# A Survey on Medium Access Control in Wireless LANs

Addendum to the First Year Report

Bogdan Roman

September 2007

# Contents

# 1 Introduction

The rapid technological advances and innovations of the past few decades have pushed wireless communication from concept to reality. Advances in chip design have dramatically reduced the size and energy requirements of wireless devices, increasing their portability and convenience. These advances and innovations, combined with the freedom of movement, are among the driving forces behind the vast popularity of wireless communication. This situation is unlikely to change, especially when one considers the current push toward wireless broadband access to the Internet and multimedia content.

With predictions of near exponential growth in the number of wireless users in the coming decades, pressure is mounting on government regulatory agencies to free up the RF spectrum to satisfy the growing bandwidth demands. Given the slow reaction to such demands and the high cost of licensing, wireless users are typically forced to make due with limited bandwidth resources, which has put an emphasis on efficient access mechanisms to the wireless medium, in short, medium access control (MAC).

The role of the MAC protocol is to determine when a node is allowed to transmit its packets. It typically controls all access to the medium. The specific functions associated with a MAC protocol vary according to the system requirements and application. Design and complexity as well as performance are also affected by the network architecture, communication model, and duplexing mechanism employed. Fig. 1 shows the MAC protocol within the simplified OSI model.

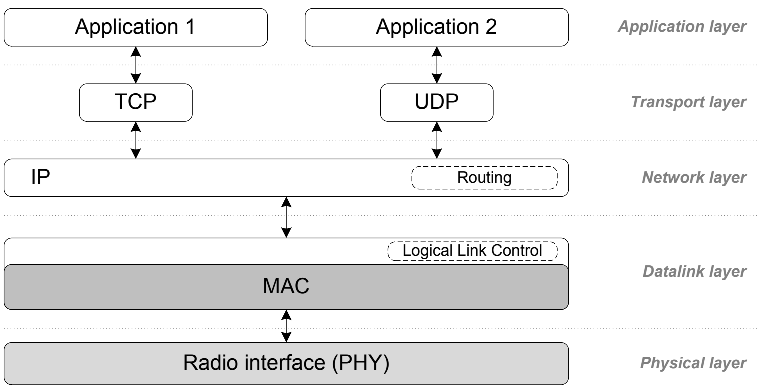


**Fig. 1**. Position of the MAC protocol within the simplified protocol stack.

Depending on the system requirements and application, the MAC protocol should provide the following main features:

• *High throughput*. A high network throughput, or high efficiency, is achieved when the protocol has low overhead and low collision probability. It is also equivalent to low frame delay.

• *Scalability*. The network throughput should be unaffected by increases in number of nodes. Naturally, the per-node throughput decreases as the number of active nodes increases but ideally, this should follow a linear decrease.

• *Topology independence*. The network throughput should be transparent to network topology changes. In high mobility scenarios, where the rate of nodes joining and leaving the network is high, as well as node velocity, the protocol should be able to maintain a high network throughput.

• *Fairness*. The protocol should not give priority to particular nodes over the others.

• *Quality of Service (QoS)*. When multiple traffic classes are defined, the protocol should be able to manage the bandwidth resources in order to assign different priorities to real-time and best-effort nodes.

• *Energy efficiency*. The complexity of the MAC protocol should be low to yield low power consumption.

It is extremely difficult to design a MAC protocol with all mentioned features. In practice, MAC protocols are usually *tweaked* towards only a subset of features depending on the architecture and application requirements. For instance, sensor networks put a high emphasis on energy efficiency and less on QoS or topology independence while densely populated broadband networks have stringent QoS and scalability requirements.

# 2 Background

## 2.1 Carrier Sense Multiple Access and IEEE 802.11

There are several fundamental MAC protocols for accessing the wireless medium.

Frequency Division Multiple Access (FDMA) divides the channel bandwidth into $N$ frequency subchannels separated by guard bands to avoid cochannel interference. A node is assigned one or more subchannels for exclusive use. The advantage of FDMA is the ability to accommodate $N$ simultaneous transmissions collision-free. However, this comes at the cost of longer transmit times, hence longer delays as well, and inefficiency when nodes lack data to transmit.

Time Division Multiple Access (TDMA) divides the channel into $N$ equal time slots that are then organized into synchronous frames. A node is assigned one or more time slots and transmission occurs in a serial fashion, nodes taking turns in accessing the channel. The advantage is that a transmission uses the full channel bandwidth but there exists the same inefficiency problem when nodes have no data to transmit. On the other hand, time slots are easier to manage and dynamic slot assignment can improve efficiency.

In contrast with FDMA and TDMA, Code Division Multiple Access (CDMA) allows multiple transmissions to occupy the channel at the same time without interference by using spatial coding techniques which spread the information bits over a broadened channel[1], allowing the information retrieval from the combined signal.

[1]This technique is called spread spectrum and in CDMA it is achieved either through direct sequence (multiplying the user data signal with a much higher rate pseudo-noise signal) or by frequency hopping (shifting the carrier frequency according to a pseudo-random sequence).

This has the advantage of increased frequency diversity, making transmissions less susceptible to fading. The tradeoff lies in increased hardware complexity and cost and difficulty to manage as well as stricter synchronization.

In ALOHA systems [1], the main feature is, oddly enough, the complete lack of medium access control. A node is allowed to send immediately if it has data to send which, of course, results in a high rate of collisions when there are multiple nodes. Slotted schemes have been shown to dramatically improve efficiency [2] and a further improvement can be obtained by allowing nodes to transmit with a certain probability, a scheme known as $p$-persistent ALOHA [3]. This decreases the collision probability but increases delay at the same time.

The MAC technology that prevailed in Wireless LANs (WLANs) is the Carrier Sense Multiple Access (CSMA) which is, in its most basic form, a listen-before-talk scheme. The most common CSMA protocols were first analyzed by Kleinrock and Tobagi [4]. Persistent CSMA, which continuously listens to detect an idle channel, results in collisions when multiple nodes are listening. Non-persistent CSMA introduce randomization to lower the collision probability by waiting for a random amount of time when the channel is detected busy, each time increasing the time interval exponentially until the channel is found idle or a maximum interval is reached. In $p$-persistent CSMA schemes [5], the channel is time-slotted but slots are not time-synchronized and nodes perform carrier sensing in each slot. If the channel is found idle, the node transmits its data with probability $p$. A busy channel will make the node wait for a random number of slots before restarting the procedure. The minimum slot length is imposed mostly by the physical layer used.

