CSMA-Radio Mobile MAC proposal

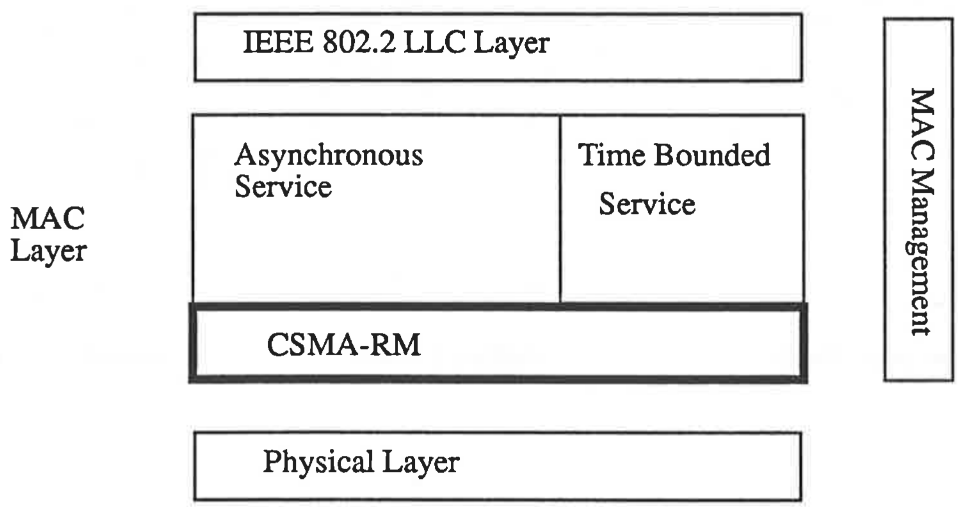
Authors: Philippe Jacquet Paul Muhlethaler
Address: INRIA, BP 105, 78153 Le Chesnay Cedex, FRANCE

1. Introduction

In the present paper we discuss the feasibility and advantages of CSMA schemes used in wireless local area networks.

CSMA schemes are decentralized, simple, robust and flexible which are very attractive features for wireless networks, since one expects physical perturbations (fading and noise) as well as logical perturbations (insertion or withdrawing of mobile STAs). Unfortunately pure CSMA schemes show some drawbacks, notably a poor reliability and a tendency to congestion and instability. Collision/Error Detection is known to cope with these problems. Error Detection provides perfect reliability. Appropriately used (Collision Resolution Protocol), Collision Detection stabilizes CSMA and also provides very good real time behavior to the protocol (promptness, access delays, reaction to bursts of traffic, etc.). These are interesting features if one wants to implement time bounded services.

It is incorrectly assumed that Collision Detection is not possible on radio networks. It is clear that collision detection cannot be performed like on wired networks, since there is a too large difference between transmission power and reception signal, so that significant comparison is impossible. But collision detection can be performed via alternative and efficient ways. In this note we present one of them, in fact a class of solutions that we call CSMA-Radio Mobile (RM).



IEEE 802.11 Protocol Architecture with CSMA-RM

In the following we introduce CSMA/RM basic medium access schemes. Our aim is not to present a complete and detailed solution to IEEE 802.11 for which there are a lot of interesting works already done. But we would like to show that CSMA/RM is flexible enough to accept various schemes of coordination functions and distribution functions. So that we will only outline general classes of acceptable solutions. Time bounded service is a challenging problem but we also restrain ourselves to the description of two general classes of solutions and how they can take advantage of CSMA-RM. CSMA-RM schemes have also been presented as a possible solution of Hiperlan-ETSI (referenced as LAURA input).

- CSMA is attractive to Radio LANs
- Collision Detection is possible in radio LANs
- CSMA accepts many coordination schemes

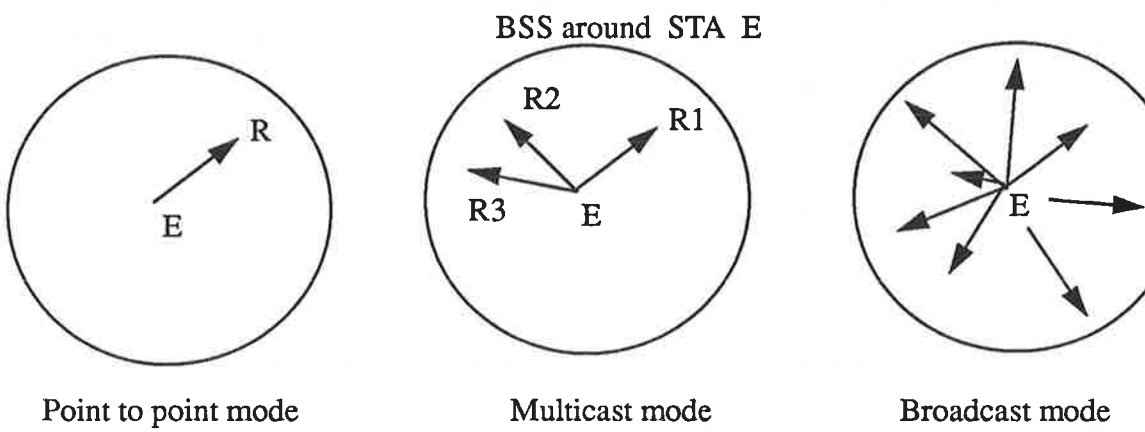
2. Radio LANs specifications

2.1 Basic MAC requirements

Usual MAC requirement is to insure connectivity in the BSS involving many STAs and one AP. MAC must achieve successful transmission of packet (MPDU) under any of the following conditions.

