

Understanding the Gap between the IEEE 802.11 Protocol Performance and the Theoretical Limits

Mathilde Durvy

School of Computer and Communication Sciences

EPFL

CH-1015 Lausanne, Switzerland

mathilde.durvy@epfl.ch

Patrick Thiran

School of Computer and Communication Sciences

EPFL

CH-1015 Lausanne, Switzerland

patrick.thiran@epfl.ch

Abstract—The ability of the IEEE 802.11 Medium Access Control (MAC) protocol to perform well in multi-hop ad hoc networks has been recently questioned. We observe levels of spatial reuse that are 30% to 50% away from the theoretical limit. The goal of this paper is to answer the following question: what prevents the IEEE 802.11 MAC protocol from operating at the limit determined by its physical layer? We identify three problems in the contention resolution mechanism of the IEEE 802.11 MAC protocol, and we show that they account for most of the gap separating the actual and optimal performances of the protocol. For each of the problems, we propose a solution that, once implemented, allows us to quantify the impact of the problem on the performance of the IEEE 802.11 MAC protocol. The resulting protocol operates 10% to 15% away from the theoretical limit. Finally, we show that reducing the overhead of the protocol to some negligible quantity brings the spatial reuse of the protocol to the theoretical limits. It also makes apparent the powerful organizing capacity of the IEEE 802.11 MAC protocol.

I. INTRODUCTION

The IEEE 802.11 MAC protocol (802.11 protocol, for short) is probably the most widely used MAC protocol for ad hoc networks. In the Wireless LAN (i.e., single-hop) setting, its behavior can be modeled accurately using a Markov chain formalism [1]. In the multi-hop setting, the performance of the 802.11 protocol has been studied extensively, mainly through simulations, and numerous changes have been proposed to improve its performance. Yet, there still remains a large gap between the fundamental limits (set by its physical layer) and the achieved performance. In this paper, we evaluate *all* the factors that contribute to bridge this gap.

After having briefly reviewed the features of interest of the 802.11 protocol and its performance in Section II, we identify three main causes of inefficiencies in Section III:

- The gagged node. A node is silenced by repeated Request-To-Send (RTS) messages that are not followed by a data packet transmission.
- The jammed node. A node is jammed by a data packet transmitted between two other nodes and cannot extract

The work presented in this paper was supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

valuable state information in control messages sent concurrently.

- The focused node. A node focuses its transmission attempts on a single of its neighboring nodes that experiences high contention.

These three inefficiencies have received very limited attention in the literature. We perform an in depth study of these inefficiencies and show that they affect severely the ability of the 802.11 protocol to schedule a high number of concurrent transmissions. On a simple network topology we observe a performance 30% to 50% away from the optimal achievable performance. We propose a remedy to address each of these inefficiencies and show that the gap to the fundamental limit is then reduced to less than 15%.

In Section IV, we show that the remaining performance gap is due to the overhead of the control messages (approximately 5%) and to the non-negligible time spent in the backoff process (necessary to limit collisions). Contrary to the simple remedies that solve the three previous inefficiencies, the modifications that would need to be brought are not realistic in practice (one would need to reduce the backoff time and the overhead message length to virtually zero) nor even desirable (the gain in spatial reuse makes the protocol very unfair). Interestingly, it shows that a decentralized protocol like 802.11 is capable of organizing the transmissions in schedules with a maximum number of concurrent transmissions.

Finally, after having reviewed the state of the art in Section V, we summarize the different factors that affect the performance of the 802.11 protocol in Section VI.

II. IEEE 802.11

In this section we describe the 802.11 protocol from a standard and an implementation point of view. We then quantify the limits imposed by its physical layer, and we study its performance in different settings.

A. Protocol

We first provide a high-level overview of the 802.11 protocol in ad hoc mode. We concentrate on the features that are relevant to the following sections. Many details are omitted and can be found in the 802.11 standard [2].

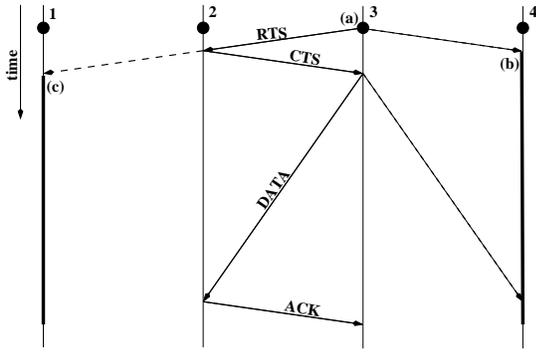


Fig. 1. A typical exchange between Node 3 and Node 2. At point (a) the backoff counter of Node 3 reaches zero, Node 3 senses the medium and finding the medium idle transmits a RTS. Upon receiving the RTS (b) (respectively, the CTS (c)) Node 4 (resp. Node 1) sets its NAV to cover the duration of the exchange between Node 3 and Node 2.

In the 802.11 protocol, a node intending to transmit first senses the medium. Physical and virtual carrier sensing mechanisms are used to determine the state of the medium. The physical carrier sensing is provided by the physical layer. The virtual carrier sensing is done at the MAC layer: it declares the medium busy if the so-called network allocation vector (NAV) value is larger than the current time (or equivalently if the corresponding counter has a non zero value). If both carrier sensing mechanisms declare the medium idle, the node transmits, otherwise it backs-off.

In the 802.11 protocol each node maintains a backoff counter. When a node backs off, it sets its backoff counter to a value selected uniformly in $\{0, 1, \dots, cw\} \times 20\mu s$, where cw is the contention window. cw is initially equal to $CW_{min} = 31$, and is doubled every time a packet sent by the node experiences a collision, up to the limit $CW_{max} = 1023$; it is reset to CW_{min} after a successful transmission by the node. Each node decrements its backoff counter if and only if it senses the medium to be idle. If the medium becomes busy, the backoff counter is frozen and its decrementing resumes only once the medium is idle again. A node may attempt to transmit when its backoff counter is zero.