One of the main reasons behind CSMA's success, as was the case for Ethernet also, in the words of Ronald Schmidt is that "it is incredibly simple, elegant and robust". However, besides simplicity the other major reason is its ability to perform well in both distributed and centralized architectures. Just as Ethernet prevailed in the wired world over other more promising solutions, thanks to advances in chip design which increased processing power and bandwidth and lowered costs making sense to move more intelligence from the network and switches into the terminals, CSMA was adopted by the IEEE as part of the WLAN standard 802.11 [6], also known as *WiFi*. The IEEE 802.11 uses a variant of CSMA called CSMA with Collision Avoidance or CSMA/CA.

At the base of the IEEE 802.11 are the landmark MACA protocols. The MACA protocol [7] introduced a handshaking mechanism called Request-To-Send/Clear-To-Send (RTS/CTS). Before transmitting the user data, the source node sends a short RTS frame including the destination address. The destination node replies with a CTS frame. This way, the hidden and exposed node problems are alleviated since both one and two hop neighbors are made aware of the incoming transmission. MACAW [8] introduced a positive ACK scheme to aid with retransmission. Before the ACK, the source sends a short control frame called Data Sending to alert exposed nodes of the incoming ACK. The MACA/BI (MACA By Invitation) [9] protocol reverses the handshake direction, the destination node sending first a request to receive (RTR) frame to which the source node responds directly with the data frame. This, of course, means that nodes will have to maintain a list of their neighbors and should predict impending transmissions to them. This is done by synchronized polling and exchanging the tables with their neighbors. MACA with piggyback reservations (MACA/PR) [10] brings channel reservation into play by having all nodes maintain a table of the reservations made by its neighbors. A reservation is made by first completing one RTS/CTS exchange and then by including the frame duration of the next data frame in the current frame's header. A positive ACK is then sent by the destination including the correct time interval. Both one and two hop neighbors update their reservation table and defer their transmissions. Provided the RTS/CTS handshake is successful, nodes can exchange data in a collision-free manner which provides better Quality of Service support.

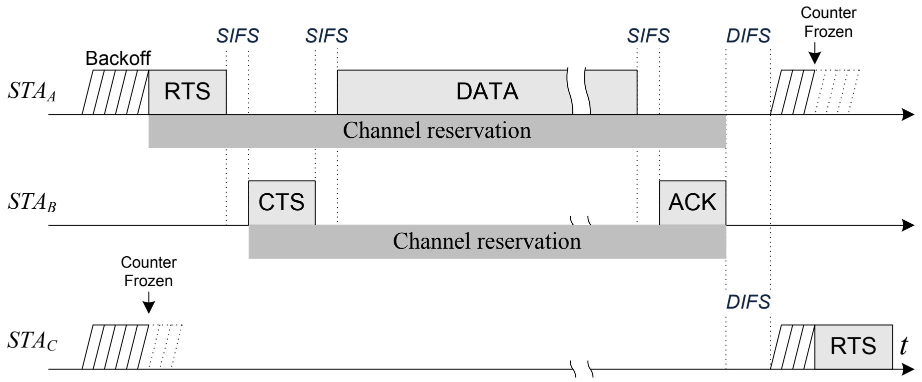


**Fig. 2**. Data exchange in IEEE 802.11. $\mathrm{STA_A}$ wins the contention and sends a data frame to $\mathrm{STA_B}$.

In the IEEE 802.11, random access through node contention, called Distributed Coordination Function (DCF), is one of the most important components. It uses a binary exponential backoff (BEB) algorithm similar to the improved slotted ALOHA protocol by Jeong and Jeon [11]. Each station starts with a contention window (CW) equal to $CWmin$ and initializes a counter to a random value in $0 \ldots CW-1$. The CW is split into backoff slots of equal duration[2]. Whenever the medium is sensed idle for more than a DCF Inter-Frame Space (DIFS) interval, the station starts the backoff procedure and decrements the counter each slot as long as the medium is sensed idle within that slot. When the counter reaches zero the node transmits its data frame. If the medium is detected busy while backing off, the counter is frozen and decrementing is resumed next time the medium is sensed idle for more than DIFS. If the counters of two or more stations reach zero at the same time then the data frames transmitted by the stations will collide. The absence of the CTS frame will indicate the sending stations that there has been a collision of the transmitted RTS frames. At this point, the colliding stations double their CW and the retransmission will take place in the same manner. By using larger CWs after collisions, the algorithm ensures a lower collision probability of retransmitted frames. The CW is doubled until a maximum $CWmax$ and is reset to $CWmin$ after a successful transmission[2]. Because of random counter initialization, the freezing and resuming of counters and depending on the number of active nodes, the actual contention session duration (i.e. the time spent backing off between two consecutive transmission attempts) is variable: it decreases when the number of nodes increases.

## 2.2 Distributed, Centralized or Hybrid Access

Most MAC protocols are tuned for either distributed (a.k.a. ad hoc), or centralized (a.k.a. infrastructure) architectures. Cellular telephony is a good example of centralized architecture and several MAC protocol types have been used throughout time. The AMPS system [14] is based on FDMA, having 832 frequency channels. Access to the control channels follows a CSMA protocol and collisions are resolved through randomized retransmission. The DAMPS, or IS-136 [15], the digital version of AMPS, is using the same channels but introduce time allocation, slotting the channel into 6 time slots to support data transmissions of up to 48 Kbps. GSM [16] is using TDMA with a frame of 8 time slots on top of FDMA while the IS-95 cellular system [17] and its successors, CDMA2000 and UMTS [18], are using CDMA, nodes being distinguished by their pseudo noise sequences.

Wireless asynchronous transfer mode (Wireless ATM) is a high-performance connection oriented switching and multiplexing system aimed at providing QoS. Communication typically follows a three step model: the node first makes a request for specific QoS to the base station (BST); the BST then allocates the uplink and downlink channels accordingly and schedules the transmissions; finally, the data is sent over the set channels according to the schedule. Notable wireless ATM MAC protocols are PRMA/DA [19] and DSA++ [20] both using frequency division duplexing and MASCARA [21] and DTDMA [22] which use time division multiplexing. The PRMA/DA and DTDMA protocols have the advantage of offering CBR, VBR and ABR traffic guarantees.