- in point to point mode: the destination node is in the cell of the source node;
- in multicast mode: the packet has several destinations in BSS;
- in broadcast mode: the packet must reach at least all nodes in its BSS.

Multi point modes are favored by the natural broadcast nature of radio transmission. Note that most of existing MAC protocols determine transmission mode via specific rules in the destination address field of the packet.



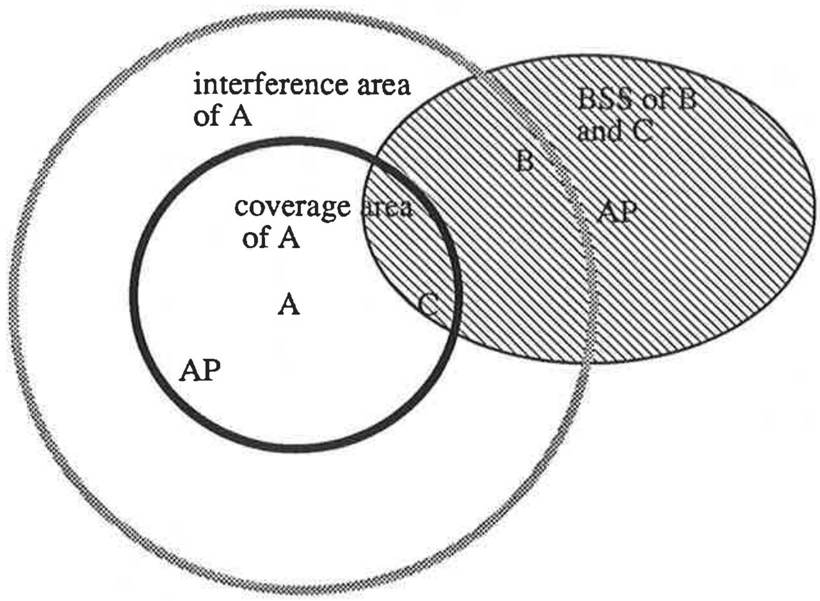
The three packet addressing modes

2.2 Basic Physical properties

The physical layer can perform three basic operations: detecting activity, receiving packet and transmitting packet. Receiving a packet may be different than detecting activity. In fact receiving a packet means reception without error (after Forward Error Correction). Error can occur because of collision and/or simply because the ratio of signal over noise is too weak. In fact, the capture effect generally transforms collisions into captures with error.

Since physical layer manages FEC, it should be able to inform MAC layer about FEC failures.
Physical layer should also be able to inform MAC layer if a packet is received out of range (for example if power reception is below certain threshold). Doing so will help MAC to make distinction between errors due to collision and interference and errors due to signal attenuation. The range of a station determines what we call the coverage area of the station. Inside its coverage area the station can perform reliable one hop communication. All stations in a BSS should be in the coverage area of their AP.

If the source is too far away and beyond the coverage area, the packet might be only detected, not received. This leads to define for any given source an interference area larger than the node coverage area. Since error-free receptions and simple detection may differ of 10 dB, interference area might be of radius twice the radius of the BSS.



Interference area versus BSS

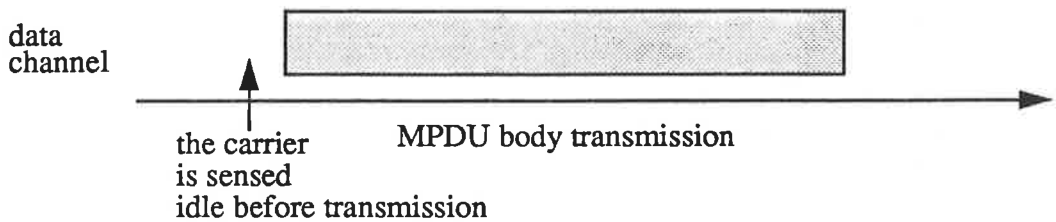
To cope with hidden nodes, physical layer can perform an enhanced Carrier Sense (for example, distributed busy tone).