Figure 1 illustrates a typical exchange in the 802.11 protocol. An exchange between a source s and a destination d consists of a request-to-send (RTS) packet from s to d , a clear-to-send (CTS) packet from d to s , a DATA packet from s to d and finally an acknowledgment (ACK) from d to s . The RTS and CTS packets are control packets that contain the time at which the exchange between s and d will finish. This information is used by nodes overhearing these packets to update their NAV value. At the end of the exchange the source s backs-off.

B. Implementation

We consider the implementation of the 802.11 protocol under the widely used ns-2 simulator [3]. The main reasons for choosing this specific implementation are twofold: (i) The code is available, contrary to the code contained in the

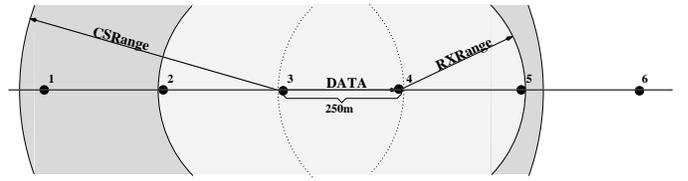


Fig. 2. Ns-2 default parameters: $P = 0.2818W$, ($RXRange = 250m$, $RXThresh = 3.652 \cdot 10^{-10}W$), ($CSRange = 550m$, $CSThresh = 1.559 \cdot 10^{-11}W$).

firmware of the 802.11 card on the market and (ii) It follows closely the 802.11 standard¹.

1) *Propagation model*: Nodes emit at a fixed power P . The power received at a distance l from the source is proportional to P/l^α . In the TwoRayGround propagation model of ns-2, $\alpha = 2$ for small values of l , and $\alpha = 4$ for larger values of l . Given the propagation model, the notions of range and power threshold are interchangeable: If the power received at the destination is larger than a threshold T it means that the distance l between the source and the destination is smaller than $\sqrt[3]{P/T}$, and vice-versa.

2) *Ranges and corresponding thresholds*: The ns-2 physical layer model is described by two ranges (in meters), or equivalently by two power thresholds (in watts):

- The Receiving Range (RXRange) (respectively, Threshold (RXThresh)) is the maximum distance from the source (resp., minimum power at the destination) at which a packet can be successfully received in the absence of interfering nodes.
- The Carrier Sensing Range (CSRange) (resp., Threshold (CSThresh)) is the maximum distance (respectively, minimum power) at which a transmitter can be detected. The CSRange is always larger than the RXRange.

All packets received at a power larger than $CSThresh$ are forwarded to the MAC layer. However, only packets received at a power larger than $RXThresh$ can be successfully processed. Collisions and capture phenomena are also detected at the MAC layer. Assume that node d is currently processing a packet from node s , and that a new packet (from node i) arrives. Denote by P_{sd} (respectively, P_{id}) the power received at node d from node s (resp., node i). If $P_{sd}/P_{id} < CPTresh$, there is a collision at node d , but if $P_{sd}/P_{id} \geq CPTresh$, node d has “captured” the channel, and can continue to process the packet from s . In ns-2 the capture threshold $CPTresh$ is set to 10. The ns-2 simulator does not support cumulative interference.

Figure 2 shows the default value of the different ranges in ns-2. Nodes in the RXRanges of the sender and the destina-

¹We noticed two differences between the 802.11 standard and the ns-2 implementation. The first difference is minor, in ns-2 the contention window is selected in $[0, cw - 1]$ while according to the standard it should be selected in $[0, cw]$. The second difference concerns the exponential increase of the contention window. According to the standard, the contention window is doubled solely before a retransmission and not when a node finds the medium busy on a transmission attempt. We modified the ns-2 implementation to meet the standard requirements before our simulations.

tion are silenced by their virtual carrier sensing mechanism after the initial RTS-CTS exchange. In addition, nodes in the CSRange of the sender are prevented to send by their physical carrier sensing mechanism. This setting offers a double protection to the data packet, but may result in wasted bandwidth. More precisely, all potential interferers (Node 5 in Fig. 2) are silenced by both physical and virtual carrier sensing mechanisms. The two options that might save bandwidth are thus: (1) to reduce the CSRange to allow more concurrent transmissions (Node 1 can now transmit concurrently) (2) to suppress the RTS-CTS exchange (i.e., remove the control message overhead) and to use only the physical carrier sensing mechanism. Option (2) is sometimes called the basic access procedure of the 802.11 protocol. Both options increase the probability of collisions on data packets and the effective gain in bandwidth has to be assessed.

C. Limits Imposed by the Physical Layer

Before running simulations, it is important to define precisely the performance metrics and to be aware of the performance limits imposed by the simulation setting. Two performance metrics are used throughout this paper.

1) *Spatial Reuse* (σ): We define the *spatial reuse* as the number of successful simultaneous transmissions per spatial unit. A spatial unit in a 1D (respectively, 2D) network topology is in meters (resp., square meters). In the network topology of Figure 2, 250m (i.e., the distance between two neighboring nodes) is a natural choice for the spatial unit. The (average) spatial reuse σ is then

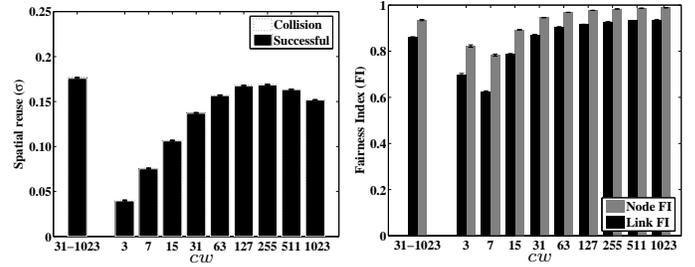
$$\sigma = r \frac{x}{S}$$

where x is the number of successful transmissions, S is the number of spatial units covered by the network topology and r is the ratio of the transmission time and the simulation time.