In WLANs, where we are not typically dealing with long range connectivity, the interest in supporting both centralized and distributed architectures is high, especially when considering multi-hop networks.

Allocation protocols can be used to assign nodes transmission schedules and a simple method of achieving this is by using TDMA based protocols with the frame size ideally equal to the number of nodes in the network to provide collision-free access. However, because of the variable rate of nodes joining/leaving the network and the need for global network parameters as input, this method would be highly inefficient. The Time Spread Multiple Access (TSMA) [23] protocol assigns multiple slots per node in a frame, by using the properties of Galois fields with respect to the maximum network degree to achieve shorter schedules and provide time bounded access. Although TSMA guarantees logarithmic grow in frame length with the number of nodes, the collision rate is high due to multiple assigned slots

[2] In IEEE 802.11a [12], the values for $CWmin$ and $CWmax$ are 16 and 1024 backoff slots respectively. Each backoff slot is $9\mu s$ but hardware implementations such as the Atheros AR5002X [13] can go as low as $6\mu s$.

per node, decreasing efficiency. A better approach, inspired by Wireless ATM, is to use dynamic allocation by employing short duration reservation slots in which nodes contend for access to data slots. With classic TDMA this consists in nodes taking turns in contending for slot reservations with the earliest node succeeding such as the protocols by Cidon and Sidi [24] or Ephremides and Truong [25]. However, this introduces unfairness that is alleviated by the use of a reordering policy which still needs global network parameters. A very notable protocol that aims to be independent of the network size, hence arbitrarily scalable, is the Five Phase Reservation Protocol, FPRP [26]. It uses a complex frame structure with two subframes: $k$ information frames each consisting of $l$ slots and a reservation frame preceding each information frame, consisting of $l$ reservation slots corresponding to the $l$ information slots. A TDMA schedule for the next $k$ information frames is created after contention takes place in the reservation frame. Although it is shown to yield virtually collision-free schedules, the amount of overhead created by the numerous hardware switches and control frame exchanges is very high making FPRP unstable in larger networks.

Another approach to dynamic allocation introduced by Sharp *et al* [27] is to permanently assign a node one time slot but permit nodes to reuse idle slots through CSMA-based contention thus increasing efficiency. In each slot, the nodes first sense the channel and if found idle, they transmit at a randomly chosen time instant, reducing collision probability. However, hidden nodes are a problem and efficiency is reduced because of the average waited time before transmission. The ADAPT protocol [28] addresses these problems by using an RTS/CTS handshake first in a priority interval by the owner node and then in a contention interval by nodes wishing to claim the slot in case it was detected as idle during the priority interval. A transmission interval is then used for data transmission. Any collisions from the contention interval are resolved with a BEB algorithm. Multicast however is not supported and the ABROAD protocol [29] makes this possible by changing the RTS/CTS handshake in the contention interval into a negative acknowledgement handshake in which destination nodes only reply with a Negative CTS in case a collision is detected during the contention interval or remain silent otherwise. The AGENT protocol [30] combines the features of ADAPT and ABROAD to support both unicast and multicast in a more reliable manner. This class of protocols may work well in fixed or highly controlled wireless networks (e.g. military) where slot assignment is effectively possible but fails in usual scenarios where nodes join and leave the network randomly.

An approach that aims to address the hidden and exposed node problems of CSMA is to use a separate channel for signalling. The Busy Tone Multiple Access (BTMA) [31] makes any node that detects a transmission on the data channel to transmit a narrow-band jam signal on a separate channel thus silencing other nodes from transmitting. This effectively solves the hidden node problem but amplifies the exposed node problem. Similar to MACA-BI, the Receiver Initiated BTMA (RI-BTMA) [32] protocol minimizes the number of exposed nodes by having only the destination node transmit the busy tone. However, the destination can only begin transmitting the busy tone after decoding the frame header and identifying the destination address with its own, hence frames are still subject to collisions during this time. The Wireless Collision Detect (WCD) [33] protocol combines BTMA and RI-BTMA, using two distinct busy tone signals on the control channel, acting as RI-BTMA when detecting itself as destination and as BTMA otherwise. A drawback of using separate channels for intermittent signalling is that some RF bandwidth is wasted and that nodes need to have multiple radios turned on.

The IEEE 802.11 approach is at its roots a distributed access scheme. IEEE 802.11 does not allocate time slots or maintain transmission schedules nor does it specify a frame size. Time is slotted (but not synchronized) only during contention when nodes perform carrier sensing and the duration of a backoff slot is very short, enough to accommodate reliable carrier sensing and RxTx turnaround. IEEE 802.11 allows easy infrastructure functionality through the use of either a simple polling scheme called Point Coordination Function (PCF) or the same DCF. The IEEE 802.11 RTS/CTS implementation not only addresses the hidden and exposed node problems but is also used for channel reservation. The header of all frame types (i.e. RTS, CTS, Data and ACK) contains a *duration* field indicating the time left until the end of the transmission (ACK included). Yet another important role of the RTS/CTS implementation in IEEE 802.11 is that it effectively reduces the duration of collisions to the duration of one RTS frame plus a DIFS interval and the random backoff deferral. This becomes very efficient, especially at higher loads, the throughput decreasing little even though collision rate increases. Contention using the DCF coupled with RTS/CTS in multi-hop and larger networks has proven to be a very efficient access method and most manufacturers of infrastructure IEEE 802.11 equipment do not implement the PCF function[3]. Considering the technological advances in hardware processing speed, bandwidth and power consumption, it makes sense to migrate intelligence from the network and switches towards the terminals.

The simplicity and robustness of the CSMA + BEB and the efficient implementation of the RTS/CTS have led to a wide adoption of the IEEE 802.11 as the standard for WLANs. The European Telecommunications Standard Institute (ETSI) also proposed a Wireless LAN standard called HiperLAN [34] with its high speed variant HiperLAN/2 [35]. HiperLAN/2 is at its roots a slotted, infrastructure-based access scheme based on a centralized TDMA scheme with a fairly complex frame structure, each slot containing up to 4 phases: broadcast, downlink, uplink and direct link. The MAC payload size is short and fixed at 48 bytes, similar to ATM systems, aiming to achieve reliable QoS. Resource allocation is connection oriented, centralized in the AP and performed dynamically according to the node requests and channel conditions. Although it was built with high performance and QoS in mind, HiperLAN/2 did not enjoy success, the complexity of the hardware and the lack of true distributed access control being among the main reasons.