- **In BSS, MPDU are point to point or multipoint.**
- **Physical layer informs MAC about activity detection and FEC failure.**
- **Interference area is larger than reception area.**

3. Basic CSMA schemes:

3.1 P-persistent CSMA

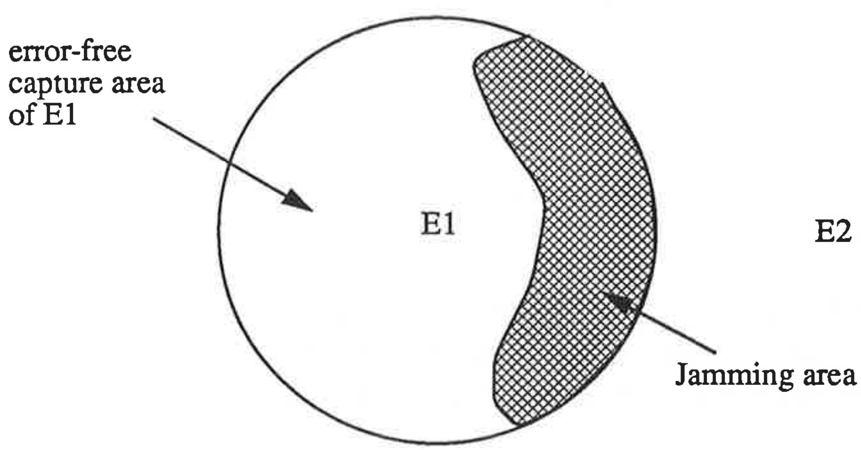
CSMA algorithms are well known: STAs sense the carrier before any attempt of transmission. A transmission is only possible when the carrier is sensed idle. Carrier sense is frequently associated to P-persistence which reduces the occurrence of collisions likely to happen after each busy period of the carrier. When a MPDU is ready to be transmitted during a busy period of the carrier, either, with probability 1-P, the packet is back-logged again for a random period, or with probability P, it is transmitted as soon the carrier becomes idle. There are other ways to achieve persistence, for example to decrement backlog periods only when carrier is idle.



CSMA MPDU transmission

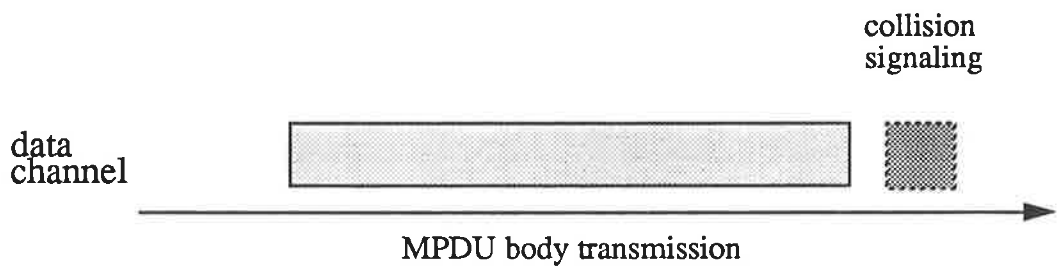
3.2 CSMA-CD (Collision/Error Detection)

Let us consider the case where two STAs E1, E2, simultaneously transmit. We suppose that E2 is sufficiently close or belongs to the coverage area of STA E1. Thus it is very likely that the coverage area of E1 will be subdivided into two subsets: the subset of the nodes which correctly capture the emission from E1 and the subset of the nodes which don't correctly capture the emission from E1. By correct capture, we mean FEC without failure. Among the latter are nodes which capture emission from E2 and those which capture transmission from E1 but with errors. This leads to defining a jamming area or area of reception with errors. It is clear that if the packet transmitted by E1 is in broadcast mode or if the destination belongs to the jamming area, then the transmission is unsuccessful.



Collision within a coverage area

In this case collision/error detection device is necessary for reliability and overall performance. Therefore one can implement CSMA/CD (CD for collision detection). Collision or Error detection can be achieved by a special signal activated by the receiver STAs. If a receiver within range of the source detects an invalidity in MPDU decoding (thus belongs to the jamming area), then it activates the collision signal after the end of reception. The source echoes the signal of collision in order to inform all node within range that a retry will take place. CSMA/CD are known to give better performance and reliability than pure CSMA. CSMA/CD allows retry procedures, either back-off algorithms or tree algorithms.



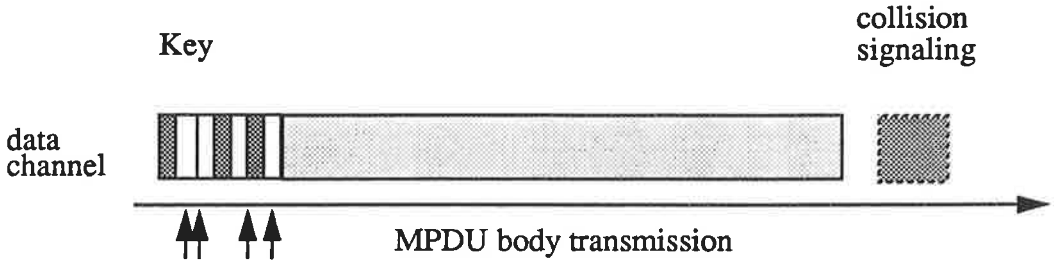
CSMA/CD MPDU transmission

The scheme is available in multi point transmission mode since collision signals can superpose without damage.

