The Industrial Electronics Handbook
Industrial Communication Systems
Bogdan M. Wilamowski, J. David Irwin

Media Access Methods

Publication details
Herbert Haas, Manfred Lindner
Published online on: 28 Feb 2011

Accessed on: 12 Dec 2018

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
3.1 Introduction ........................................................................................................... 3-1
3.2 Full-Duplex Media Access.................................................................................... 3-2
3.3 Synchronous Access Arbitration Concepts ....................................................... 3-2
   Static Timeslot Mechanisms • Dynamic Timeslot Mechanisms
3.4 Statistic Access Arbitration Concepts ................................................................. 3-3
   Aloha Mechanisms • Pure Aloha • Slotted Aloha
3.5 Carrier Sense Mechanisms with Exponential Backoff ........................................ 3-4
   Using Collision Detection • Using Collision Avoidance •
   Token Mechanisms • Polling Mechanisms
3.6 Other Media Access Issues.................................................................................. 3-6
   Access Fairness • Access Priorities • Quality of Service •
   Hidden Stations
References.................................................................................................................... 3-7

3.1 Introduction

The technical details of media access are generally implemented in OSI layer 1 and 2. Usually the nature of OSI layer 1 determines the operation conditions and feasible access mechanisms that can be implemented in layer 2.

Media access on full-duplex media is often a simple task because each station can send data independently of each other. On the other hand, media access on half-duplex media is typically nontrivial and requires solutions for the following issues:

- Access arbitration
- Collision treatment
- Access fairness
- Access priorities for selected stations or messages
- Quality of service (QoS) requirements (queuing)
- Real-time requirements

Furthermore, most interesting for each access mechanism is the question about media utilization. Typically, there is a natural trade-off between real-time support and maximum media utilization, which might be solved only by increasing the complexity of the system (for example, by introducing a base station which is typically used in wireless data/voice communications).

The subsequent sections provide a detailed problem description, especially considering half-duplex media access in more detail, and explain possible solutions. Historical and current examples are also given.
3.2 Full-Duplex Media Access

If the communication medium supports full-duplex access, most communication problems are inherently solved. Full duplex can be achieved using the following methods:

- Separated wire pairs within cables, one for transmitter (TX), the other for receiver (RX). This is the most important solution in practical implementations (e.g., used in 100BaseT).
- Separated TX and RX fibers in optical communication environments.
- Separated frequencies, one frequency for TX, the other for RX.
- Separated wavelengths, one for TX, one for RX, within fibers. This method is typically more expensive than using separate fibers.
- Echo cancellation—both stations send and receive at the same time over the same wires. This scheme has been implemented, for example, on Gigabit Ethernet on twisted pair media (1000BaseT).
- Alternating TX and RX timeslots—this method is, for example, supported on ISDN (Basic Rate Interfaces, BRI) lines.

When full-duplex communication is supported, a station willing to transmit a frame does not need to consider the transmission states of the other station. Instead, the current station can immediately initiate the transmission. Any issues regarding QoS, blocking, or real-time requirements must be solved by the actual network node (e.g., an Ethernet switch) using queuing and scheduling to decide what should be forwarded first to the next link.

Consequently, half-duplex media deserve closer attention, especially because several modern systems such as wireless networks are typically half duplex. Furthermore, the half duplex mode is preferred in many real-time communication systems (e.g., Ethernet Powerlink) in order to avoid bridges which would introduce some additional latency.

3.3 Synchronous Access Arbitration Concepts

Access on half-duplex media requires algorithms that select exactly one station at each time for transmission. For a particular station, the problem arises when to send data. The algorithms can be divided in deterministic and statistical approaches.

Obviously, access arbitration is the central mechanism that must be tuned and controlled to achieve or improve fairness, precedence, QoS, and real-time.

3.3.1 Static Timeslot Mechanisms

A technically simple approach to handle half-duplex media access is to assign a dedicated timeslot of fixed size to each station. If there are \( N \) timeslots within a periodic frame, then the media can support up to \( N \) stations (however, usually \( N - 1 \) because one timeslot is often used for management purposes).

This mode is also known as time division multiple access (TDMA) and is generally best suited for isochronous services such as telephony, where the communication patterns can be described by calls or circuits (steady data rates, blocking, non-bursty). TDMA is implemented, for example, in GSM, UMTS, DECT, or TETRA.