[3]Other reasons for this are given in Section 2.4.

## 2.3 The Quest for High Efficiency (Throughput)

Increasing the bandwidth efficiency has probably been the most researched topic in the MAC protocols field. This section will not concentrate on artificial methods of increasing throughput, such as on-the-fly hardware compression, packet aggregation or block ACK-ing, as these apply to all MAC protocols in the same manner and can increase the network throughput to values higher than the channel bit rate. The interest is in presenting the attempts made at optimizing the access control mechanisms that increase efficiency, i.e. minimize overhead to approach the channel bit rate while maintaining the key protocol features mentioned in Section 1.

In a landmark paper by Bianchi [36], the analytical model for the IEEE 802.11 was devised, based on a two-dimensional Markov chain making an innovative key approximation that all stations have equal probability of collision. The model was later on improved by Ziouva and Antonakopoulos [37] and ultimately by Foh and Tantra [38] to account for freezing and resuming of counters as specified in the standard. As Bianchi showed in the initial paper, because of the nature of the BEB algorithm with freezing counters, the only way to make the efficiency of IEEE 802.11 scale with the network size is through adaptive methods. There have been a tremendous number of proposed solutions to improve the efficiency of the IEEE 802.11.

MACAW [8] proposes that colliding nodes should increase their CWs multiplicatively while successful nodes should decrease their CWs linearly instead of resetting to $CWmin$. Song *et al* [39] propose an exponential increase exponential decrease (EIED) algorithm while Kuo *et al* [40] propose a Multiplicative Increase Multiplicative Decrease (MIMD) algorithm of the contention window size, based on the same principles of MACAW. In addition, MACAW proposes a backoff broadcast mechanism so that the CW value used by successful nodes is copied by the neighbors. A similar scheme that redistributes successful backoff parameters to neighboring nodes is taken by Kwon *et al* [41] which broadcasts the backoff timer so that recent successful nodes can use smaller CWs while competing nodes decrease their CW exponentially by monitoring the channel for idle slots. Deng *et al* [42] show that such schemes are still far from a theoretical limit and proposes a hybrid Linear/Multiplicative Increase and Linear Decrease (LMILD) algorithm in which nodes overhearing collisions will also increase their CWs but linearly. A different approach at optimizing the CW size is taken by Wang *et al* [43]. In this work, the protocol monitors the channel for $c$ consecutive successful transmissions and exponentially decreases the CW, obtaining a lower collision probability. A newer proposed scheme, called P-DCF [44], monitors the channel for idle durations and constructs a history of past transmissions, ultimately selecting an optimal backoff time dynamically and reducing the collision probability.

Through a thorough analysis, Cali *et al* [45, 46] prove that the IEEE 802.11 operates far from the protocol capacity depending on the number of nodes and propose a dynamic tuning of the CW size by estimating the number of active nodes in the network. The estimation is done locally by observing the channel status for collisions and idle states. Bianchi and Tinirello [47] show that an improved method of estimating the number of active nodes is by using Kalman filtering. The scheme is also compared against a simpler auto regression moving average (ARMA) filtering and shows better tracking ability. An even more sophisticated but however up to 33% more accurate estimation of the number of active nodes is proposed by Zheng and Zhang [48] in which a sequential Monte Carlo technique is applied, namely the unscented particle filter which add unscented transformation to particle filtering. A different and simpler approach is made by Li *et al* [49] to adaptively tune the decrementing of the backoff timer with a probability $p_d < 1$ instead of every backoff slot to yield lower collision probability. The authors improve the scheme [50] by introducing exponential increase of CWs in case of collisions.

Instead of using BEB or variants of BEB, Borgonovo *et al* [51] propose a scheme based on Reliable Reservation ALOHA (RR-ALOHA) for slotted access in which nodes can send data in one or multiple slots. Channel monitoring is performed and along user data, nodes also include frame information consisting in the perceived slot allocation which achieves a two-hop awareness among nodes but increases overhead. A scheme devised CSMA/CP (Collision Prevention) [52] is proposed by You *et al* in which binary countdown[4] is used as contention algorithm on a separate control channel. An RTS/CTS handshake is employed as well to overcome the inherent issue of binary countdown schemes which is subject to hidden node interference. However, similar to many other multi-channel proposed solutions, such as the ones presented in the previous section or newer ones like DBTMA [53], DBTMA-DA [54], Bi-MCMAC [55] or CA-CDMA [56], the usage of a separate channel for control messages or contention decreases efficiency considerably especially considering the need for guard bands and multiple transceiver. An interesting solution that uses binary countdown but no additional channels nor RTS/CTS is SYN-MAC [57]. In this work, the authors propose the usage of a feedback slot at the end of the binary countdown contention which effectively addresses the hidden node problem.

## 2.4 Quality of Service

The provision of reliable Quality of Service in wireless LANs is another actively researched topic. In the case of IEEE 802.11, the PCF scheme may be used in a centralized architecture. However, as many studies such as [58–60] show, PCF is unreliable, may reach instability and the QoS ability is minimal. Consequently, most manufacturers of infrastructure equipment do not implement PCF at all, DCF offering similar performance. The DCF itself, while being simple and robust, does not differentiate traffic priorities. While QoS can, in theory, be achieved easier by

[4] In binary countdown schemes, the winner is selected based on a $k$-bit binary number. A node gives up contention when detecting a higher number. E.g. given four numbers 0010, 0100, 1001, 1010, the first two nodes will give up after comparing the first (highest) bit, and the third gives up after the third bit. Finally, 1010 wins the contention.

always using a coordinating node, achieving dynamic QoS is more attractive and bandwidth efficient since nodes can reclaim bandwidth dynamically without waiting for polling signals. Again, the technological advances, especially in processing power, make a good argument to migrate intelligence into the terminal in this case as well.