3.3 Collision reduction and avoidance CSMA

An ultimate improvement of CSMA and CSMA/CD, is Collision Reduction-Avoidance schemes which makes collision detection general and efficient. There are several ways to achieve collision avoidance. The most general one, consists in transmitting a preamble of bursty transmissions before every packet. Each burst lasts more than a propagation delay plus carrier sense time. The bursts are synchronized to a coded on-off time sequence, called transmission key which belongs to the STA. During the silences between the bursts, the transmitter listens for other nodes' activity. For example, if no other nodes' activity is sensed during this period the source sends its packet. Otherwise the source quits key transmission and re-schedules a new attempt as soon the carrier becomes idle (at least waiting for other node to complete the transmission of their keys). There is no synchronization needed between any pair of transmitters, excepted the one given by carrier sense.

If the first bit of each transmission key is always set at "1", we call this bit the synchronization bit, after which by carrier sense all other stations in interference area will be aware that a key is in current transmission. If we suppose that the transmission keys of every pair of nodes within range of each other, are always different and that there is no hidden nodes, then we have an almost perfect access scheme called "Collide and win".



If the transmitter detects an alien signal
during a listening period it resigns transmission

1 0 0 1 0 1 0 Transmission key

Collision reduction-avoidance MPDU transmission

Note the presence of the eventually activated collision signaling. Collision signaling is still possible in the case of spurious collision, like with hidden nodes, for example. Such collisions are resolved via classic resolution (for example usual back-off or tree algorithms). In this perspective the key can also be used by another station to find invalidity in the packet decoding.

In passing, we note that CSMA-RM keeps the global properties of CSMA schemes: simplicity, flexibility, etc.

CSMA-RM can be tuned in
- pure CSMA (with or without P-persistence);
- CSMA-CD via collision tail signals;
- CSMA-(CD/CA) with keys of transmission.

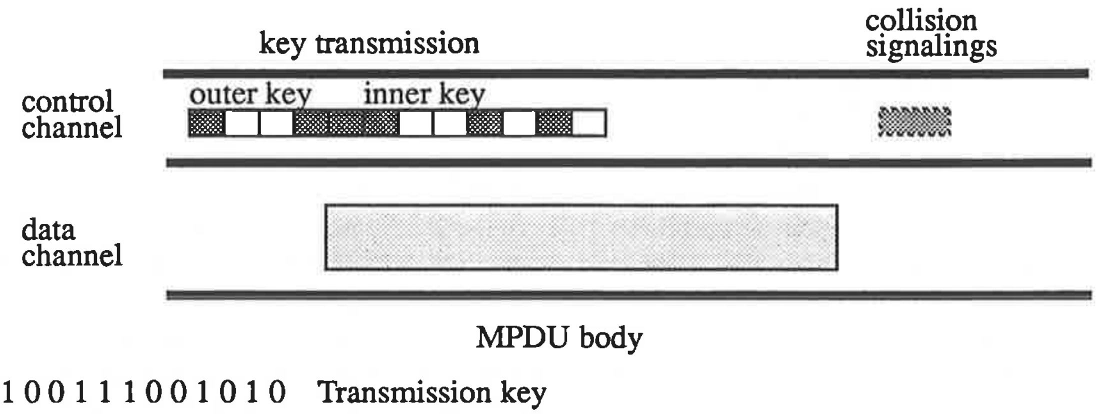
4. Comment on CSMA-RM schemes

4.1 Physical layer independence

Note that if the keys are all zeros (all bits set at zero) we simply have CSMA/CD. If the collision/error signal is never activated, we have pure CSMA. The scheme is decentralized and does not care about overlapping BSS's that share the same physical medium. The scheme is easily tunable (size of the key, etc.), depending on physical layer constraints. For example, the scheme is easily adaptable on physical medium where each data channel is divided into several parallel sub-channels, as illustrated in the figure below with two sub-channels.

More generally, if the medium is shared between several channels which allow parallel packet transmissions, then the scheme can be duplicated on every channel. Thus multi-channel management is simplified.

The strategy based on transmission of keys takes advantage of the relatively short propagation delays that one can expect in BSS. Note that such key strategy are also used in wired very local area networks (for example, VAN in car engineering: ISO TC22/SC3/WG1).