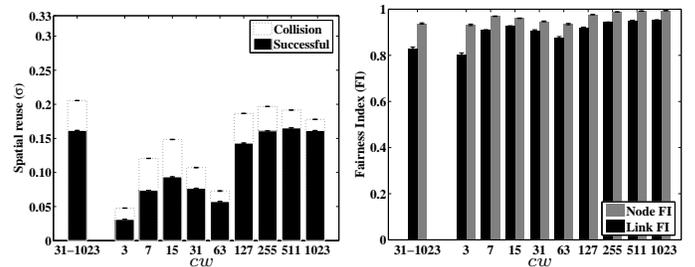
The *maximal spatial reuse* depends on two physical layer parameters, the CSRange and the RXRange. On the line topology of Figure 2 where the CSRange is set to 550m and the RXRange to 250m, at most one link out of four can be active simultaneously, hence the maximal spatial reuse is 1/4. However, if one reduces the carrier sensing range to the minimal value of 250m (i.e., the RXRange) one link out of three can be active simultaneously and the maximum spatial reuse is 1/3. In general, finding the maximum spatial reuse that can be achieved in a network is NP-Complete (by equivalence with the maximal independent set problem [4]) and can only be solved for network topologies that are small or have specific symmetries (such as the line topology).

2) *Fairness Index* (FI): To assess the MAC layer fairness of the 802.11 protocol we use the *Jain's Fairness Index* (FI) [5]. Consider a network with n nodes and denote by (x_1, x_2, \dots, x_n) the number of packets sent per node under a given medium access control protocol. The (node) fairness index of the protocol is

$$FI = \frac{(\sum_i x_i)^2}{n \sum_i x_i^2}.$$



(a) CSRange = 550m



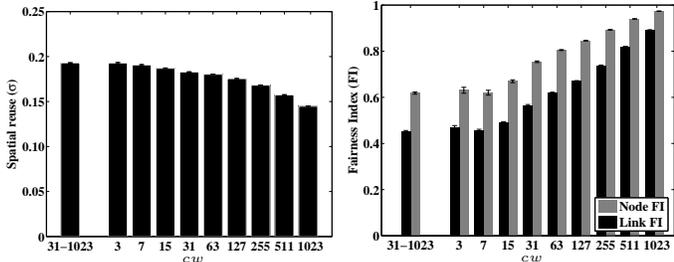
(b) CSRange = 250m

Fig. 3. Performance of the 802.11 protocol with RTS-CTS handshake. The left-most bar of each figure corresponds to the variable contention window case (i.e., $CW_{min} = 31$ and $CW_{max} = 1023$), the other bars correspond to fixed contention window values (i.e., $cw = CW_{min} = CW_{max}$). In the figures representing the spatial reuse, the dashed line extension of each bar corresponds to the additional spatial reuse that would be achieved if all data packets were received correctly. At CSRange=550 there is barely any collision (on DATA packets); the dashed lines coincide with the bars.

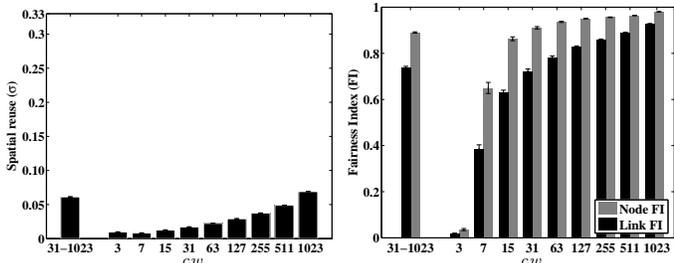
The *maximum fairness index* is 1. It corresponds to a network where all nodes access the channel equally. Replacing 'node' by 'link' in the previous definition, we obtain the link fairness index of the protocol.

D. Performance Overview

The numerical results presented throughout this paper are obtained using the ns-2 implementation [3] of the 802.11 protocol on a line network topology of 50 nodes where two neighboring nodes are at a distance of 250m. Our simulations use the default ns-2 parameters (Fig. 2). However, to make the results complete, we also consider a less conservative value of the CSRange (i.e., CSRange = 250m instead of 550m). We choose this simple topology for several reasons: (i) to render the limitations of the ns-2 physical layer negligible (in Fig. 2 only Node 5 can create a collision at Node 4; if Node 5 does not emit, even the combined interference of all other nodes is not enough to create a collision at Node 4), (ii) to enable an exhaustive study with respect to the CSRange parameter (the behavior of the protocol is the same for all CSRange in [250, 500) and for all CSRange in [500, 750) and setting $CSRange \geq 750$ simply scales down the spatial reuse of the protocol), (iii) to pick a topology where the performance



(a) CSRange = 550m



(b) CSRange = 250m

Fig. 4. Performance of the 802.11 protocol using the basic access procedure.

limits are known (Section II-C) and thus make it easier to explain and quantify the problems of the 802.11 protocol. However, it is important to emphasize that the problems of the 802.11 protocol highlighted in this paper are not specific to this topology. The same problems are observed in 2D-topologies as demonstrated in Section VI.

All simulations assume saturated traffic conditions (i.e., nodes have always a packet to send and successively transmit to each of their one-hop neighbors). The packet size is 1500bytes. To provide accurate results, each simulation is repeated 50 times (using different random seeds). The simulations run for 50s. The presented results consist of the values averaged over the 50 experiments, and the 95% confidence intervals. All figures follow the same format, displaying the performance metric as a function of the contention window, cw .

Figure 3 shows the performance of the 802.11 protocol with RTS-CTS handshake. The spatial reuse, σ , achieved by the 802.11 protocol is at most equal to half the optimal spatial reuse of $1/3$. The protocol performs slightly better with a CSRange=550m as all the packets sent are received correctly. For a CSRange=250m the number of packets sent is larger, but the spatial reuse is reduced by collisions. In both cases, the highest level of spatial reuse is reached for the scenario where the contention window size is variable. However, a carefully selected (fixed) contention window size can result in similar levels of spatial reuse and a better fairness.