An interesting solution to effectively prioritize traffic in distributed architectures is proposed by Sobrinho and Krishnakumar [61] in which QoS is achieved by the usage of unmodulated bursts with the burst duration proportional to the node's elapsed waited time. The authors show that a dynamic TDMA schedule is finally created thus offering time bounded access. The scheme is however limited to single-hop and small networks, large networks causing unacceptable delays due to the long bursts duration. Numerous attempts at improving the QoS ability of the IEEE 802.11 have been made. Deng and Chang [62] initially proposed a simple scheme based on shorter inter-frame spaces and smaller CWs to differentiate priorities. Lower priority nodes would be silenced sensing the channel busy with transmissions started earlier by higher priority nodes while smaller CWs statistically increase channel access for high priority nodes. This however can introduce starvation[5] or bandwidth stealing[6] in case of many best-effort nodes. Xiao [63] proposes backoff-based priority schemes by differentiating between several backoff parameters: $CWmin$, CW increasing factor and the maximum backoff stage. However, the great number of parameters to be tuned make this solution unattractive for real applications. Yan *et al* [64] propose an adaptive priority scheme in which nodes monitor the channel extracting transmission duration and data rate information and maintaining a transmission history table for all neighbors. The optimum CW size is then dynamically computed. Hidden nodes however can severely affect the scheme introducing unfairness and breaking QoS. Of note are the survey papers on QoS protocols and enhancements by Zhu *et al* [60] and Ni *et al* [59]. An interesting point that results from these works is that most of the studied schemes are focusing on tuning either the inter-frame spaces or CW limits or a combination of the two.

The recently approved IEEE 802.11e standard [65], which enhances the initial IEEE 802.11 protocol to support better QoS, introduces the concept of transmit opportunity (TXOP), which refers to the time duration in which a node is allowed to send a data frame, and several sets of concurrent fixed-value backoff parameters corresponding to the different traffic classes: $CWmin$, $CWmax$, arbitration inter-frame space (AIFS) and TXOP limit. Notable works which analyze the performance of the IEEE 802.11e and provide improved analytical models based on Bianchi's model [36] are those of Kong *et al* [66], Xiao [67] and Clifford *et al* [68] in which it is shown that the IEEE 802.11e improves the QoS capability of both

[5] When the number of real-time nodes increases, the bandwidth consumption can reach the maximum protocol throughput and low priority nodes may be denied channel access, an effect knows as starvation.

[6] Depending on the implementation, a great number of low-priority nodes could end up stealing bandwidth from higher priority nodes thus breaking QoS.

PCF and DCF, at the same time improving on some of their known issues such as unfairness due to nodes employing different data rates or late timer countdown. Clementson *et al* [69], Wiethölter and Hoene [70] and Ni [71] propose good simulation models for IEEE 802.11e while Clifford *et al* [68] and Dangerfield *et al* [72] show comprehensive real world performance figures. These works provide good comparison models and reference points for future research on QoS.

Li and Xiao [73] enhance the IEEE 802.11e by using a global data parameter control scheme integrated with a measurement-based admission control scheme in which the AP dynamically controls best-effort data parameters of nodes based on traffic condition and nodes listen to available budgets from the AP and decide to accept or reject the streams. The two schemes improve fairness and priority differentiation but at the cost of increased complexity both at the AP and the nodes. Ryu *et al* [74] propose a scheme based on pseudo-collision of so called *super-slots* which are virtual slots consisting of $D$ backoff slots, where $D$ is specified for each traffic class. The scheme yields lower collision probability than IEEE 802.11e between different priority nodes, thus alleviating starvation. However, it is not scalable with the network size and dynamic tuning of $D$ becomes necessary. Ansel *et al* [75] point out the poor VBR support of IEEE 802.11e and propose an algorithm using adaptive queue length estimations to tune the time allocation schedule for each node achieving better fairness but at the cost of decreased scalability and increased starvation. A different approach at addressing the starvation and bandwidth stealing in IEEE 802.11e WLANs is that of Paschos *et al* [76] which attempts to achieve a TDMA behavior similar to wireless ATM. Starvation and collision probability are reduced by using distributed admission control algorithms for each traffic class through local timers and a fixed virtual frame duration $F_d$. However, because of the fixed $F_d$, the scheme's scalability is impaired and bandwidth stealing still remains an issue at higher loads.

# 3 Comments, Open Questions and Motivation

Wireless capability has become increasingly available in many mobile devices such as laptops, smart-phones, handhelds and vehicles. Coupled with the current push towards wireless broadband access to the Internet and multimedia content, we will soon witness very large and highly loaded networks. High mobility networks, such as vehicular networks, experience rapidly changing topologies imposing harsh conditions for wireless communication. Such factors make efficient access control to the medium a high priority.

## 3.1 High Efficiency and Scalability

Many solutions for improving efficiency and scalability in wireless LANs involve adaptive algorithms. A common problem in the presented algorithms is that channel status or network state measurements are necessary which means the station needs to keep its radio turned on. With local measurements, the adaptation time is also long which does not play well in high mobility or variable loads networks, decreasing tracking ability considerably. Another problem is unfairness which strikes real scenarios where nodes leave and join the network. The complexity of such schemes is also high, especially for the schemes that need to maintain not only local but also global network state tables. These are among the main reasons why such adaptive schemes have not been adopted, the industry usually aiming for simplicity and robustness.

The non-adaptive schemes, such as the ones based on binary countdown, although providing low collision probability, fail to provide high efficiency, particularly at high bit rates. Solutions involving separate control or contention channels are inherently inefficient since an entire channel is practically wasted, especially when considering guard bands, and the hardware complexity is high requiring multiple radios (hence also high power consumption).

Most studied protocols, including recent ones, offer results for low data bit rates such as 1-2 Mbps and seldom up to 11 Mbps yet bit rates of 54 Mbps have been available for more than half a decade. One aspect that is infrequently mentioned in literature is that the PHY preamble and most of the PHY header have fixed duration, as do the inter-frame spaces. This makes the overhead more important as the bit rate increases and also puts an upper limit to the maximum attainable throughput. Hence, although at low bit rates many of the protocols show good efficiency, the same does not hold for high bit rates.

The literature survey has showed that scalability is a common issue, most results being presented only for small networks of up to 20 nodes, sometimes 50 nodes. Comparison is also done mainly against the IEEE 802.11 without RTS/CTS, a.k.a basic mode, which is known to be inefficient at moderate or higher loads and non-scalable with the network size. The BEB algorithm itself imposes limits to the scalability of the MAC protocols; good scalability can be obtained by using large CWs but that inherently introduces long delays at lower loads. Alternatives to BEB, such as SYN-MAC [57], show better scalability but unfortunately fail to provide high efficiency at high bit rates.

## 3.2 Reliable Quality of Service

Starvation and bandwidth stealing as well as scalability are still major issues in all BEB-based (and similar) protocols, which includes the improved IEEE 802.11e, due to the nature of the BEB algorithm that freezes and resumes counters at different transmission attempts instants. The solutions not involving BEB (or similar), although providing better QoS through allocation schemes either fail in the efficiency front because of large overhead or have a high degree of complexity.

