Dynamic P-Persistent Backoff for Higher Efficiency and Implicit Prioritization

J. Barceló, B. Bellalta, C. Cano, M. Oliver
Dept. of Information and Communication Technologies
Universitat Pompeu Fabra, Passeig de Circumval.lació 8, 08003 Barcelona
Email: {jaume.barcelo,boris.bellalta,cristina.cano,miquel.oliver}@upf.edu

***Abstract*—This article studies the efficiency of backoff algorithms. The fraction of channel time devoted to successful transmissions is maximized when the stations choose the optimal transmission probability. The binary exponential backoff algorithm does not come close to optimal channel efficiency, thus a new backoff mechanism that attains near-optimal efficiency is proposed. This algorithm is called Dynamic-P-Persistent backoff and is based on the observation that, under optimal efficiency conditions, the fraction of channel slots busy with collisions is constant. The stations monitor the channel to estimate the fraction of collision slots and adjust their transmission probabilities consequently. As opposed to previous backoff proposals, DPP does not require any estimation of the number of concurrent active stations. Further, DPP offers implicit prioritization that reduces the delay of real time and interactive traffic while maintaining optimal throughput for background traffic.**

I. INTRODUCTION

Wireless networks build upon the IEEE 802.11 [1] standard and its different flavors are growing and proliferating at universities, enterprises and homes. In each of these networks, the stations and access points share a common channel to transmit data. Being the air a broadcast channel, the participants in the network should avoid to transmit simultaneously. If two participants do transmit at the same time a collision occurs and the data of both senders might be lost. It is the duty of the Medium Access Control (MAC) layer to handle collisions and minimize their impact on performance.

This is not a new problem; it already appeared in early Aloha [2] and Ethernet [3] networks. There are two general techniques that effectively improve the efficiency of this kind of networks. The first one consists on sensing the channel before transmitting (Carrier Sense Multiple Access, CSMA [4]). If the channel is sensed busy, it means that there is an ongoing transmission and the other participants will refrain from transmitting to avoid a collision. Further, limiting the instants at which the participants can begin a new transmission, also reduces the number of collisions. The time is divided in slots and transmissions are allowed only at the beginning of each slot. There is a collision if two or more stations choose the same slot to transmit. To reduce the probability of a collision, it is necessary to randomize the selection of the time slot at which a given station transmits.

In P-persistent protocols, the stations involved in a collision retransmit in the following slot with probability P. With probability $1-P$ the retransmission is postponed for the next slot. This operation repeats until the station finally retransmits. In a more sophisticated backoff algorithm, the stations involved in a collision draw a random number from a contention window (*e.g.* a number between 0 and 31) and then wait for that number of slots before re-attempting transmission. If the random values are selected from a contention window that doubles after each failed attempt, the mechanism is called Binary Exponential Backoff (BEB). A variant of this scheme called Truncated BEB (T-BEB) is the contention algorithm of choice for IEEE 802.11 networks.

IEEE 802.11 medium access comes in two different flavors. The most simple (Basic Access) consists on a two-way handshake in which the sender transmits a packet and waits for the receiver to explicitly acknowledge the correct reception with a short packet. When a collision occurs, a considerable amount of time is wasted since the senders cannot detect the collision while they are transmitting. This implies that the senders will not immediately interrupt transmission when a collision occurs. Conversely, the transmitters will send the whole packet and will only realize that a collision has happened because of the lack of acknowledgement.

To prevent collisions, RTS/CTS can be used. It is a more elaborated four-way handshaking mechanism in which the sender requests permission to send (Request-To-Send) and the receiver grants the permission (Clear-To-Send) effectively reserving the channel for the duration of the transmission and acknowledgement. This approach also solves the hidden terminal problem. The hidden terminal problem occurs when two terminals that can not hear to each other have a packet ready to transmit. If this is the case, the carrier sense mechanism will not work and both stations will transmit simultaneously. The problem arises when the receiver is in the hearing range of both transmitting stations and the collision occurs.