Figure 4 shows the performance of the 802.11 protocol

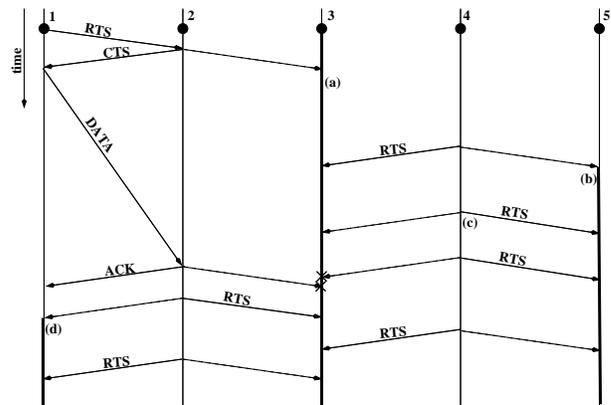


Fig. 5. The gaged node situation: (a) The NAV of Node 3 is set until the end of the exchange between Nodes 1 and 2. (b) Node 4 sends a RTS to Node 3. Node 3 does not reply with a CTS because of its NAV, however Node 5 sets its NAV to protect a potential transmission between Nodes 4 and 3. (c) Node 4 is now the only node that can attempt to transmit and does so by sending repeated RTS to Nodes 3 and 5. As a consequence at the end of the exchange between Nodes 1 and 2, Nodes 3 and 5 have been gaged by the repeated RTS of Node 4. (d) In addition, the RTS from Node 2 to Node 3 silences Node 1. At this point no transmission is possible in the network and the repeated RTS of Nodes 2 and 4 can maintain Nodes 1, 3 and 5 in a gaged situation for a long time.

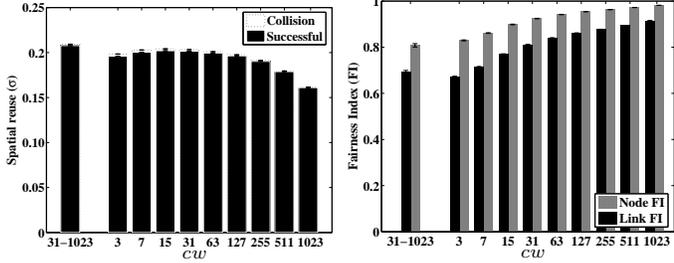
using the basic access procedure (without RTS-CTS handshake). At a CSRange of 550m the transmitter can use its carrier sensing mechanism to determine the state (idle, receiving or sending) of the receiver. As a result, in terms of spatial reuse, the basic procedure performs as well (at high cw) or even better (at low cw) than the access procedure with RTS-CTS. Indeed, at low cw the basic access procedure benefits from its reduced overhead and achieves a spatial reuse 23% away from the optimal spatial reuse of $1/4$. At a CSRange = 250m the transmitter cannot infer the state of the receiver, hence the number of collisions on DATA packets is high and the spatial reuse is poor.

III. IEEE 802.11: DOES IT PUT THE CORRECT STATE IN THE NETWORK?

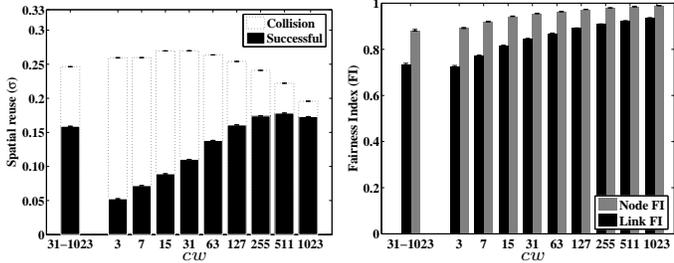
The emphasis of this section is on identifying awkward situations that happen under normal operation of the 802.11 protocol with RTS-CTS. We describe precisely three such situations: the *gaged* node situation, the *jammed* node situation and the *focused* node situation. To each situation we associate a remedy and implement it in [6]. The remedies proposed might not be unique and often consist of minimum effort solutions. However, they are helpful in quantifying the impact of each of these situations on the performance of the 802.11 protocol. The remission of the gaged, jammed, and focused node situation brings the 802.11 protocol performance very close to optimum.

A. The Gaged Node Situation

1) *Observation*: The number of DATA packets sent by the 802.11 protocol is too low to reach a high level of spatial



(a) CSRange = 550m



(b) CSRange = 250m

Fig. 6. Performance of the 802.11 protocol with a reduced NAV on RTS. Upon receiving a RTS packet not for it, a node sets its NAV to 'the current time + a CTS transmission time', and not to 'the current time + CTS + DATA + ACK transmission times'

reuse. In addition, it is hard to provide a logical explanation for the levels of spatial reuse observed at a CSRange of 250m (Figure 3(b)). Why should the levels of spatial reuse be lower at $cw = 31$ and $cw = 63$ than at $cw = 15$ and $cw = 127$?

2) *Problem*: A node is gagged (silenced) by repeated RTS (intended to other nodes) that are not followed by a data packet transmission. The gagged node situation arises when a node sets its NAV to protect a DATA packet that is never sent. In this situation, the virtual carrier sensing that is based on the NAV value does not reflect the real state of the channel anymore. Figure 5 shows an example of a gagged node situation. The gagged node situation is especially likely when the number of RTS received correctly is high, as nodes need to decode successfully the RTS messages to update their NAV value.

3) *Remedy*: The 802.11 standard (p.79) mentions that 'A station that used information from an RTS frame as the most recent basis to update its NAV setting is permitted to reset its NAV if no PHY-RXSTART.indication is detected from the PHY during a period' slightly larger than the time needed to send a CTS. However, this recommendation is not currently implemented. A possible remedy is thus to implement the NAV reset as proposed by the 802.11 standard; another, simpler remedy requiring less modification in the code is to modify the NAV on RTS to cover only the CTS exchange. We refer to the first solution as the *reset NAV* solution and to the second