Many of the proposed solutions still involve adaptive algorithms to reduce unfairness and to provide better real-time traffic prioritization. As discussed above, these algorithms introduce various important drawbacks, making them attractive mostly from a scientific standpoint only. Dynamic QoS which does not suffer from polling delays, variable contention times, high complexity, starvation or bandwidth stealing is still the subject of ongoing research.

## 3.3 Notable Recent Works

Among the many studied works related to improving medium access control, important adaptive schemes are those by Bianchi and Tinnirelo [47], which introduces a novel Kalman filtering method, and of Zheng and Zhang [48], which introduces a particle filtering method with a considerable improvement in tracking ability.

Notable analytical models are that of Foh and Tantra [38], which perfects the initial IEEE 802.11 model by Bianchi [36] providing a solid base for any analysis of or comparison to the IEEE 802.11, and the QoS models by Xiao [67] which fully extend the models by Bianchi and Foh and Tantra with the QoS features introduced in the IEEE 802.11e.

A study of the CA-CDMA protocol by Muqattash and Krunz [56] shows that CDMA alone is not suited for WLANs due to several issues such as very strict synchronization, pseudo-noise signals correlation and vulnerability to hidden nodes, and that CSMA techniques considerably improve performance. Wu *et al* [57] show that a scheme based on a completely different collision resolution algorithm than the binary backoff employed in the IEEE 802.11, namely binary countdown with feedback, can offer lower collision probability and high robustness and improves performance at lower bit rates. The improved schemes by Ansel *et al* [75] and Paschos *et al* [76] show that medium access control schemes that make use of TDMA-like scheduling schemes are able to provide better QoS while maintaining high bandwidth efficiency than schemes which lack time allocation properties.

Although not directly related to improving the key MAC features listed in Section 1, the medium sensing mechanism developed by Weiss and Jondral [77] for spectrum pooling in cognitive radio networks, which uses the OFDM PHY in conjunction with selective activation of the OFDM subcarriers, and which is later analyzed by the authors in more depth [78], shows that it is possible to achieve very short durations for sensing multiple frequency sub-bands of the same RF channel.

## 3.4 Conclusion

The design of a new MAC protocol should take into account more than the MAC layer. In almost all cases, the PHY plays a crucial role that may finally render a seemingly high-performance MAC scheme low-performance or even unfeasible. A good understanding of the PHY is a requirement for a good MAC design. As was the case with the IEEE 802.11, the best MAC designs are those that take into account or even improve the PHY as well.

A MAC protocol that is highly tweaked to provide a certain functionality, such as QoS or high throughput, usually lack performance in other aspects making them too specific to the application, architecture, topology or communication model. History has shown that the adopted designs have usually been those which are scalable with the technological advances and which have a high level of generality and simplicity providing fair overall performance and allowing future improvements.

# References


[1] N. Abramsom, “Development of the ALOHANET,” *IEEE Trans. Inf. Theory*, vol. 31, no. 2, p. 11923, 1985.

[2] L. Roberts, “ALOHA packet system with and without slots and capture,” *Comput. Commun. Rev.*, vol. 5, no. 2, p. 2842, Apr. 1975.

[3] A. Leon-Garcia and I. Widjaja, *Communication network: Fundamental concept and key architecture*. McGraw-Hill, 1999.

[4] L. Kleinrock and F. Tobagi, “Packet switching in radio channels. i. carrier sense multiple access models and their throughput delay characteristics,” *IEEE Trans. Commun.*, vol. 12, p. 14001416, Dec. 1975.

[5] D. Bertsekas and R. Gallager, *Data Networks*. Prentice-Hall, 1992.

[6] 802.11-1999 (R2003), *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Std., 1999, Reaffirmed June 2003.

[7] P. Karn, “MACA A New Channel Access Protocol for Packet Radio,” in *ARRL/CRRL Amateur Radio 9th Comput. Net. Conf.*, 1990.

[8] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, “MACAW: A Media Access Protocol for Wireless LAN’s,” *Proc. ACM SIGCOMM’94*, 1994.

[9] F. Talluci and M. Gerla, “MACA-BI (MACA by invitation): A wireless MAC Protocol For High Speed Ad Hoc Networking,” in *ACM SIGCOMM’94*, 1994.

[10] C. R. Lin and M. Gerla, “Real-Time Support In Multihop Wireless Networks,” *Springer J. Wireless Netw.*, vol. 5, no. 2, pp. 125–135, Mar. 1999.

[11] D. Jeong and W. Jeon, “Performance of an exponential backoff scheme for the slotted-ALOHA protocol in local wireless environment,” *IEEE Trans. Veh. Technol.*, vol. 44, no. 3, p. 470479, Aug. 1995.

[12] 802.11a-1999 (R2003), *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications - High-speed Physical Layer in the 5 GHz Band*, IEEE Std., 1999, Reaffirmed June 2003.

[13] Atheros Communications, “AR5002X,” 2003. [Online]. Available: http://www.atheros.com/pt/bulletins/AR5002X.pdf

[14] ANSI/EIA/TIA, “Mobile stations and station compatibility specification,” EIA/TIA, Tech. Rep. 553, 1989.

[15] ——, “800MHz TDMA cellular radio interface–Mobile station-base station compatibility–Digital control channel,” EIA/TIA, Tech. Rep. IS-136, 1994.

[16] M. Rahnema, “Overview of the GSM systems and protocol architecture,” *IEEE Commun. Mag.*, vol. 43, pp. 92–100, 1993.

[17] EIA/TIA, “Mobile station-base station compatibility standard for dual-mode wideband spread-spectrum cellular system,” EIA/TIA, Tech. Rep. IS-95, 1993.

[18] M. Oliphant, “Radio interfaces make the difference in 3G cellular systems,” *IEEE Spectr.*, pp. 53–58, Oct. 2000.

[19] J. Kim and I. Widjaja, “PRMA/DA: A new media access control protocol for wireless ATM,” in *IEEE ICC96*, Dallas, TX, Jun. 1996, pp. 1–19.

[20] D. Petras and A. Krämling, “MAC protocol with polling and fast collision resolution for ATM air interface,” in *IEEE ATM Workshop*, San Fransisco, CA, Aug. 1996.