Due to the additional control messages, RTS/CTS access places an additional overhead on the channel that penalizes performance. For this reason, the rest of the article focuses on the Basic Access two-way handshaking mechanism. To simplify the analysis, it is considered that all the participating stations share a common broadcast channel, and each station can hear the transmissions of all the other stations.

After this first introductory section, the remaining of the paper is organized as follows. Sec. II reviews previous art and highlights the contribution of this paper. Sec. III describes T-BEB and proposes a general framework to assess the efficiency of backoff mechanisms in general. This framework is used to derive the optimum efficiency, which can be used as a benchmark to compare backoff schemes. It is observed that the maximum efficiency is a function of both the packet length and the number of contending stations. Further, it can be concluded that T-BEB performs less-than optimal in most of the cases. The finding that the fraction of collision slots is constant when optimal transmission probability is used is crucial to derive a near-optimal backoff algorithm.

Sec. IV introduces Dynamic-P-Persistent (DPP) backoff protocol. It is a variant of P-Persistent backoff that constantly monitors the number of collision slots and adjusts the transmission probability to attain optimal collision probability. Since the collission probability is independent of the number of active stations, this proposal delivers near-optimal performance for any number of competing stations. It is noticeable that the estimation of the number of backlogged stations is not required.

Sec. V presents simulations results to support the analysis of the previous sections. A first simulation shows how the stations adjust their transmission probability as the number of stations varies. This simulation offers an intuitive understanding of the behaviour of the mechanism in a dynamic environment. Then, extensive simulations assess the efficiency of DPP and show how close it is to the upper bound obtained in Sec. III.

The proposed backoff scheme comes with advantageous implicit prioritizing features that are explored in Sec. VI. DPP benefit stations that generate real-time and interactive traffic and penalizes those that are permanently active sending background traffic.

Finally, Sec. VII summarizes the paper and provides some concluding remarks.

II. RELATED WORK

The Truncated Binary Exponential Backoff is a protocol to control multiple-access broadcast channels. It is a distributed access mechanism in the sense that each station independently executes the algorithm to decide whether to transmit or not in a given time slot. Each station selects a number from a contention window and waits for that number of slots before attempting transmission. The contention window doubles after each failed transmission attempt and resets to its minimum value after a successful transmission. It is called Truncated, because when reaching a maximum backoff stage (m) the contention window does not double any more. Additionally, a packet is dropped after reaching the maximum number of retransmission attempts (R). The properties of BEB and T-BEB have been extensively studied in [5]–[7] to cite a few.

CSMA and T-BEB are widely used in WLAN since they are at the core of the Distributed Coordinated Function (DCF) defined in IEEE 802.11. Any improvement in the backoff mechanisms would traduce in increased performance of the ubiquitous WiFi networks. Moreover, CSMA and T-BEB also appear as an ingredient of many MAC layer proposals supporting upcoming networks such as (Mobile) Ad-Hoc Networks [8], Sensor Networks and Personal Area Networks [9].

The studies are performed under saturation conditions, *i.e.* each station has always a packet to transmit. This is the maximum load that can be offered to the network and it is assumed that it is the maximum strain to which the network may be exposed. The properties of interest include fairness (both short-term and long-term), stability and efficiency. In this paper the focus is placed on efficiency (the fraction of channel time devoted to successful transmissions). Given a data rate, this metric can be translated to throughput which is widely used in the literature.

The backoff protocols put the stations on hold thus diminishing the chances that a station attempts transmission in any given slot. The backoff effectively influences the frequency with which stations transmit. Another way to interpret the effect of the backoff is to understand that it tunes the transmission probability.

In [10], it was already stated that the optimal transmission probability is a function of the packet length (l) and the number of competing stations (n). A p-persistent backoff mechanism was also suggested to study the behaviour of T-BEB. The maximum efficiency of T-BEB was estimated by minimizing the average virtual transmission time. Similarly to our work, an algorithm to tune the transmission probability to improve the efficiency was proposed. The main difference resides in that the estimation of the number of competing stations is not required in our algorithm.