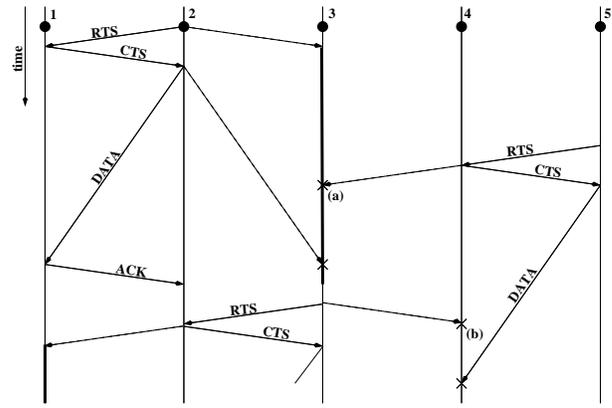


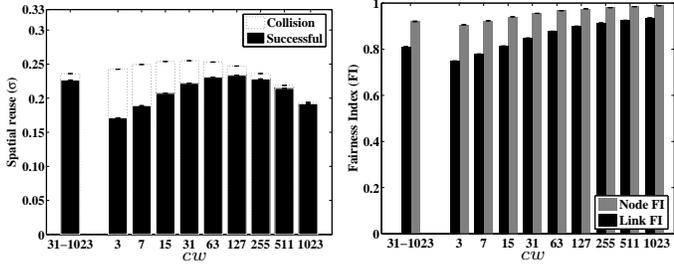
Fig. 7. The jammed node situation: (a) Node 3 is jammed by the reception of two simultaneous packets (the DATA packet from Node 2 to Node 1 and the CTS packet from Node 4 to Node 5). It cannot update its NAV based on the latest information contained in the CTS. (b) At the end of the exchange between Nodes 1 and 2, Node 3 is not aware of any ongoing transmission, it sends a RTS that collides with the DATA packet sent from Node 5 to Node 4.

as the *reduced NAV* solution. We implemented both solutions [6] and found that in terms of performance they are similar, although the reduced NAV solution performs slightly better than the reset NAV solution at a CSRange of 550m, whereas the opposite is true at a CSRange of 250m. Figure 6 shows the performance of the 802.11 protocol using the reduced NAV on RTS solution. Compared to the current implementation of the 802.11 protocol, the number of RTS sent drops as fewer RTS need to be retransmitted (e.g., at $cw = 63$ the number of RTS sent is divided by 3). Due to the absence of gagged nodes, a successful RTS is now more likely to be followed by a successful CTS and a DATA packet. As a result, the abnormal behavior of the actual 802.11 protocol at a CSRange of 250m disappears completely. The spatial reuse achieved by the modified 802.11 varies more smoothly and is higher than the spatial reuse of the actual 802.11 protocol with RTS-CTS. The fairness of the modified protocol is only slightly lower than the fairness of the actual 802.11 protocol with RTS-CTS. The comparison of Figures 4 and 6 shows that solving the gagged node problem is enough to bring the performance of the 802.11 protocol with RTS-CTS above the performance of the basic access procedure and this both in terms of spatial reuse and fairness.

B. The Jammed Node Situation

1) *Observation*: At low CSRange the number of concurrent transmissions is higher but the benefit is reduced due to collisions on DATA packets. Indeed, from Figure 6(b), it is obvious that if we could limit the collisions on data packets, the 802.11 protocol would achieve a significantly higher spatial reuse.

2) *Problem*: A node is jammed by a data packet (initiated and terminated at two other nodes). As a consequence, it cannot extract information from control messages sent concurrently and cannot update its network allocation vector. This loss of information or of state gives it a wrong view

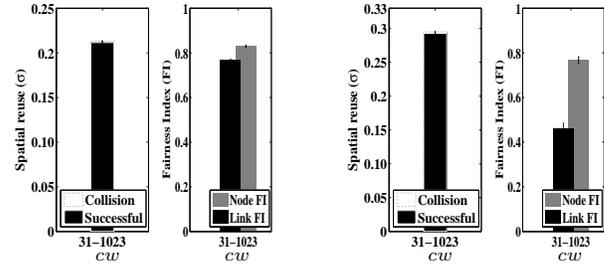


(a) CSRange = 250m

Fig. 8. Performance of the 802.11 protocol with a reduced NAV on RTS and an additional control channel.

of the network. In particular, such a node cannot rely on its virtual carrier sensing mechanism to determine if the medium is busy, and is likely to create a collision the next time it transmits. Figure 7 shows a typical instance of the jammed node situation. The jammed node situation occurs every time a node cannot decode the content of a CTS that is followed by a DATA packet. A node can be jammed by a data packet or by a control message. However, given that control messages are much shorter than data packets, the first occurrence is more common.

3) *Remedy*: It is difficult to suppress collisions between control messages as they are part of the contention resolution mechanism. Moreover, a node is more often jammed by DATA packets as they are typically much longer than control messages. A solution to the jammed node problem is thus to create a separate channel for control messages. In fact, the 802.11 protocol has 3 (802.11b) to 12 (802.11a) non-overlapping channels. Typically only one of these channels is used in a multi-hop ad hoc network. It is thus possible to add a control channel, basically 'for free'. Several works already report that the addition of a control channel increases the throughput of the 802.11 protocol (see for example the recent works of [7], [8]) but they do not identify the jammed node situation. Again, we selected the implementation that requires the minimum change in the functioning of the 802.11 protocol. Each node has only one transceiver and a node switches to the data channel only to transmit or receive DATA and ACK packets. The protocol is otherwise unchanged. In Figure 7, after the RTS/CTS handshake between Node 1 and Node 2, only these two nodes switch to the DATA channel. Consequently, Node 3 which is listening to the control channel would be able to decode the CTS sent by Node 4, update its NAV, and the collision at Node 4 would be avoided. Figure 8 reports the performance of the 802.11 protocol with a data and a control channel. At a CSRange of 550m it is not necessary to add a control channel to overcome the jammed node situation as nodes can use their physical carrier sensing to recover the state of the channel. Figure 6(a) confirms that at a CSRange of 550m the DATA packets do not suffer from collisions.



(a) CSRange = 550m

(b) CSRange = 250m (we use an additional control channel)

Fig. 9. Performance of the 802.11 protocol with a reduced NAV on RTS and a backoff counter per link.

On the contrary, the performance of the 802.11 protocol at a CSRange of 250m greatly benefits from the addition of a control channel.

C. The Focused Node Situation

1) *Observation*: It is somewhat surprising that despite a relatively low fairness and an additional degree of freedom, the 802.11 protocol using a variable contention window is unable to achieve a higher spatial reuse than its fixed contention window counterpart.