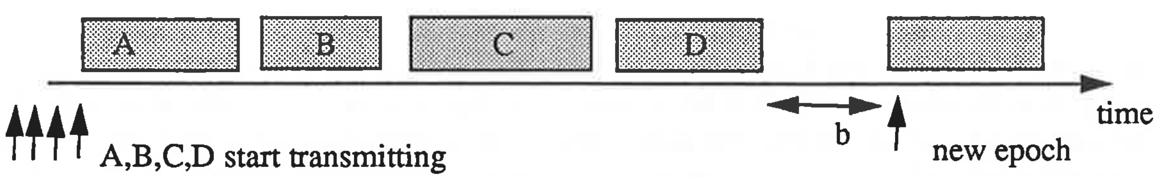
Alternative transmisson on control and data channels

4.2 CSMA general behaviour.

For simplicity of presentation we will investigate the mode "collide and win" presented above, because its behavior is easy to grasp and illustrates many properties of these schemes. This can be extended to a large number of collision resolution schemes or, to a lesser extend, to pure CSMA schemes.

For fairness reasons, each STA with a new MPDU makes its first attempt after observing a persistent idle state of the channel during a duration of at least b. Doing so, one eliminates the risk of possible starvation.

As illustration, let us consider four transmitters A, B, C and D, wanting to access the medium at the same moment. We suppose that the transmission keys are in decreasing order if we consider them as integers in a binary representation (key(A)>key(B)>key(C)>key(D)). If the four STAs are within the range of each other, the resolution of the conflict is trivial: the node with the strongest key will win the first access, therefore A. After the transmission of A, all the remaining transmitters again compete, B wins, then C and finally D. After the end of this epoch, a new one can start after



Collide and Win example

time b. Note that MPDUs can be of different length. In the figure above, the keys are included in MPDU boxes to simplify the presentation.

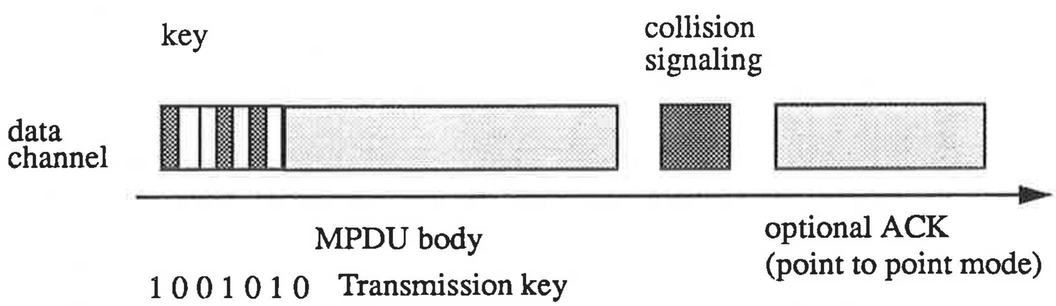
In the case all STAs being within range of each other, the scheme resembles a distributed M/G/1 perfect scheduler in terms of behavior and performance. If some STAs are not within range of some others, then simultaneous transmissions naturally take advantage of spatial reuse. Access delays are very good and, in distribution, are not influenced by the number of connected nodes (like with perfect scheduler). There is no need of network synchronization and the algorithm is limited sensing (stations can eventually switch off when idle in order to save battery power).

4.3 Reliability and extension of CSMA-RM schemes

CSMA-RM scheme is basically broadcast mode oriented in the sense that all MPDUs are treated as if they were all one-hop broadcast MPDUs. Therefore the same reliability is provided to packet transmission regardless of their transmission mode (mono-point or multipoint)

It is clear that if the broadcast problem is solved, then point to point is also solved. But it may happen that a point to point MPDU transmission be considered in collision within BSS, although it is correctly received by its destination. For example, when the targeted destination is sufficiently close to the source, while other remote receivers in the BSS detect and signal a collision. But if the destination immediately transmits a specific acknowledging MPDU and if the source receives it correctly, then the latter will automatically quit the collision resolution procedure. Note that it is not possible to use such feature for multicast transmissions: ACKs will automatically collide.

The transmission of this ACK MPDU will be done without collision management (key set at zero and no care to collision signal). The ACK may improve the performance of MAC protocol in terms of access delays when bit error rate is high, but it implies an overhead. If implemented, some specific mechanisms should be added in MAC sublayer.



