QoS in IEEE 802.11

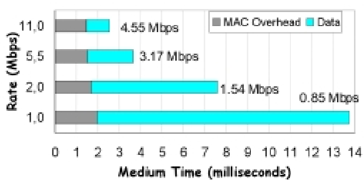
Issues

Some issues are important for quality of service: the first one mentioned is the difference of performances expired by nodes based on their position in the network. Indeed, considering as a topology a chain, nodes placed at extremities see different channel conditions with respect to stations placed internally. The same phenomenon happens considering a topology in which nodes are located on a grid: in this situation, stations in the center that are able to detect transmission of other two nodes, see the channel always busy while those other nodes can transmit every time.

A second issue is the frame overhead. Each data packet and ACKs are composed by preambles and data: if data actually can be sent at an higher rate, preambles are always sent at the basic rate to be heard by everyone. Moreover, preambles are quite large in terms of size:

- 24 Byte for data packets;
- 14 Byte for ACKs.

Due to this fact, the overhead is more relevant if the data contained in packets is sent to high rates:



The third most important issue is fairness. It can be provided:

- *per-packet* if each node can transmit the same number of packets in a given period (usually long) of time;
- *per-time* (also called temporal fairness) if each node can access the channel for the same amount of time.

In general MACs protocol provide either per-packet and temporal fairness. Consider two stations n_1 and n_2 that transmit 10 packets of length L in a given period T; with per-packet fairness throughputs will be:

$$T_{n_1} = \frac{10 \cdot L}{T} \qquad\qquad T_{n_2} = \frac{10 \cdot L}{T}$$

If temporal fairness is provided, those stations have two different transmission time for each packet, $T_{x_{n_1}}$ and $T_{x_{n_2}}$; of course:

$$\sum T_{x_{n_1}} = \sum T_{x_{n_2}}$$

If the transmission rate is R, throughputs will be:

$$T_{n_1} = \frac{T_{x_{n_1}} \cdot R}{T} \qquad\qquad T_{n_2} = \frac{T_{x_{n_2}} \cdot R}{T}$$

It is very important because, different models of fairness allows to achieve a better throughput in particular conditions:

. per-packet fairness allows to reach the same throughput for all stations if packets have the same size and they are transmitted at different bit rate;

. temporal fairness allows to reach the same throughput for all stations if packets have different size, but they are transmitted at the same bit rate;

. if packets have different size and are transmitted at different bit rate it is difficult to predict which scheme allows to achieve a better performance.

In general per-packet and temporal fairness are equal with packets of the same size transmitted at the same rate.

The anomaly effect

CSMA-CA provides per-packet fairness and in a multi-rate environment it happens the so called *anomaly effect*: it is an inefficiency in which all nodes achieve the same throughput as all of them are transmitting at the lowest rate. The anomaly effect is more relevant if fastest and slowest sources are present in the same number in the network. Possible solutions are essentially two:

. force the protocol to provide temporal fairness;

. use relay nodes.

To have temporal fairness OAR (Opportunistic Auto Rate) is used: it looks at performances and not to all user's requirements. The CTS is changed a bit, adding a field that contains the SNR: the source that receive that CTS, after have sent an RTS, can change the data transmission rate accordingly to the CTS. In this case, if channel conditions are good, the transmission can happen with an higher transmission rate, but if they change the rate is quite fast adapted to new conditions.

The use of relay nodes implies multi-hop transmissions: it can help to deep with slow stations since relay nodes are connected to other stations with fast links and the interference range is reduced, but this solution has some drawbacks. Indeed, the delivery delay increases and the overhead is very effective because headers and preambles are always sent at the basic rate for each hop.

QoS

Some mechanism introduced are:

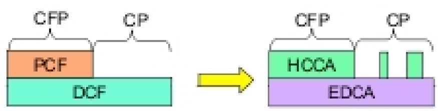
- Deng;
- Black Burst;
- adaptation of 802.11 parameters;
- 802.11e.

Deng

It is an extension of IEEE 802.11 that allows to provide QoS: backoff intervals and and IFSs are computed accordingly to priority levels (4 classes available). Higher priority is associated to PIFS while DIFS has low priority: if the load is very huge, the 55% of the bandwidth is dedicated to higher priority traffic, but in the other case both high and low priority traffic can obtain the bandwidth needed.