2) *Problem*: In the current implementation of the 802.11 protocol, each node has a backoff counter and a contention window value. The contention window is doubled each time the node needs to retransmit a packet and reset to its minimum value on a successful packet transmission. However, a node has typically several neighbors and can send traffic on each of the corresponding links. Consider a node with two links, one with a high level of collisions and one with a low level of collisions. Assume that the node attempts to transmit on the link with the high level of collisions. After each packet collision, the node doubles its contention window, waits for its backoff counter to reach zero and retransmits. Consequently, the node attempts less and less frequently to access the channel even though it might be able to send successfully on its other outgoing link. The node implicitly assumes that all its outgoing links experience the same contention as the link where it is currently trying to transmit. In other words, a backoff counter per node maintains the node focused on a single neighbor and it cannot take advantage of the link diversity.

3) *Remedy*: We propose to use a backoff counter per link instead of per node. Each link has its own contention window. The rules that regulate the decrease of the backoff counter are otherwise unchanged. A backoff counter per link introduces contention among the different links of the same node. In the previous situation, the retransmissions on the link with a high level of collisions would alternate with successful transmissions on the good quality link. Figure 9 shows the performance of the 802.11 protocol with a per

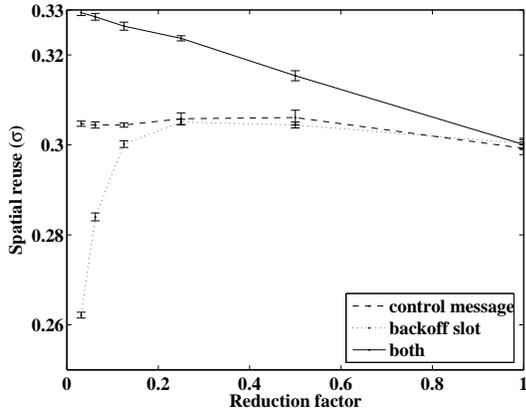


Fig. 10. Spatial reuse of the modified 802.11 protocol (variable cw , CSRange = 250m) as a function of the overhead size.

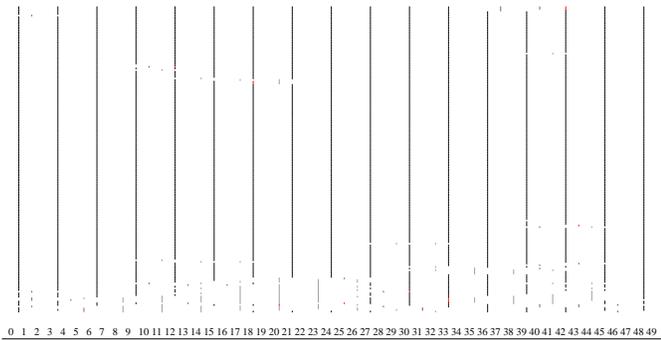


Fig. 11. Activity of the different links over a period of 5s at a CSRange of 250m. The x axis is the spatial coordinate of the node, the y axis is the time. A transmission between node i and j from time t_1 to time t_2 is depicted by a vertical line from $(\frac{x_i+x_j}{2}, t_1)$ to $(\frac{x_i+x_j}{2}, t_2)$.

link backoff counter and a variable contention window. At a CSRange of 550m, a per link backoff does not change the spatial reuse of the protocol, however it improves its fairness. At a CSRange of 250m, the increase in spatial reuse is spectacular and the performance of the 802.11 protocol is now only 10% away from the optimal spatial reuse of one-third. However, the fairness of the protocol is severely reduced. The 802.11 protocol can thus use the additional degree of freedom granted by a variable contention window to trade-off fairness for spatial reuse.

IV. IEEE 802.11: CLOSING THE GAP

The gagged, jammed and focused node situations demonstrate how important it is to maintain an accurate network state at each node and show how a wrong perception of the channel state can affect the performance of the 802.11 protocol. At a CSRange of 250m, the suppression of these three situations brings the performance of the 802.11 protocol with RTS-CTS within 10% of the optimal spatial reuse. Although such a level of performance might be satisfactory in practice it is important to understand what prevents our modified 802.11

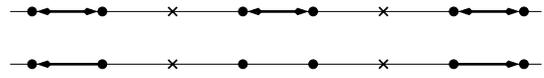


Fig. 12. On the top of the figure the schedule of maximal spatial reuse on a small topology. The nodes represented by a cross sense the medium as busy (virtual or physical carrier sensing) and therefore must remain silent. When the the middle transmission finishes, only a transmission on the same link is possible (bottom).

protocol to operate at the limits set by its physical layer. In this last section, by lack of space, we restrict our study to the CSRange of 250m as it brings the highest gain in terms of spatial reuse.

A. Overhead of the Control Messages and Backoff Mechanism

Figure 10 shows the spatial reuse achieved by our modified 802.11 protocol in the variable contention window case as a function of a reduction factor. The reduction factor is applied to different sources of overhead. A reduction factor of 1 corresponds to the overhead of the actual 802.11 protocol, a reduction factor of 1/2 corresponds to a reduction of the actual overhead by a factor 2, etc.

Not surprisingly, the overhead of the RTS/CTS handshake is responsible for some of the wasted bandwidth. However, the dashed curves of Figure 10 show that even if one could reduce the time spent to send the control messages to some negligible amount, it would still remain impossible to achieve the optimal spatial reuse. We observe a gain of performance of at most 5%.

Another cause of wasted bandwidth is the time spent in the backoff process. The actual 802.11 protocol uses a backoff slot of $20\mu s$. One way to reduce the backoff time is thus to reduce the length of the backoff slot. The dotted curves of Figure 10 show that reducing the value of the backoff time has limited benefits on the spatial reuse of the protocol. In particular, at low values of the backoff slot we observe a poor spatial reuse, which is due to an increase of the collisions on data packets.