Optional ACK Transmisson

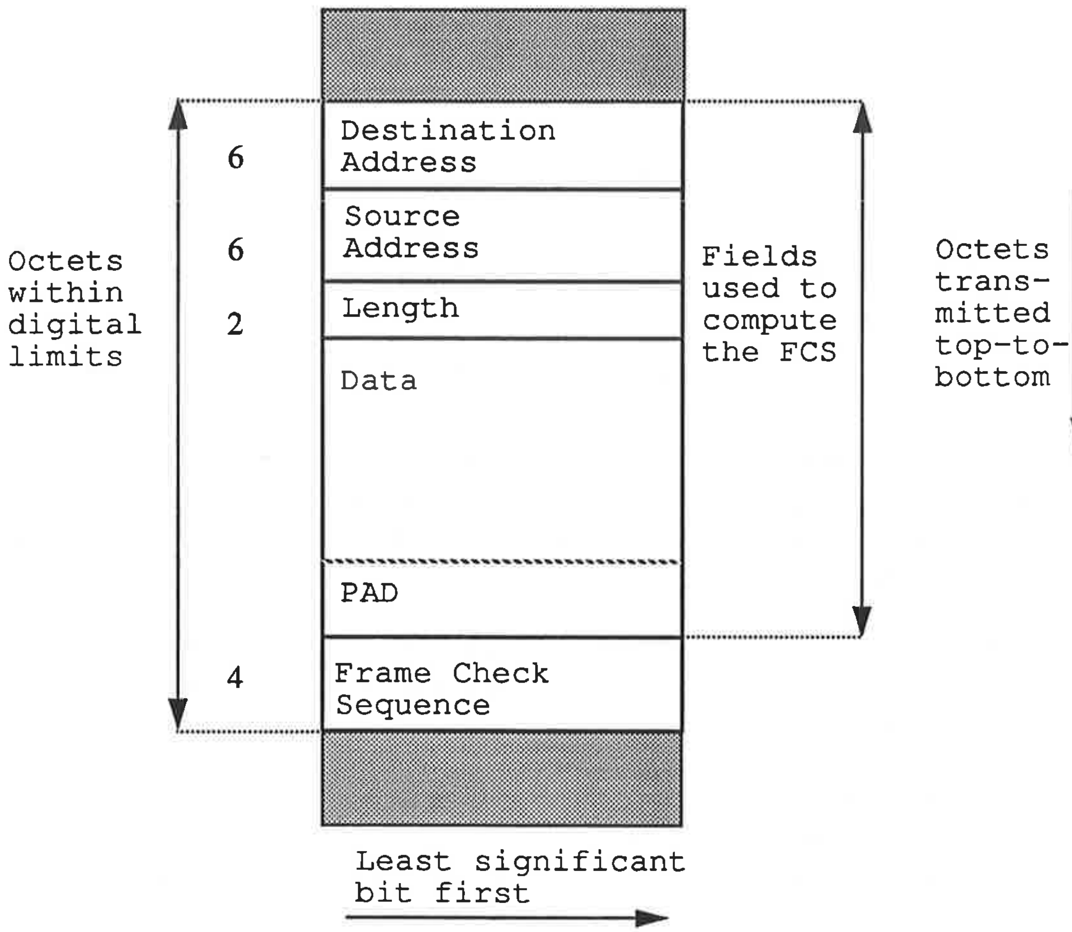
In summary, the reliability of data transfers when provided at low level is a great advantage in term of performance. Packet recovery at upper layer (ARQ) relies on timers whose actual minimal delay is within one second: if MAC looses in average one packet per one thousand, the recovery procedure may divide by two the actual peer to peer throughput. It is also important to provide reliability for multipoint communications, since network management is greatly dependable on such mode of transmission in most MAC proposals (see section about time bounded).

- CSMA-RM can be adapted to most wireless media
- parameters are tunable to optimize performance
- optimal reliability for all modes (mono-point and multipoint).

5. MSDU format, coordination function and distribution system

5.1 MSDU format and addressing

CSMA-RM does not need any special MPDU format and does not impose any particular coordination function and distribution system. Below is the description of basic ISO 802.3 format.



MAC PDU Format

If auxiliary MPDUs are used for medium access, then a special field indicating the type of MPDU should be added. If one type of such MPDUs is envisaged like the immediate ACK above described, then no new field may be necessary. In this case the immediate ACK can be a MPDU with length field equal to zero.

Addressing is IEEE over 48 bits which offers bridging facilities and enough flexibility for the alternative choice between universal addresses and local addresses. Note that with local address management, ISO 8802 gives the possibility to deal with network ID versus node ID.

48 bit address format

I/G	U/L	46 bits address

```
I/G=0 Individual address
I/G=1 Group address
U/L=0 Globally administered address
U/L=1 Locally administered address
```

Address Field Format

5.2 Coordination function and distribution system

CSMA/RM is decentralized thus it can support any coordination and distribution system. Connection in the BSS can operate either direct peer to peer or via a store and forward on AP. This can be done on decentralized mode or centralized mode. On decentralized mode each STA must know the subset of its BSS which is directly within range and the subset which is only attainable via the AP. On centralized mode only the AP has the knowledge and automatically stores and forward if necessary. The management of the coordination function must be done via dedicated broadcast MPDUs.

The distribution system is independent of CSMA/RM basic schemes and it resembles to network management. Numerous options exist. With ad-hoc networks one can implement intra-forwarding facilities where every node is eligible to be a relay between two STAs mutually out of range. The relay management can be done either via source routing or hop by hop. When a backbone network with APs exists, the relay management should be reduced to a hierarchical bit to decide whether or not a MPDU has to be forwarded on backbone. Schemes similar to learning routers can be implemented in the AP architecture.

Insertion of a new STA in a BSS must be done via spontaneous broadcast packets either from STA or from AP within range. If several APs are within range, the STA should choose one (for example the one received with the most power). There also exists the possibility for a STA applying to several BSS as for Hiperlan (ETSI RES10).