Previous efforts focused on inferring the number of stations from the number of empty, busy and collision slots. Specifically, [11] shows that the number of active stations can be expressed as a function of the collision probability encountered on the channel. Additionally, it proposes an extended Kalman filter coupled with a change detection mechanisms to estimate the number of contending stations *n*. A notable advancement was presented in [12] in which a bayesian approach was adopted to estimate the number of competing terminals.

Other works [13] assume that the number of contending stations is known (either using one of the estimation techniques cited above or assuming that the information is directly available at the AP) and then compute the optimal – fixed – contention window. A fixed (as opposed to T-BEB's exponentially-growing) optimal contention window increases performance both in terms of efficiency and fairness.

Another line of research consists on cross-layer techniques that combine BEB, Tree Algorithms [14] , and successive interference cancellation [15]. However, these studies maximize the number of successful slots while neglecting the fact that empty slots are much shorter than collision slots. In Sec. III it is explained that the different duration of the slots is of paramount importance in computing channel efficiency.

Finally, there is a game-theoretical approach presented in [16]. It is extended in [17] to include Virtual-CSMA, a technique that helps to estimate the conditional collision probability. This estimation is used to compute the number of contending stations (n) which, in turn, is used to obtain the minimum contention window as

$$CW_{min} = [n \cdot RAND(7, 8)]. \quad (1)$$

The contributions of this paper are as follows. First, it provides a general framework to study the efficiency of the backoff protocols. From this framework, the optimal transmission probability is derived and the optimal efficiency is compared to the efficiency obtained when using T-BEB. The comparison shows that there is room for improvement and that it is possible to design a backoff algorithm that performs better than T-BEB. It is observed that the fraction of slots containing a collision is independent of the number of contending stations when optimal transmission probability is used. Conversely, the fraction of slots containing collisions increases with the number of stations when T-BEB is used.

Inspired by this observation, a variant of the P-Persistent backoff algorithm is proposed. It is called Dynamic P-Persistent backoff (DPP) and dynamically adjusts the transmission probability to reach the optimal (constant) target fraction of collision slots. Thus the problem of estimating the number of contending stations is suppressed and substituted by an easier one which is estimating the fraction of collision slots. This estimation is performed using an exponential moving average estimator based on direct channel observations.

In addition to being simpler than the other optimization proposals mentioned in this section, DPP also presents advantageous implicit prioritization properties. The behaviour of DPP reduces the delay suffered by real-time traffic and interactive traffic in the presence of background traffic, when compared to the other backoff solutions. While previous research focused on either optimization or prioritization, DPP presents simultaneous improvements in both fields.

III. BINARY EXPONENTIAL BACKOFF AND PERFORMANCE ANALYSIS

This section introduces T-BEB which is part of the popular suite of protocols IEEE 802.11. This protocol is an example of CSMA algorithm in which the stations transmit without any previous knowledge about other stations intentions to transmit. The second part of this section assesses the performance of T-BEB, and finds the theoretical efficiency upper bound for this sort of algorithms.

A. Binary Exponential Backoff

The MAC mechanism used in IEEE 802.11 networks is called Distributed Coordination Function (DCF). Although the standard considers also a centralized alternative - the Point Coordination Function - it has been sparsely implemented.

In T-BEB, when a station that has its MAC queue empty receives a packet from the upper layer, it is allowed to transmit the packet after sensing the channel empty [1]. Otherwise, when the MAC queue is not empty or a packet arrives to the Head-Of-Line (HOL) of the MAC queue after the previous packet is successfully transmitted, the station has to backoff.

The backoff consists on a random draw from a Contention Window (CW) and waiting for that number of slots before transmitting. For the first transmission attempt the minimum congestion window is used (CW_{min}). If there is a collision, the congestion window doubles ($CW = 2 \cdot CW_{min}$) and the station randomly chooses a new number and waits for that number of slots before re-attempting transmission. The CW doubles after each collision until it reaches a maximum value CW_{max}. After a successful transmission the value of CW is reset to its minimum. Vanilla IEEE 802.11 takes the values 32 and 1024 for its minimum and maximum contention windows, respectively.