B. Asymptotic Behavior as the Overhead Tends to Zero

We now reduce the control message overhead and the backoff time simultaneously. Reducing both overhead factors simultaneously makes it possible to operate at lower backoff times and thus to increase the number of transmission attempts without increasing the number of collisions on data packets. Figure 10 (plain curve) shows that the performance of the 802.11 protocol tends to the fundamental limit of 1/3 and ultimately achieves it. Indeed, Figure 11 demonstrates that, after a short initialization phase, the protocol spends most of the time in the single transmission schedule of maximal spatial reuse. The explanation is straightforward. At small backoff time, an ending data transmission is almost immediately replaced by a new one as the nodes spend almost no time in the backoff process. It is thus unlikely that another transmission ends before the completed transmission is replaced. Figure 12 confirms that once the schedule of maximal spatial reuse is reached, an ending transmission can only be replaced by a

transmission on the same network link. Consequently, the schedule of maximal spatial reuse is guaranteed to persist. The 802.11 protocol organizes, in a completely decentralized manner, the transmissions towards schedules of maximal spatial reuse. Unfortunately, in practice the high number of collisions destroy this organization.

In general, at low backoff time, the spatial reuse is high (provided most of the collisions are avoided), but the fairness of the protocol is poor. However, as we increase progressively the time spent in the backoff process, the level of organization in the network decreases, the fairness increases and the spatial reuse decreases. By operating at the appropriate backoff value it is thus possible to trade-off spatial reuse for fairness.

V. RELATED WORK

In this section we review our findings and position them with respect to the abundant related work.

The 802.11 protocol is probably the most well-known decentralized MAC protocol. It is also one of the few that is actually implemented in industrial product. Nevertheless, detailed experimental studies of the 802.11 protocol reveal poor performance in multi-hop ad hoc networks [9], especially at high traffic load.

The first category of works on the 802.11 protocol identifies situations where its contention resolution mechanism is vulnerable or unfair. The hidden node situation [10] is probably the most (in)famous and the most studied of these situations. In Figure 2 (with a CSRange=250m), Node 5 is 'hidden' from Node 3 as it cannot detect its transmission. The hidden node situation is partially, but not completely [11], addressed by the use of the RTS/CTS handshake before the DATA packet transmission. The exposed node situation is often mentioned as the counterpart of the hidden node situation. In Figure 2 (with a CSRange=550m), Node 1 is 'exposed' to Node 3 transmission as it could transmit a packet without interfering with the reception at Node 4. The exposed node problem disappears at a CSRange of 250m. The jammed and the gagged node situations were only recently studied in the work of Ray et al. [12], [13] and deserve more attention. We propose an alternative view of these two situations, which complements and extends this work. In particular, we use a ns-2 implementation of the 802.11 protocol (instead of a Matlab implementation) which allows us to (i) give more complete simulation results, (ii) investigate the effect of additional parameters such as the carrier sensing range and the contention window, and (iii) implement the remedies proposed in Section III. Many of the aforementioned situations result in short or long-term unfairness as demonstrated by [14].

The second category of works on the 802.11 protocol modifies the protocol to improve its fairness or its throughput. Multi-Channel proposals [15], [16], [7], [8] fall in this category. The proposals differ in the number of transceivers needed and the number of channels supported, but almost all proposals target an increase in the 802.11 protocol throughput. These proposals are in general much more complex (but also

maybe more efficient) than the one presented in this paper, whose only purpose was to reduce the number of collisions on data packets. In addition, several papers study the fairness of the 802.11 protocol. We mention a few that relate directly to the focused node situation. Indeed, the use of a backoff counter per link has first been proposed in the context of the MACAW protocol to provide per link fairness in wireless LAN (single-hop networks). More recently, [17] proposes a per (multi-hop) flow scaling of the contention window to achieve a better flow fairness. The disadvantage of this proposal is that it requires the destination of the multi-hop flow to propagate appropriate state information to all intermediate nodes. Finally, the 802.11e draft standard relies on a backoff counter per traffic class to provide quality of service in the 802.11 framework. In this paper, we show that a per-link backoff solves a specific problem, namely the focused node situation and results in a gain in the spatial reuse of the protocol. However, it is not clear that a per-link backoff solves the fairness problem of the 802.11 protocol in the multi-hop setting.

The third category of works on the 802.11 protocol investigates the effect of different parameters of the protocol on its performance. Several works study the effect of the carrier sensing range [18], [19], [20] and the effect of the contention window size [21], [22] on the spatial reuse of the 802.11 protocol and argue that a suboptimal choice of these parameters can severely impacts the spatial reuse of the protocol.

Our work is unique in the sense that it belongs to these three categories simultaneously and identifies all the steps needed to bring the performance of the 802.11 protocol to the theoretical limits. Finally, it appears that the asymptotic behavior of the 802.11 protocol presented in this paper reinforces the theoretical predictions of [23], [24].

VI. CONCLUSION

This paper proposes a global view of the 802.11 protocol. We use a systematic approach where we start from the actual implementation of the protocol, highlight an unwanted behavior of the protocol, describe its origin, and propose a solution. The remedies proposed in this paper might not be unique, but they are essential to show that the resulting protocol can achieve the maximal level of spatial reuse.

Table I summarizes the problems identified and the variation of performance that can be attributed to each of them. To convince the reader that these problems are not specific to the line topology, Table I also presents the same results on a random 2D topology. In both topologies the resolution of the gagged, jammed, and focused node situations increases significantly the spatial reuse of the 802.11 protocol. To further increase the spatial reuse and reach the limit imposed by the physical layer, one has to reduce the overhead (backoff and control messages) to some negligible quantity. In the line topology, there is a clear trade-off between fairness and spatial reuse. In the random topology, most of the unfairness is due to the variation in node degree and is inherent to the

	Line topology (50 nodes)					Random 2D topology (100 nodes)				
	variable cw		fixed $cw(=x)$			variable cw		fixed $cw(=x)$		
	σ	FI (node/link)	σ	FI	x	σ	FI	σ	FI	x
Current implementation	0.16	0.94 / 0.83	0.16	0.99 / 0.95	511	0.08	0.41 / 0.24	0.06	0.70 / 0.46	127
Gagged node solved	0.16	0.88 / 0.73	0.17	0.98 / 0.92	511	0.08	0.40 / 0.23	0.08	0.57 / 0.35	31
Jammed node solved	0.22	0.92 / 0.81	0.23	0.98 / 0.90	127	0.10	0.58 / 0.34	0.10	0.67 / 0.39	31
Focused node solved	0.29	0.76 / 0.45	0.24	0.96 / 0.87	127	0.12	0.55 / 0.23	0.11	0.62 / 0.30	31
Reduced overhead	0.33	0.72 / 0.39	0.32	0.78 / 0.47	7	0.14	0.55 / 0.22	0.14	0.60 / 0.27	7