In some cases (e.g., TETRA), multiple stations that wish to communicate with each other can statistically share a particular timeslot. The major disadvantage is that one station cannot utilize unused timeslots easily (without significant increase of the overall management complexity; an example is GSM). Therefore, TDMA is not a good choice to transmit bursty data traffic.

3.3.2 Dynamic Timeslot Mechanisms

In order to achieve better media utilization, TDMA can be implemented with dynamically negotiated timeslot sizes. Every additional station registering at the central base station will be assigned a dedicated
timeslot at the expense of the timeslot sizes of other existing stations. That is, assuming a fair environment, the base station would create new timeslots by dividing the whole time period (typically called a superframe) by the total number of stations.

Alternatively, timeslot sizes may remain constant but the base station may assign a set of $D$ timeslots for particular stations. That is, a single station may be assigned a single timeslot or $D$ timeslots, assuming the superframe consists of $N > D$ timeslots.

This solution achieves a mix of both real-time behavior and good trunk utilization and is therefore common in modern wireless cellular network solutions, including WiMAX, GSM, and UMTS.

### 3.4 Statistic Access Arbitration Concepts

#### 3.4.1 Aloha Mechanisms

The DARPA-founded Aloha media access protocol has been invented in 1971 to interconnect several minicomputers distributed over the islands of Hawaii via a single wireless channel [Abr70]. Although today Aloha is seldom used practically, it is fundamental to understand the Aloha mechanisms (especially the drawbacks) before studying modern concepts such as carrier sense multiple access (CSMA).

All classic Aloha methods described below used fixed frame sizes (cells) and did not support carrier sense (CS).

#### 3.4.2 Pure Aloha

The basic idea of Aloha is implemented in Pure Aloha:

1. Every station may transmit whenever it wants
2. Upon collision with another station (no acknowledgment received on a dedicated broadcast channel), the involved station may retransmit the message after a random time

Due to this rather anarchistic access method, the collision probability is relatively high. Let $T$ denote the fixed message length in seconds and $R$ the rate of both messages and retransmissions. Assuming that the starts of messages and retransmissions can be described via a Poisson process, the probability of zero messages within time $T$ is $\exp(-RT)$. However, a collision occurs if a particular message is sent at time $t$ and another message appears within $[t−T, t+T]$. That is, a collision occurs if another (or more) message(s) appear within a time interval of $2T$ s. Therefore, the collision probability is $1 − \exp(-2TR)$. Hence the average number of retransmissions per second is $R[1 − \exp(-2TR)]$ and adding the message rate $r$ leads to $R = R + R(1 − \exp(-2TR))$, or equivalently, $r = R \exp(-2TR)$.

Multiplying both sides by $T$ gives a relation between the channel utilization $rT$ and total channel traffic. The channel utilization reaches a maximum value of $1/2e = 0.184$ when $RT$ reaches 0.5. That is, the channel is saturated as soon as two messages per message duration time $T$ are sent on average. At maximum, only 18.4% of the channel bandwidth can be utilized.

Since the message rate $r$ is the product of the number of stations $k$ times the average single-station message rate $\lambda$, the maximum number of stations is determined solely by $\lambda$ and $T$, that is, $k_{\text{max}} = 1/2e\lambda T$.

#### 3.4.3 Slotted Aloha

The throughput of Aloha can be doubled if all transmissions must take place within fixed timeslots of size $T$, which is usually equal to the frame size. That is, all stations maintain a synchronized clock and transmissions must start at the beginning of a timeslot. In this situation, collisions can only occur within a single timeslot $T$ and not within $2T$ as in the case of Pure Aloha. Therefore, using the same considerations as above, the maximum channel utilization is 36.8%.

Today, Slotted Aloha is still used, for example, in RFID networks because the algorithm can be implemented on simple ASIC structures and the throughput requirements are relatively low.
3.5 Carrier Sense Mechanisms with Exponential Backoff

3.5.1 Using Collision Detection

The most important achievement was made by Robert Metcalfe in 1972 when he created Ethernet-based on Carrier Sense Multiple Access with Collision Detection (CSMA/CD). Actually, three important improvements [Cha00] were made compared to the Aloha principle:

