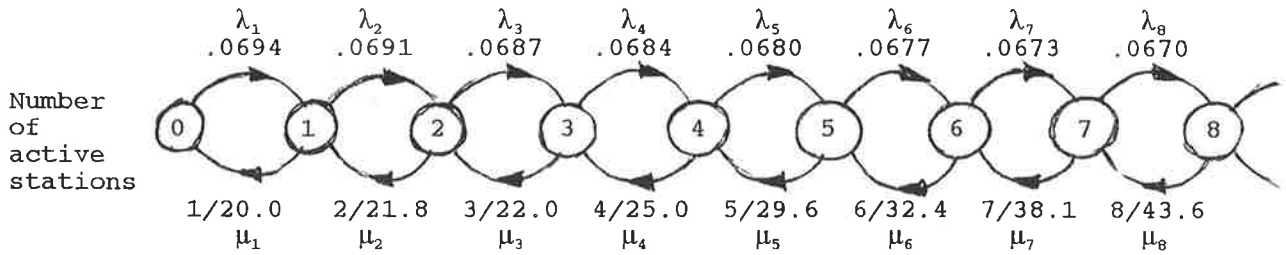
5. ERROR RECOVERY

A slotted CSMA technique with an enhanced p-persistent or collision avoidance scheme will still be subjected to the undetected collision phenomenon. Due to the capture effect, a collision does not always cause that both packets are mutilated. A packet which is mutilated by a collision or a noise burst, will be discarded on the MAC layer. So a mutilated transmission gives the loss of a packet on the LLC.

When LLC type 2 (connection-oriented service) or LLC type 3 (acknowledged connectionless service) is present above the MAC, there will be a consistent protocol stack with regard to collisions. It is also possible not to implement LLC type 2 or 3 and to provide a recovery procedure on a higher layer. A major Network Operating System product has, for common network applications, a recovery for request-response traffic on the higher level layers (transport-application). Without response a new request is made after a certain period. After a collision this approach results in a time-out situation for the workstation in question. With a high load there will be more such time-outs and the number of medium access attempts is reduced, which prevents a fallback in the total network throughput at cost of station throughput.

Another outstanding issue is Compatibility requirements section of the PAR (Project Authorization Request - version May 1991). The PAR specifies that the MAC Service Data Unit loss rate shall be less than $4 * 10^{-5}$ for an MSDU length of 512 octets (for at least 99.9 % of the time and in 99.9 % of the total geography of the service area). Does the PAR specify a clear requirement with regard to loss of packets by collisions? If this maximum loss rate of $4 * 10^{-5}$ is a MAC layer requirement, then some sort of an Acknowledge or Collision detection protocol implementation is necessary.