802.11e

The standard approved in 2005 tries to enhance the standard 802.11 to improve and manage QoS. It provides 8 different traffic categories and has improved DCF and PCF by introducing a new access scheme: HCF (Hybrid Coordination Function), composed by EDCA (Enhanced Distributed Channel Access) and HCCA (HCF Controlled Channell Access). As shown



in the picture, even during the contention period the HCCA can operate while PCF can not.

Main features introduced are:

- FEC techniques (while 802.11 implement just stop & wait);
- direct communication/side traffic;
- WARP (wireless address resolution protocol);
- AP mobility (in principle any user can act as AP).

Definitions

- QBSS: BSS with QoS support;
- QSTA: WSTA with QoS support;
- Hybrid Coordinator: similar to PC (Point coordinator), it is the central controller and it can be placed either at on of the QSTAs or at AP (QAP);
- TXOP (Transmission Opportunity): a new feature, very important, that describe the time in which a QSTA can transmit; with TXOP, nodes can send more than one packet if the period is larger enough (the limit, or maximum duration, is called TXOP_limit) and it is implemented both in CP (contention period) and CPF (contention period free).

EDCA EDCA is backward compatible with basic DCF and it introduces the following mechanism:

- the TXOP is allocated by contention;
- there are 4 Access Categories: at MAC layer 4 queues are implemented as virtual stations that try to access the channel; each one uses different parameters (enumerate below) and different CSMA/CA istances;
- different IFSs (also AIFS, Arbitrary IFS), CW_{min}, CW_{max} and TXOP_limit to characterize traffic.

To priority traffic is assigned values of CW_{min} low in this way they have he opportunity to access the channel sooner; the value of CW_{max} is not too much relevant.

The 4 access categories, implemented by queues, act as independent stations, therefore they sense the channell independently and perform backoff procedures independently too. If, in a given istant, the backoff counter reaches 0 in two different queues, the TXOP is granted to the queue with the highest priority. When a collision occurs, it is treated as an external collision and queues perform backoff in the usual way. Once the station is able to access the channel, it can transmit until the TXOP_limit; moreover, acks can be sent either frame by frame or just after a burst of packets: this allows to reduce the overhead. To avoid collisions, in case of burst acks, SIFS is used between packets.

Pros	Cons
reduced overhead	increase delay jitter
increase throughput	difficult choice of TXOP_limit
fairness among same priority queues	(not longer than the largest frame)

HCCA HCCA is based on polling of QSTAs by the HC (Hybrid Controller), it is backward compatible with DCF/PCF and the access of the channel is deterministic. HCCA can operate both in CP and CPF and the main features are:

. the HC allocates TXOPs by using opportune frames (CF-Poll frame) if the channel is free for PIFS;

. the time in which TXOPs are polled in CP is called CAP (Controlled Access Period);

. there are 8 traffic categories.

Now the HC has the knowledge about how much the station will transmit: in PCF was not possible. Moreover the coordinator is also informed on the state of the queues of a given station and can poll nodes in the right way performing the best choice, but it requires a non static polling list.

Behaviour of HC

. HC may allocate TXOPs to itself to transmit MSDUs whenever it wants, however only after having sensed the channel idle for PIFS;

. in CP, the HC can send the CF-Poll frame after a PIFS idle period, thus starting a CAP;

. In CFP, only the HC can grant TXOPs to QSTAs by sending the CF-Poll frame;

. The CFP ends after the time announced by HC in the beacon frame or by the CF-End frame from HC.

Behaviour of QSTA

. In CP QSTAs can gain a TXOP thanks to a CF-Poll frame issued by HC during CAPs, otherwise they can use EDCA;

. In CFP, QSTAs do not attempt accessing the channel on their own but wait for a CF-Poll frame from the HC.

The HC indicated the TXOP duration in a specific field of the CF-Poll frame called QoS-control field: in this way stations can be kept silent by NAV when they detect such a frame.

Signalling and Polling The signalling is provided essentially in two ways:

. by means of the *connectionless queue state indicator*;

. by means of the TSPEC (Traffic Specification): it allows to provide some CAC (Control Admission Control).

Polling is the crucial step in HCCA: performances depend essentially on how the polling is implemented although the way in which QSTAs are polled is not characterized. In general QSTAs can send periodically updated to HC to inform it on the queue size and their TXOP desired; moreover, to initiate a new communication, they can send a frame called ADDTS (Add Traffic Stream).