1. Before starting a transmission, a station must sense for an existing carrier to avoid immediate collisions. This is the CS part of the protocol.
2. Even during a transmission, the transmitting station must sense for a collision. This is the CD part of the protocol. Ten Megabit IEEE 802.3 and Ethernet use Manchester coding with a negative DC level; during collisions, each station measures an increase of the average DC level. When the collision is detected, the sender must continue the transmission for the duration of 32 bits. This jam signal ensures that all involved senders back off for a random time.
3. If the network diameter is relatively large so that the signal propagation time \( t_p \) becomes significant compared to the transmission delay \( t_a \), frames “detach” from the senders and collisions may appear somewhere on the cable.
4. To guarantee that all collisions are detectable, it is required that frames have a minimum size so that any collision signal can propagate back to the sender. In general, the dimensions of the network size and the minimum frame size must satisfy \( 2t_p \leq t_a \).
5. The random backoff time is actually the most important innovation. After detection of a collision each sender calculates a backoff time, which is the product of the slot time (a fundamental base time unit) times a randomly chosen number from the interval \([0 \ldots 2^k]\), with \( k = \min\)\{number of retransmission attempts, 10\}. The slot time is defined as the transmission delay of a 64 byte frame.

This truncated binary exponential backoff algorithm ensures that the probability of repeated collisions is exponentially reduced. The backoff is truncated by \( k = 10 \), so up to 1024 potential timeslots can be occupied by stations (theoretically).

These improvements enable Ethernet to reach throughputs close to the physical bandwidth, in particular 95% with small frames and close to 100% with large frames [Bog88]. It is important to understand that collisions in CSMA/CD-based network technologies are part of the arbitration process; the occurrence of collisions does not indicate a bad design.

3.5.2 Using Collision Avoidance

In certain environments, most importantly wireless media within the RF range, the concept of CD is not implementable because

1. The ratio between TX and RX powers is typically “dramatic” (e.g., 110 dBm), so that “listen while talk” would fully saturate (if not damage) the RX circuits.
2. Due to the long distances, the transmission time is often smaller than twice the signal propagation time; hence, collision signals (i.e., the sum of two or more frames) might not reach each sender.

Therefore, wireless technologies, such as IEEE 802.11 WLANs, utilize a collision avoidance (CA) method with the following properties:

- Collisions should be avoided if possible. Therefore, a sender willing to transmit a frame, must first listen whether the medium is non-occupied, and additionally start a random timer. Only after expiration of that timer, the frame can be sent. This initial backoff method reduces the collision probability if multiple stations wait for the medium to become idle.
• Each frame must be acknowledged by the sender. CSMA/CA interprets the absence of acknowledgment frames as collision. Obviously, acknowledgment frames have a higher priority than data frames and should not suffer from the initial backoff.

• In order to give certain (management) frames precedence over data frames, data frames must not only await the initial backoff delay but also a constant inter-frame space (the DIFS in IEEE 802.11). Management frames, for example, acknowledgments, are sent immediately after the medium is idle and after a short interframe space (SIFS in IEEE 802.11, where SIFS < DIFS) is over.

In case of collisions, again a truncated binary exponential backoff algorithm ensures that the probability of repeated collisions is exponentially reduced.

Generally, the throughput of CSMA/CA transmissions is approximately 50% lower than the nominal bit rate because each packet must await an acknowledgment.

### 3.5.3 Token Mechanisms

Statistic but collision-free medium access methods can be implemented via a token mechanism. The principle is simple: A single “token” (e.g., a bit flag) is handed over from one station to the other, allowing each station that possesses the token to initiate a data transmission. After a transmission, the station must forward the token to the next station.

Logically, the network must follow a ring topology because the single token must “rotate” through all stations. This method has been implemented in IEEE 802.4 (token bus, the physical medium is a bus) or IEEE 802.5 (token ring, also the physical medium is a ring).

Obviously token-based access arbitration provides statistic access to the medium because idle stations simply forward the token without any frame being sent. However, note that the arbitration itself is deterministic, i.e., the order is foreseeable.

In general, token mechanisms prevent channel occupation, i.e., other than with CSMA/CD-based networks, it is not possible that particular stations are preferred by a random backoff algorithm during specific network conditions.

Moreover, the rotating token provides each station periodic transmission opportunities with guaranteed maximum interval duration. This maximum token rotation time (MRT) depends on the number of stations and the maximum transmission delay per station.

Since the token is always forwarded deterministically and the MRT is a known constant, token-based access arbitration is well suited in real-time environments. Today, token mechanisms are mainly used in field bus systems, such as Profibus (IEC 61158/IEC 61784).