a. Transition rate diagram



b. Histogram

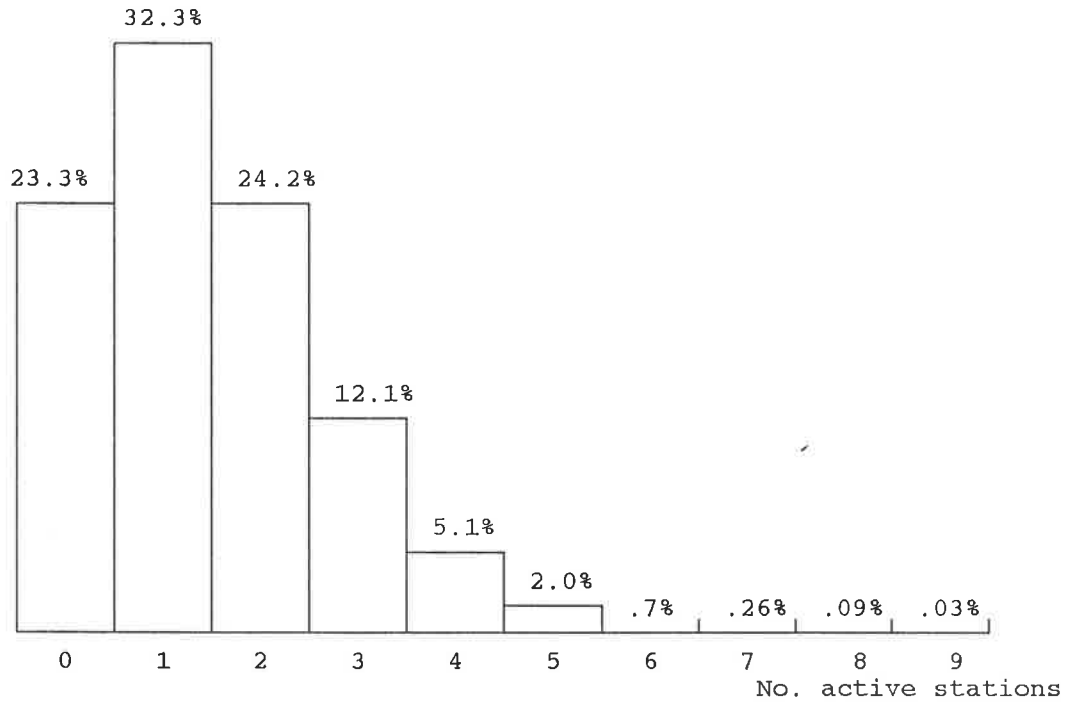


Fig. 7: Transition rate diagram (a) and histogram of the number of active workstations (b) as an example for a (peak hour) situation with 200 workstation users, which evoke 3,000 request-response cycles (125 sequences of 24 request-response cycles) per station per hour.

- (If 1 workstation is active then 24 request-response cycles take 200 msec;
- if 2 workstations are active then 24 request-response cycles take 218 msec;
- if 3 workstations are active then 24 request-response cycles take 220 msec;
- if 4 workstations are active then 24 request-response cycles take 250 msec;
- if 5 workstation are active then 24 request-response cycles take 296 msec;
- if 6 workstations are active then 24 request-response cycles take 324 msec;
- if N workstations are active then 24 request-response cycles take $.054 * N$ msec)

6. CONCLUSION

For CSMA the normalized propagation delay is a fundamental parameter for the throughput performance. For slotted nonpersistent CSMA, with a data rate in the range of 1 - 4 Mbps, a carrier detect time of 15 μ sec and 512 octet packets, a throughput in the range 78 % (4 Mbps) - 90 % (1 Mbps) is possible.

The performance behavior of slotted CSMA with p-persistent or collision avoidance scheme can be improved by respectively an adaptation of the p-persistence factor and an adaptation of the backoff slot range. The improved behavior gives a better throughput for a variable load and more robustness at heavy load conditions.

In a practical request-response situation the throughput performance will differ significantly from well known theoretical figures because of a request-response interaction between server and workstations and a buffered load. In practical throughput scores the overhead bits at various levels is discarded.

Enhanced CSMA systems can provide a high throughput performance, while a relatively simple and inexpensive control provisions are applied. It is a question if other non-CSMA-like systems are more efficient under real life LAN circumstances.

Although enhanced CSMA systems can provide a high throughput with error recovery above the MAC layer, it is an issue to provide a highly reliable service on the MAC layer. There exists a question relating to the compatibility requirements specified in the PAR.

REFERENCES

- [1] IEEE P802.11/91-17, *Local Area Network Market Analysis and Forecast*, K. Biba; February 1991
- [2] IEEE P802.11/91-35, *802.11 MAC Layer - Some Proposed Characteristics*, P. Altmaier and J. Mathis; March 1991
- [3] IEEE P802.11/91-44, *A Short Tutorial on CSMA*, P. Altmaier; May 1991
- [4] IEEE P802.11/91-49, *Comments on CSMA*, D. Buchholz and L. Hamilton; May 1991
- [5] *Performance Analysis of Local Computer Networks*, J.L. Hammond and P.J.P. O'Reilly, Section 9.4 - 9.5: Protocols that Use Carrier Sensing - Protocols that Use Carrier Sensing with Collision Detection, pg. 320 - 339; Addison Wesley, 1986
- [6] *Telecommunication Networks: Protocols, Modeling and Analysis*, M. Schwartz, Section 8 - 9: Polling and Random Access in Data Networks, Local Area Networks, pg. 403 - 480; Addison Wesley, 1987
- [7] *Multiple Access Protocols*, R. Rom and M. Sidi, Chapter 4: Carrier Sensing Protocols, pg. 79 - 103; Springer-Verlag 1990