Note that all the features described above have no impact on the MPDU format. All the informations needed for coordination and distribution should be encapsulated into the general DATA field described in format. The use of encapsulation is encouraged in order to keep safe 8802.3 format and to keep the architecture flexible and easy to implement.

- ISO 8802 for MSPDU format and addressing.
- CSMA-RM is flexible in CF and DS
- Management can be handled via encapsulation in DATA field

6. Time bounded services

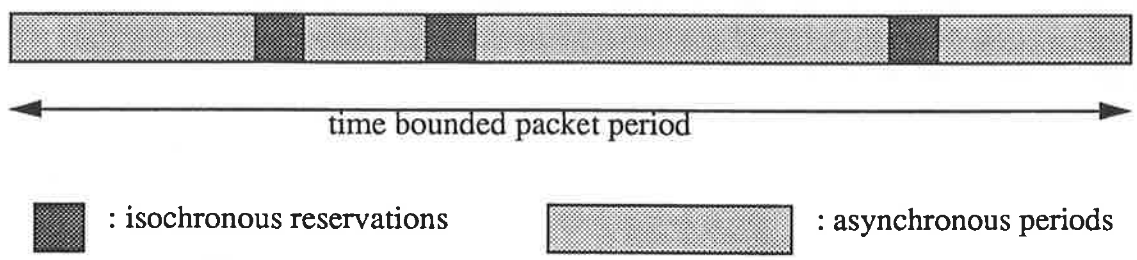
Our decentralized scheme is general enough to support various time bounded managements.

Time bounded services consist in isochronous packet delivery (voice, multi-media). There are two main approaches:

- 1. periodic bandwidth reservation between asynchronous and isochronous traffics;
- 2. asynchronous time bounded traffic rendered isochronous via priority and buffering.

6.1 Reservation schemes

During the normal utilization of the network, every node has the knowledge of the reservations done within one or two hops distance around. Thus asynchronous and isochronous traffics will never collide. This knowledge is generally stored under the form of a local reservation vector containing the initialization times of every connection.



Periodic pattern of reserved bandwidth

The reservation vector is a list of coefficients of the form (connection ID, x), where x is the initialization time (modulo the time bounded period). In case of different time bounded periods, the coefficients should also indicate the period T: (connection ID, x, T).

Reservation is simplified by the usage of broadcast transmissions. Calls for connection and periodic updates has to be carried out by broadcast packets so that all other users within range be aware of the status of the reservation.

The call set up consists of finding a good initialization time and to properly insert it in the reservation vector of each STA within range of source STA and recipient STAs. Multiple recipients is supported per connection. The source sets up a connection ID and broadcast it to the recipients in a broadcast packet containing recipient ID's. So far all STA within range will be informed of the existence of a new connection. The determination of the initialization time will be done by comparison of reservation vectors of source and recipients. This comparison is done either on a decentralized mode (vector exchange), or on a centralized mode (for example on AP).

Source and recipients update or quit, their reservations via broadcast packets referring to coefficient (connection ID,x). Thus, the reservation vector of the other STAs within range are updated. By STAs within range we mean all STAs, including those which eventually don't belong to the same BSS. The use of broadcast packets allows to manage several connection via single packets.

Calls and updates are asynchronous packet exchange and are done during asynchronous periods, so that there is no confusion between time bounded service management and isochronous packets (thus less information carried by isochronous packets).

The problem of a common reference time when exchanging initialization times can be solved via the usage of timing pairs (two consecutive packets: the second one containing the actual transmission time of the first one).

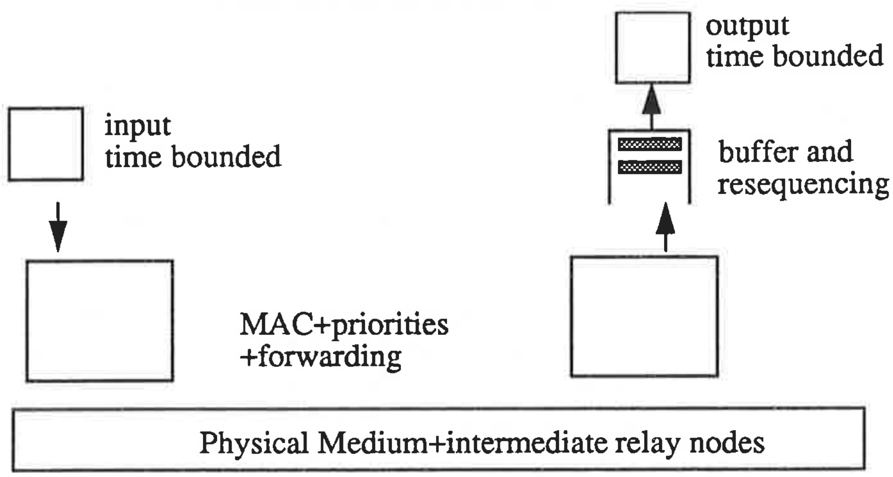
The advantages of reservation schemes is that isochronous delivery is perfect and without delay. The disadvantages: the use and the management of reservation vectors may be heavy, costly overhead if update periods too short or if number of recipients is too high (audio conference)

6.2 Priority and buffer schemes

No special bandwidth allocation but asynchronous packets are transformed into isochronous packets via buffering and sequencing (like in ATM).