### 3.5.4 Polling Mechanisms

Polling mechanisms logically resemble token passing; the only difference is that a dedicated central polling station is required (while token mechanisms do not need dedicated stations, except for token monitoring and reestablishment tasks).

Instead of a token that rotates through all stations, a central polling instance polls each station whether it wants to transmit data. If the station does not want to transmit, the central station receives some sort of negative acknowledgment and immediately polls the next station.

The polling arbitration mechanism has the following distinguishing properties:

• There is no token that can be lost or that needs to be monitored and maintained. Therefore, polling methods are usually simpler to implement than token schemes.

• This simplicity allows for scalability: The central polling instance may not only poll normal stations but also other polling instances that are responsible to poll a subset of stations. This way, a hierarchical tree of polling realms can be implemented.

• The central polling instance is a single point of failure.
Polling-based access arbitration has been used with IEEE 802.12 and is still implemented in several real-time systems. The standard IEEE 802.11 defines an optional polling mode for real-time services though it is only seldom implemented because queue-based CSMA/CA is practically sufficient. Another more recent example is Ethernet POWERLINK, an Ethernet-based real-time protocol.

### 3.6 Other Media Access Issues

Although the problem of access arbitration is the most important issue, this short section shall highlight other issues that must be considered with each implementation.

#### 3.6.1 Access Fairness

Regarding the issue of fairness, the two main questions are

- Can the arbitration mechanism prevent single stations from dominating the medium?
- Are there scenarios where single stations are not scheduled for a long duration (“starvation”)?

Obviously, only statistic arbitration methods may result in unfair situations. One example for a fairness-enforcing mechanism is to introduce a post-backoff in CSMA/CA-based systems. This concept requires that every station that successfully occupied the channel must start a post-backoff timer after transmission, so that unsuccessful stations are luckier in the next arbitration cycle.

#### 3.6.2 Access Priorities

In many applications, several stations or frames (especially control frames) must be given a higher access priority. If statistical access methods are used (e.g., CSMA), priorities can be implemented by tuning backoff times. In deterministic access schemes (e.g., Token) priority information (flags) might be signaled within the token. Obviously, TDM methods provide the same access priority to each station.

#### 3.6.3 Quality of Service

Again, the problem of QoS is basically an issue with statistical access schemes because token or polling principles inherently provide constraints on delay. Obviously, QoS is not an issue with (synchronous) TDM techniques.

With statistical access schemes, prioritized queuing and consequently tuned backoff is the only method to implement QoS. For example, IEEE 802.11e defines up to eight queues per station (practically, WiFi MultiMedia only requires four queues) to differentiate higher or lower priority packets. Each queue can be considered as a virtual station with its own backoff constraints. That is, the truncated binary exponential backoff algorithm is implemented for each queue separately while the range of possible random numbers (the contention window, CW) has a different size in each queue, depending on the priority.

This QoS mechanism does not guarantee that high-priority frames are always preferred when low-priority queues are also filled; rather, this mechanism provides statistical QoS—which turned out to be sufficient for practical interactive voice and video services.

#### 3.6.4 Hidden Stations

In wireless environments, some stations may interfere with sessions between other pairs of stations. This is especially an important issue when the disturbing stations cannot detect the peers of existing sessions.

For example, assume that station A communicates with station B. Another station C can detect A but not B. Therefore, CSMA/CA on C will not take care about B and station A will suffer from collisions caused by the simultaneous transmissions of B and C. In this example, stations B and C are “hidden” from each other.
One (optional) solution defined in IEEE 802.11 (WLAN) is to announce a transmission via a dedicated request-to-send (RTS) frame, which are confirmed by the receiver by a clear-to-send (CTS) frame. Both frames, RTS and CTS, carry an estimation of the total handshake duration. Hidden stations will receive either the RTS or the CTS frame and shall be silent during the announced handshake duration.

That is, the total transaction consists of one RTS, one CTS, one data, and one acknowledgment frame. Clearly, this four-way handshake decreases the throughput of plain CSMA/CA by approximately 50%; only 25% of the bandwidth remains for data throughput.

In reality, the RTS/CTS mechanism is typically only used in point-to-multipoint wireless bridging scenarios, where the concurrent non-headend bridges use directional antennas (pointing to the single head-end bridge) and therefore cannot see each other.

References