With the IEEE 802.11e [18] standard amendment for Quality of Service support, the values of CW_{min} and CW_{max} can vary. However, the essence of the T-BEB remains the same.

Four our analysis we will consider traffic sources that are saturated, *i.e.* each active station has always a packet ready to transmit. Intuitively, if there is only one active station in the network, it is expected to transmit one slot in every 16 slots.

It is apparent that an efficiency problem exists, since only one of every 16 slots is used while the rest remain empty. Nevertheless the problem is not as acute as it may seem at a first glance, because an empty slot is much shorter than a busy slot. Actually, the duration of an empty slot is $20 \mu s$ in IEEE 802.11b while the duration of a successful slot is in the order of ms. The exact value of the latter depends on the length of the data contained in the packet.

As the number of stations increases, the number of empty slots decreases. Additionally, there are chances that two or more stations transmit on the same slot and that the transmissions are lost due to collision. A slot containing a collision is even longer than a successful slot. Therefore it is critical to reduce the number of collisions.

T-BEB reacts to collisions by doubling the contention window, thus diminishing the transmission rate of the stations. This reaction reduces the load on the network and should decrease the collision probability. Note, however, that it is necessary that there is one collision for the algorithm to realize that the network is highly loaded. Since the value of CW is reset to CW_{min} after a successful transmission, the station has to learn about the network congestion conditions for every packet, and every time there has to be a collision for the station to adjust its CW value. This is a relatively high price to pay for adjusting the CW to its optimal value.

It is shown in [13] that small contention windows are desirable when the number of contending stations is low, to reduce the number of empty unused slots. Conversely, for a large number of stations, larger contention windows offer better performance because reduce the collision probability. The framework provided by IEEE 802.11e can be used to dynamically tune the values of CWmin and CWmax to adapt to the number of contending stations. However, as explained in the previous section, this strategy requires previous estimation of the number of active stations n [19].

This qualitative analysis of T-BEB can help to understand the trade-off in choosing the right CW. A quantitative analysis

---

[1] The channel has to be sensed for a DIFS (Distributed-coordination-function Inter Frame Space).

of the algorithm can be obtained using Markov Chains and the assumption that, regardless of the number of retransmissions, a packet collides with constant probability [7]. Using that model, it is possible to compute the probability that a given station attempts transmission in a given slot (τ). This probability can then be used to obtain the probability of an empty, successful and collision slot. With these values, the overall performance of T-BEB can be evaluated and compared to other mechanisms.

The backoff process pursues the random distribution of the transmission attempts among the slots. An important goal is to maximize the number of successful transmissions while minimizing the collision probability. It is also important to keep the number of empty slots relatively low. However, an empty slot is much more desirable than a collision since the duration of the empty slots is orders of magnitude lower than the duration of a collision.

B. Efficiency of CSMA Algorithms

In CSMA algorithms, the stations autonomously decide whether to transmit or not. The probability that a station transmits (τ) is the key parameter to compute the probability of empty (P_e), successful (P_s) or collision [2] (P_c) slot. For a given number of contending stations n:

$$P_e = (1-\tau)^n, \tag{2}$$

$$P_s = n\tau(1-\tau)^{n-1}, \tag{3}$$

$$P_c = 1 - P_e - P_s. \tag{4}$$

The probability that a station transmits τ can be derived from [7] and is:

$$\tau = \frac{2(1-2p_{cc})}{(1-2p_{cc})(CW_{min}-1) + p_{cc}CW_{min}(1-(2p_{cc})^m)},$$
$$p_{cc} = 1-(1-\tau)^{n-1}. \tag{5}$$

where p_{cc} is the conditional collision probability, *i.e.* the probability that a collision occurs given that one tagged station is attempting transmission. CW_{min} is the minimum congestion window and m the maximum backoff stage.