[21] N. Passas, S. Paskalis, D. Vali, and L. Merakos, “Quality-of-service-oriented medium access control for wireless ATM networks,” *IEEE Commun. Mag.*, pp. 43–50, Nov. 1997.

[22] D. Raychaudhuri and N. Wilson, “ATM-based transport architecture for multiservices wireless personal communication networks,” *IEEE J. Sel. Areas Commun.*, vol. 12, no. 2, p. 14011414, Oct. 1992.

[23] I. Chlamtac and A. Faragó, “Making transmission schedules immune to topology changes in multihop packet radio networks,” *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 23–29, Feb. 1994.

[24] I. Cidon and M. Sidi, “Distributed assignment algorithms for multihop packet radio networks,” *IEEE Trans. Comput.*, vol. 38, no. 10, p. 13531361, Oct. 1989.

[25] A. Ephremides and T. Truong, “Scheduling broadcasts in multihop radio networks,” *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 456–460, Apr. 1990.

[26] C. Zhu and M. S. Corson, “A Five-Phase Reservation Protocol (FPRP) for Mobile Ad Hoc Networks,” in *IEEE INFOCOM’98*, vol. 1, 1998.

[27] B. Sharp, A. Grindrod, and D. Camm, “Hybrid TDMA/CDMA protocol self-managing packet radio networks,” in *IEEE ICUPC95*, Tokyo, Japan, Nov. 1995, pp. 929–933.

[28] I. Chlamtac, A. Faragó, A. Myers, V. R. Syrotiuk, and G. Zaruba, “ADAPT: A Dynamically Self-Adjusting Media Access Protocol Control for Ad Hoc Networks,” *IEEE GLOBECOM’99*, vol. 1a, pp. 11–15, Sep. 1999.

[29] I. Chlamtac, A. Myers, V. R. Syrotiuk, and G. Zaruba, “An Adaptive Medium Access Control (MAC) Protocol For Reliable Broadcast In Wireless Networks,” in *IEEE ICC’00*, vol. 3, 2000.

[30] A. Myers, G. Zaruba, and V. R. Syrotiuk, “An Adaptive Generalized Transmission Protocol for Ad Hoc Networks,” *ACM Mob. Netw. and Apps.*, vol. 7, no. 6, pp. 493–502, Dec. 2002.

[31] F. Tobagi and L. Kleinrock, “Packet switching in radio channels. ii. the hidden terminal problem in carrier sense multiple-access and the busy-tone solution,” *IEEE Trans. Commun.*, vol. 23, no. 12, p. 14171433, Dec. 1975.

[32] C. Wu and V. Li, “Receiver-initiated busy tone multiple access in packet radio networks,” *Proc. ACM SIGCOMM’87*, vol. 17, no. 5, p. 336342, Aug. 1987.

[33] A. Gummalla and J. Limb, “Wireless collision detect (WCD): Multiple access with receiver initi ated feedback and carrier detect signal,” in *IEEE ICC00*, vol. 1, New Orleans, LA, Jun. 2000, pp. 397–401.

[34] ETSI, “Radio Equipment and Systems (RES); HIgh PErformance Radio Local Area Network (HIPERLAN); Functional specification,” ETSI, Tech. Rep., Jun. 1996.

[35] ——, “Broadband radio access networks (BRAN); HiperLAN Type 2; System Overview,” ETSI, Tech. Rep. 101 683, V1.1.1, Feb. 2000.

[36] G. Bianchi, “Performance Analysis of the IEEE 802.11 Distributed CoordinationFunction,” *IEEE J. Sel. Areas Commun.*, vol. 18 (3), no. 2, 2000.

[37] E. Ziouva and T. Antonakopoulos, “CSMA/CA performance under high traffic conditions: throughput and delay analysis,” *Elsevier J. Comput. Commun.*, vol. 25, 2002.

[38] C. Foh and J. Tantra, “Comments on IEEE 802.11 saturation throughput analysis with freezing of backoff counters,” *IEEE Commun. Lett.*, vol. 9, 2005.

[39] N. O. Song, B. J. Kwak, J. Song, and L. E. Miller, “Enhancement of IEEE 802.11 Distributed Coordination Function with Exponential Increase Exponential Decrease Backoff Algorithm,” in *IEEE VTC’03 Spring*, New Orleans, 2003.

[40] W. Kuo and C. J. Kuo, “Enhanced backoff scheme in CSMA/CA for IEEE 802.11,” in *IEEE VTC’03 Fall*, Orlando, 2003.

[41] Y. Kwon, Y. Fang, and H. Latchman, “A novel MAC protocol with fast collision resolution for wireless LANs,” in *IEEE INFOCOM03*, 2003.

[42] J. Deng, P. K. Varshney, and Z. J. Haas, "A New Backoff Algorithm for the IEEE 802.11 Distributed Coordination Function," in *Commun. Net. and Distributed Sys. Modeling and Simulation (CNDS)*, San Diego, CA, 2004.

[43] C. Wang, B. Li, and L. Li, "A New Collision Resolution Mechanism to Enhance the Performance of IEEE 802.11 DCF," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, 2004.

[44] N. Choi, Y. Seok, Y. Choi, S. Kim, and H. Jung, "P-DCF: enhanced backoff scheme for the IEEE 802.11 DCF," in *IEEE VTC'05 Spring*, vol. 3, Stockholm, Sweden, 2005.

[45] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 9, 2000.

[46] ——, "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit," *IEEE/ACM Trans. Netw.*, vol. 8, no. 6, 2000.

[47] G. Bianchi and I. Tinnirello, "Kalman Filter Estimation of the Number of Competing Terminals in an IEEE 802.11 network," *IEEE INFOCOM'03*, vol. 2, pp. 844–852, Mar. 2003.

[48] D. Zheng and J. Zhang, "A unscented particle filtering approach to estimating competing stations in IEEE 802.11 WLANs," *IEEE GLOBECOM'05*, vol. 5, Dec. 2005.

[49] Y. Li, K. Long, W. Zhao, and F. Yang, "RWBO(pdw): A Novel Backoff Algorithm for IEEE 802.11 DCF," *J. Comput. Sci. Technol.*, vol. 20, no. 2, 2005.

[50] Y. Li, L. Ke-Ping, Z. Wei-Liang, and C. Qian-Bin, "A Novel Random Backoff Algorithm to Enhance the Performance of IEEE 802.11 DCF," *Springer J. Wireless Personal Commun.*, vol. 36, 2006.