TABLE I

802.11 PERFORMANCE FOR THE VARIABLE CONTENTION WINDOW CASE (VARIABLE cw) AND FOR THE FIXED CONTENTION WINDOW x (FIXED $cw = x$) THAT GIVES THE HIGHEST SPATIAL REUSE AT A CSRANGE OF 445M. WE CONSIDER TWO TOPOLOGIES, A LINE AND A RANDOM 2D TOPOLOGY (125 NODES ARE RANDOMLY DEPLOYED ON A 2500X2500 SQUARE AREA, THE ISOLATED NODES ARE THEN REMOVED TO KEEP A CONNECTED COMPONENT OF 100 NODES, THE AVERAGE NODE DEGREE IS 3.5 AND VARIES BETWEEN 1 AND 8). IN THE 2D TOPOLOGY, A CSRANGE OF 250M WOULD RESULT IN A VERY LOW NODE DEGREE AND A POOR CONNECTIVITY. THIS OBSERVATION, TOGETHER WITH THE FACT THAT A CSRANGE OF 445M DOES NOT CHANGE THE RESULTS FOR THE LINE TOPOLOGY, MOTIVATED THE CHOICE OF A CSRANGE OF 445M FOR THE 2D TOPOLOGY.

network topology. For such a topology it is possible to improve the spatial reuse of the protocol without further reducing its fairness.

All the results presented in this paper were obtained by simulation and although the ns-2 implementation of the IEEE 802.11 MAC protocol is very accurate its physical layer model (wireless channel, interference model, etc.) is quite simplistic. The ultimate goal of this work is thus to be able to reproduce the experiments presented in this paper on a testbed and to infer the impact of the gagged, jammed and focused node situations on a real network.

REFERENCES

- [1] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," in *IEEE Journal On Selected Areas In Communications*, vol. 18, no. 3, 2000.
- [2] *IEEE802.11, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Std., Aug 1999.
- [3] Network Simulator 2 (ns-allinone-2.28). [Online]. Available: <http://www.isi.edu/nsnam/ns/>.
- [4] M. Garey and D. Johnson, "Computers and intractability: A guide to the theory of NP-completeness," in *Freeman*, New York, 1983.
- [5] R. Jain, *The Art of Computer Systems Performance Analysis*. John Wiley and Sons, Inc., 1991.
- [6] Simulator with remedies to the gagged, jammed, and focused node situations. [Online]. Available: <http://icapeople.epfl.ch/mdurvy/research.html>.
- [7] J. So and N. Vaidya, "Multi-Channel MAC for Ad Hoc Networks: Handling Multi-Channel Hidden Terminals Using A Single Transceiver," in *MobiHoc*, 2004.
- [8] A. Raniwala and T.-C. Chiueh, "Architecture and Algorithms for an IEEE 802.11-Based Multi-Channel Wireless Mesh network," in *IEEE INFOCOM*, 2005.
- [9] P. Gupta, R. Gray, and P. Kumar, "An experimental scaling law for ad hoc networks," UIUC, May 2001. [Online]. Available: citeseer.nj.nec.com/gupta01experimental.html
- [10] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II - the hidden terminal problem in carrier sense multiple-access modes and the busy-tone solution," in *IEEE Transactions on Communications*, 1975, p. 23(12):1417-1433.
- [11] K. Xu, M. Gerla, and S. Bae, "How Effective is the IEEE 802.11 RTS/CTS Handshake in Ad Hoc Networks?" in *GLOBECOM*, 2002.
- [12] S. Ray, J. Carruthers, and D. Starobinski, "RTS/CTS-induced congestion in ad hoc wireless lans," in *WCNC*, 2003.
- [13] —, "Evaluation of the masked node problem in ad-hoc wireless LANs," in *IEEE Transactions on Mobile Computing*, (to appear).
- [14] M. Garetto, J. Shi, and E. W. Knightly, "Modeling media access in embedded two-flow topologies of multi-hop wireless networks," in *MobiCom*, 2005.
- [15] A. Nasipuri and S. Das, "Multichannel CSMA with signal power-based channel selection for multihop wireless networks," in *IEEE Vehicular Technology Conference*, 2000.
- [16] N. Jain, S. Das, and A. Nasipuri, "A Multichannel CSMA MAC protocol with Receiver-Based Channel Selection for MultiHop Wireless Networks," in *IEEE IC3N*, 2001.
- [17] K. Sundaresan, H.-Y. Hsieh, and R. Sivakumar, "IEEE 802.11 over Multi-hop Wireless Networks: Problems and New Perspectives," in *Ad Hoc Networks*, vol. 2, 2004, pp. 109-132.
- [18] F. Ye, S. Yi, and B. Sikdar, "Improving Spatial Reuse of IEEE 802.11 Based Ad Hoc Networks," in *IEEE GLOBECOM*, 2003.
- [19] F. Ye and B. Sikdar, "Distance-Aware Virtual Carrier Sensing for Improved Spatial Reuse in Wireless Networks," in *IEEE GLOBECOM*, 2004.
- [20] J. Deng, B. Liang, and P. Varshney, "Tuning the Carrier Sensing Range of IEEE 802.11 MAC," in *IEEE GLOBECOM*, 2004.
- [21] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism," in *IEEE/ACM Trans. on Networking*, 8(6):785-799, December 2000.
- [22] —, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," in *IEEE/ACM Trans. on Networking*, 8(6):785-799, December 2000.
- [23] X. Wang and K. Kar, "Throughput Modelling and Fairness Issues in CSMA/CA Based Ad-Hoc Networks," in *INFOCOM, Miami*, 2005.
- [24] M. Durvy and P. Thiran, "A Packing Approach to Compare Slotted and Non-Slotted Medium Access Control," in *INFOCOM, Barcelona*, 2006.