We define the efficiency as the fraction of time that the channel is used for successful transmissions. It is understood that the time that the channel remains empty or busy with collisions is wasted.

$$\phi = \frac{T_s P_s}{T_e P_e + T_s P_s + T_c P_c}. \tag{6}$$

In Eq. 6 we can observe that the duration of empty, successful and collision slots also affect the observed efficiency. While T_e is constant and defined in the standard, T_s and T_c are a function of the length of the frames. The duration of successful and collision slots are similar, thus the duration of a collision can be approximated to the duration of a successful slot $T_c \approx T_s$. Using the approximation and substituting Eqs. 2 - 4 into Eq. 6 we obtain:

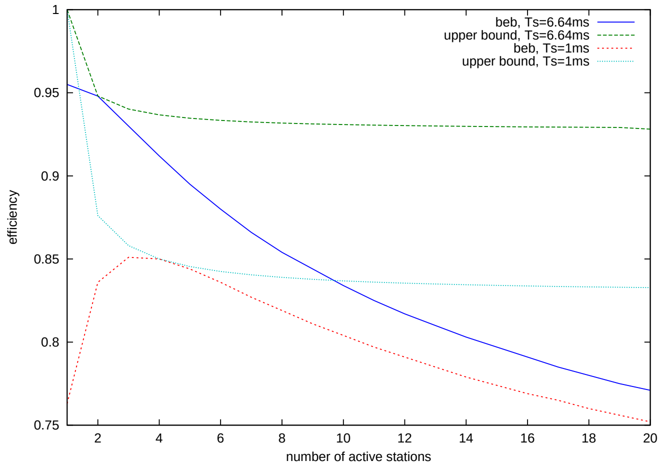

Fig. 1. This figure compares the performance of BEB to the theoretical maximum for different values of successful slot duration T_s.

$$\phi = \frac{n\tau(1-\tau)^{n-1}}{1 - \frac{T_s - T_e}{T_s}(1-\tau)^n} \tag{7}$$

From Eq. 7 it can be observed that the efficiency increases when using large frames. Given a number of contending stations n and a successful slot duration T_s, the optimal transmission probability τ that maximizes efficiency satisfies:

$$\frac{d\phi}{d\tau} = \frac{(1-\tau)^{n-1} + (n-1)\tau(1-\tau)^{n-2}}{1 - \frac{T_s - T_e}{T_s}(1-\tau)^n} - \frac{\frac{T_s - T_e}{T_s} n\tau(1-\tau)^{2(n-1)}}{(1 - \frac{T_s - T_e}{T_s}(1-\tau)^n)^2} = 0 \tag{8}$$

In Fig. 1, the efficiency using optimal values of τ is plotted. Fig. 2 shows that when using an optimal transmission probability, the collision probability is (almost) independent of the number of active stations. This interesting property can be used to derive a near-optimal contention algorithm based on a variant of the P-Persistent mechanism explained in the introduction.

IV. DP-PERSISITENT CSMA

The observation that the collision probability is almost constant when the transmission probability τ is optimal can be exploited to increase the efficiency to values closer to the theoretical optimum.

The proposal consists on observing the channel to estimate the collision probability. Then the stations adapt the transmission probability τ to adjust the collision probability to the target (optimal) collision probability.

---

[2] The notation P_c is used in this paper to denote the probability that a slot is busy with collision. This is different to the conditional collision probability (p or p_c in many papers) which is the probability that a collision occurs conditioned to the event that a tagged station attempts transmission.