[51] F. Borgonovo, A. Capone, M. Cesana, and L. Fratta, "ADHOC MAC: a new, flexible and reliable MAC Architecture for ad hoc Networks," *IEEE WCNC'03*, vol. 2, pp. 965–970, Mar. 2003.

[52] T. You, C.-H. Yeh, and H. Hassanein, "A new class of collision-prevention MAC protocols for ad hoc wireless networks," in *IEEE ICC'03*, 2003.

[53] Z. J. Haas and J. Deng, "Dual busy tone multiple access (DBTMA)-a multiple access control scheme for ad hoc networks," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 975–985, Jun. 2002.

[54] Z. Huang and C.-C. Shen, "A comparison study of omnidirectional and directional mac protocols for ad hoc networks," *IEEE GLOBECOM 2002*, vol. 1, pp. 57–61, Nov. 2002.

[55] T. Kuang and C. Williamson, "A bidirectional multi-channel MAC protocol for improving TCP performance on multihop wireless ad hoc networks," in *7th ACM Intl. Symp. on Modeling, Analysis and Simulation of Wireless and Mobile Sys.*, 2004, pp. 301–310.

[56] A. Muqattash and M. Krunz, "CDMA-based MAC protocol for wireless ad hoc networks," in *4th ACM Intl. Symp. on Mob. Ad Hoc Net. & Comp.*, Jun. 2003, pp. 153–164.

[57] H. Wu, A. Utgikar, and N.-F. Tzeng, "SYN-MAC: A Distributed Medium Access Control Protocol for Synchronized Wireless Networks," *Springer J. Mobile Netw. Appl.*, vol. 10, 2005.

[58] S. Mangold, S. Choi, G. Hiertz, and O. Klein, "Analysis of IEEE 802.11e for QoS Support in Wireless LANs," *IEEE Wireless Commun. Mag.*, vol. 10 (6), 2003.

[59] Q. Ni, L. Romdhani, and T. Turletti, "A Survey of QoS Enhancements for IEEE 802.11 Wireless LAN," *Wiley J. of Wireless Commun. Mob. Comput.*, vol. 4, no. 5, pp. 547–566, 2004.

[60] H. Zhu, M.Li, I. Chlamtac, and B. Prabhakaran, "A survey of Quality of Service in IEEE 802.11 networks," *IEEE Wireless Commun. Mag.*, vol. 11, no. 4, pp. 6–14, 2004.

[61] J. Sobrinho and A. Krishnakumar, "Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 8, 1999.

[62] D. Deng and R. Chang, "A Priority Scheme for IEEE 802.11 DCF Access Method," *IEICE Trans. Commun.*, vol. 82, no. 1, 1999.

[63] Y. Xiao, “Backoff-based Priority Schemes for IEEE 802.11,” *IEEE ICC’03*, vol. 3, pp. 1568–1572, May 2003.

[64] S. Yan, Y. Zhuo, S. Wu, and W. Guo, “Priority backoff algorithm for ieee 802.11 dcf,” in *IEEE ICCCAS 2004*, Jun. 2004, pp. 423–427.

[65] 802.11e 2005, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications - Medium Access Control (MAC) Quality of Service Enhancements*, IEEE Std., 2005.

[66] Z. Kong, D. Tsang, B. Bensaou, , and D. Gao, “Performance analysis of IEEE 802.11e contention-based channel access,” *IEEE J. Sel. Areas Commun.*, vol. 22, p. 20952106, Dec. 2004.

[67] Y. Xiao, “Performance Analysis of Priority Schemes for IEEE 802.11 and IEEE 802.11e Wireless LANs,” *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1506–1515, 2005.

[68] P. Clifford, K. Duffy, J. Foy, D. J. Leith, and D. Malone, “Modeling 802.11e for data traffic parameter design,” *IEEE/ACM WiOpt’06*, pp. 1–10, 2006.

[69] F. Clementson, M. Nilsson, L. Staalhagen, and F. Hermodsson, “Performance Modeling of IEEE 802.11e using OPNET Modeler,” in *OPNETWORKS-2004*, Washington, DC, Aug. 2004.

[70] S. Wiethölter and C. Hoene, “Design and Verification of an IEEE 802.11e Simulation Model in ns-2.26,” Technical University Berlin, Tech. Rep. TKN-03-019, Nov. 2003.

[71] Q. Ni, “ns-802.11e EDCF for IEEE 802.11e Wireless LAN,” Nov. 2002. [Online]. Available: http://www-sop.inria.fr/planete/qni/,November2002.

[72] I. Dangerfield, D. Malone, and D. J. Leith, “Experimental evaluation of 802.11e edca for enhanced voice over wlan performance,” in *IEEE/ACM 4th Intl. Symp. on Modeling and Optimization in Mob., Ad Hoc and Wireless Net.*, 2006, pp. 1–7.

[73] H. Li and Y. Xiao, “Voice and video transmissions with global data parameter control for the IEEE 802.11e enhance distributed channel access,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 15, no. 11, pp. 1041–1053, Nov. 2004.

[74] J. Ryu, Y. Seok, Y. Choi, T. Kwon, , and J.-M. BONNIN, “S-EDCF: EDCF based on SuperSlot and Pseudo Collision,” in *IEEE MobiCOM’05*, 2005.

[75] P. Ansel, Q. Ni, and T. Turletti, “FHCF: An Efficient Scheduling Scheme for IEEE 802.11e,” *Springer J. Mobile Netw. Appl.*, vol. 11, pp. 391–403, 2006.

[76] G. S. Paschos, I. Papapanagiotou, S. A. Kotsopoulos, and G. K. Karagiannidis, “A New MAC Protocol with Pseudo-TDMA Behavior for Supporting Quality of Service in 802.11 Wireless LANs,” *EURASIP Journal on Wireless Comm. and Netw.*, pp. 1–9, 2006.

[77] T. Weiss and F. Jondral, “Spectrum pooling: An Innovative Strategy for the Enhancement of Spectrum Efficiency,” *IEEE Commun. Mag.*, vol. 42, pp. 8–14, Mar. 2004.

[78] T. Weiss, J. Hillenbrand, A. Krohn, and F. Jondral, “Mutual interference in OFDM-based spectrum pooling systems,” *IEEE VTC’04 Spring*, vol. 4, pp. 1873– 1877, 2004.