The main constraint induced by time bounded services is the imperative isochronous packet delivery. But there is flexibility in packet loss and delay (interactive voice can support 200 ms overall constant shift).

With the buffer scheme asynchronous packets are stored in specific buffers. The purpose of such buffers is to resequence and readjust timings between packets and to recover from eventual delay discrepancy (MIC traffic, H-221). If MAC performance are acceptable for time bounded packets, delay in buffer can be small.

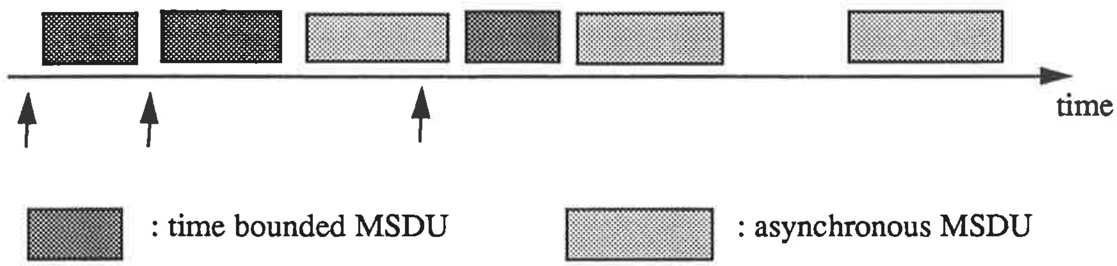


Time bounded service delivery

To make MAC performance acceptable one can make usage of priorities so that time bounded packet have privileged access on the medium. Below is an example using CSMA Collide and Win.

We define two priority levels: a high level for time bounded, and a low level for asynchronous traffic. The priority is indicated in the transmission key: the first significant bit (after the synchronization bit) will be set to "1" for high priority packet, otherwise it will be set at "0" for low priority packet. Therefore any node holding a packet of high priority will transmit it with a key in any case larger than any other packet of lower priority.

Time bounded packet will be transmitted without any persistence slot. With such enhancement the traffics of lower priorities are literally transparent to the traffic of higher priorities. The only delay that a time bounded packet may incur is eventually the length of a lower priority packet.



Collide and Win with priorities

Connection set up and synchronization will be done via lower priority MSDUs exchange so that appropriate initialization time be automatically found between existing time bounded traffics.

In the above example, the vertical arrows indicate initialization times of time bounded MSDUs.

Equivalent schemes can be proposed with other collision resolution protocols. For example with tree algorithm the splitting sequence can replace the transmission key and contain several "1" in a row for time bounded packets. To a lesser extend one can also tune P-persistence according to class of traffics with pure CSMA. For example, time bounded packets may have 1-persistence.

Because of buffer resequencing, time bounded packets may be simply destroyed if access delay exceeds a given threshold. The high priority level can only be used to carry alarm packets or equivalent emergency traffic.

The advantage of the priority-buffer scheme is that it does not generate overhead nor extra management (it has no impact on asynchronous traffic of other STA), it is insensitive to the number of recipients per connection, is easy to handle. The disadvantage is that it introduces a global shift delay in packet delivery and packet losses remain possible due to overdelay if no priority is implemented.

Time bounded services can be carried out either via:
- **management of centralized or decentralized reservation vectors**
- **or via decentralized priorities and MAC buffers.**

7. Conclusion

Collision detection and resolution are very attractive features which provide promptness (minimize medium access delays) and stability to the network, even when used in distressful conditions. Another consequence of the scheme is an exceptional reliability provided to data transfer and this reliability extends without loss to broadcast and multicast traffics. This last property makes the scheme particularily adapted to time bounded implementations.

Warning

Some of the items presented in this note are under patent. In particular the collision detection and management schemes are patented items.

Bibliography:

- INRIA 93a, "Busy Tone Control Signal," working paper ETSI RES10, TTG;
- INRIA 93b, "Collision Detection," w.p. ETSI RES10, TTG;
- INRIA 93c, "Coexistence, Interoperability and Control," w.p.ETSI RES10, TTG;
- INRIA 93d, "Prioritisation in Hiperlan, solutions at MAC sub-layer," w.p.ETSI RES10, SAG;
- ISO 8802.3 Specifications
- Massey, "Collision Resolution Algorithms and Random-Access Communications," in Multi-User Communications Systems, G. Longo Editor, CISM Courses and Lectures No 255, Springer Verlag, Wien-New York, 1981, pp. 73-137.
- Bertsekas, Gallager, "Data Networks," Prentice Hall, Inc, 1987