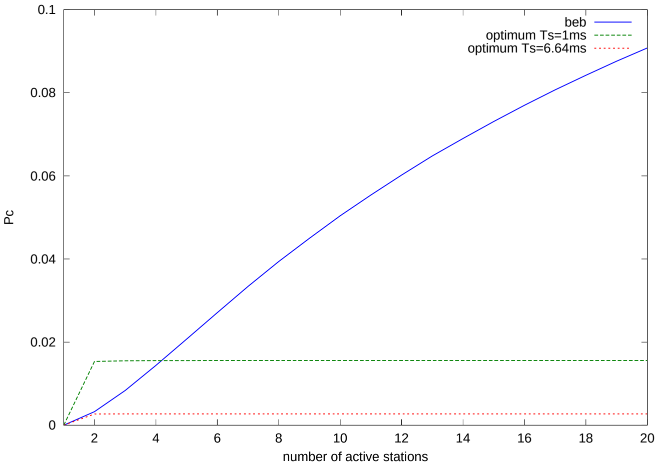

Fig. 2. This figure compares the collision probability obtained when using BEB with one that would be obtained when using optimal transmission probability.

Algorithm 1 Transmission probability adaptation

```
{ $\tau$ is the transmission probability}
{ $\hat{P}_c$ is the estimated collision probability }
{ $P_c^T$ is the target collision probability }

{ $\tau$ and $\hat{P}_c$ are initialized }
$\tau \leftarrow \tau_0$
$\hat{P}_c \leftarrow \hat{P}_{c0}$
while There are packets ready to transmit do
    Sense the channel
    {Moving exponential average is used to update $\hat{P}_c$}
    if Collision then
        $\hat{P}_c \leftarrow \epsilon + (1 - \epsilon) \cdot \hat{P}_c$
    else
        $\hat{P}_c \leftarrow (1 - \epsilon) \cdot \hat{P}_c$
    end if
    {$\tau$ is updated using AIMD}
    if $\hat{P}_c < P_c^T$ then
        $\tau \leftarrow MIN\left[\tau + \alpha(P_c^T - \hat{P}_c), \tau_{max}\right]$
    else
        $\tau \leftarrow \frac{\tau}{1+\mu(\hat{P}_c - P_c^T)}$
    end if
end while
```

Algorithm 1 explains how the transmission probability is distributedly adjusted to attain the optimal collision probability. $\hat{P}_c$ is the estimated collision probability and is computed as an Exponential Moving Average (EMA) based on the observation of the channel. Then, the estimated collision probability ($\hat{P}_c$) is compared to the target collision probability (P_c^T).

If $\hat{P}_c > P_c^T$, the transmission probability (τ) is decremented. Otherwise, the transmission probability is increased. We adopt an Additive Increase Multiplicative Decrease (AIMD) approach for the tuning of τ. The reason for this choice is that it provides long-term fairness among competing flows, even when they begin with different values of τ.

It can be observed that Algorithm 1 includes a number of parameters ($P_c^T, \tau_0, \hat{P}_{c0}, \epsilon, \alpha, \mu, \tau_{max}$). Each of this parameters conditions the overall performance of the backoff mechanism, and the selection of these parameters also involve some kind of trade-off. In the following, we summarize and discuss the values of these parameters.

P_c^T is the target collision probability, *i.e.* the collision probability that delivers optimal performance. Unfortunately, P_c^T is a function of the duration of a successful transmission (T_s). Assuming a data rate of 11Mbps, T_s takes values from 0.6 ms (when the frame carries no data) to 9.9 ms (when the payload is maximum, 2304 bytes). The actual packet size distribution in WLAN [20] is trimodal, being most of the packets smaller than 100 bytes or larger than 1470 bytes, with a lower fraction around 600 bytes. Since the duration of a collision is approximately equal to the duration of the longest packet involved in the transmission, the conservative decision of assuming a payload size of 1500 bytes is adopted.

If the payload size is 1500 bytes, the duration of a slot containing a successful transmission is 6.64ms and the optimal collision probability (as described in Sec. III) is 0.0027. Therefore, the target collision probability P_c^T is set to 0.0027.

Since the minimum contention window in IEEE 802.11b is 32 (the stations would transmit every 16 slots on average if there were no collisions), a value of 1/16 have been chosen as initial transmission probability τ_0. The initial estimated collision probability $\hat{P}_{c0}$ is set to the target collision probability P_c^T. As the station senses the channel, it will obtain a finer value of $\hat{P}_c$ that can be used to adapt τ and take it closer to the optimal value.

The EMA estimator uses the parameter ϵ. It must take values between 0 and 1. A high value of ϵ gives more weight to what has happened in recent slots and makes the estimation to react faster to new conditions (*i.e.* addition or suppression of a contending station or changes in transmission probability τ). However, since collisions happen seldom, a high value of ϵ can easily lead to excessive oscillations that would set τ far from its optimal value. Thus a value of 0.001was chosen for ϵ.

The parameters α and μ represent the Additive Increase and Multiplicative Decrease of τ respectively. As happens with ϵ, a higher value offers prompt reactions but also increases the risk of larger oscillations that penalize performance. Their values $\alpha = 0.01$ and $\mu = 0.05$ were choosen empirically, after observing their impact in simulation results.

Finally, there is a need to limit the maximum transmit probability τ_{max}. The purpose of τ_{max} is to prevent τ to grow to 1 in the special case in which there is only one active station. A transmission probability of 1 would boost the efficiency to 100% but would hamper the entry of a new contender. A value $\tau_{max} = 1/8$ is a good compromise to guarantee high efficiency when there is only one station while leaving 7 out of 8 slots free for the new contender to successfully transmit.

TABLE I

PARAMETER VALUES

P_c^T	τ_0	$\hat{P}_{c0}$	ϵ	α	μ	τ_{max}
0.0027	1/16	0.0027	0.001	0.01	0.05	1/8

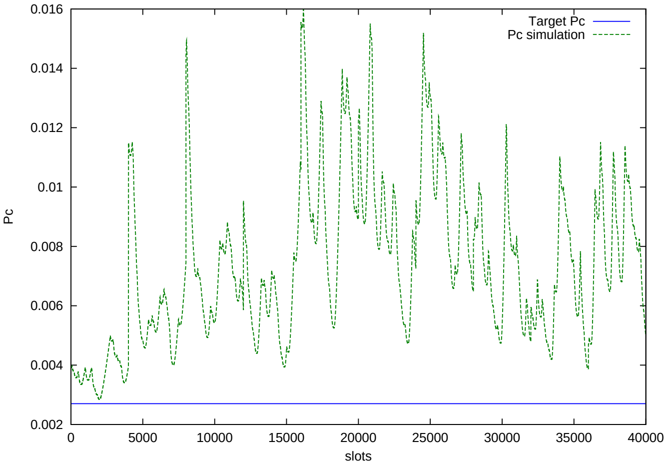

Fig. 3. The actual P_c is compared to the target P_c^T. The number of active stations is increased from 2 to 11. A station is added every 4000 slots

Table I summarize the parameters and its values.

V. SIMULATION RESULTS

Using the algorithm and parameters described in previous section, simulations[3] can be used to observe the results obtained using the proposed alternative backoff algorithm. First we present a toy scenario in which the number of stations is increased from two to eleven. The increments happen every 4000 slots. The case with only one station is omitted in the figures because it presents results so different from the other cases that obfuscate the resultant plots. When there is only one station the collision probability is equal to zero, and the transmission probability tends to τ_{max}.

The following plots show the actual collision probability compared to the target collision probability (Fig. 3), the actual transmission probability compared to the optimal transmission probability (Fig. 4) and the actual efficiency compared to the achievable maximum (Fig. 5).

In Fig. 3 it can be observed that that the backoff algorithm tries to keep the collision probability close to the (constant) target collision probability for any number of stations. When the number of stations increases (at slot 4000, 8000, etc.) a spike appears in the actual collision probability. It takes some time for the stations to detect the increased number of collisions and reduce the transmission probability and thus adjust the collision probability to a value closer to to the desired one. A careful observer would notice that the actual collision probability (P_c) is larger than the target collision probability (P_c^T). There are two causes for this misadjustment: (a) the estimator fails to capture the instant collision probability (b) The τ parameter tuning is a slow iterative process. Nevertheless, P_c is close enough to P_c^T to offer excellent efficiency.

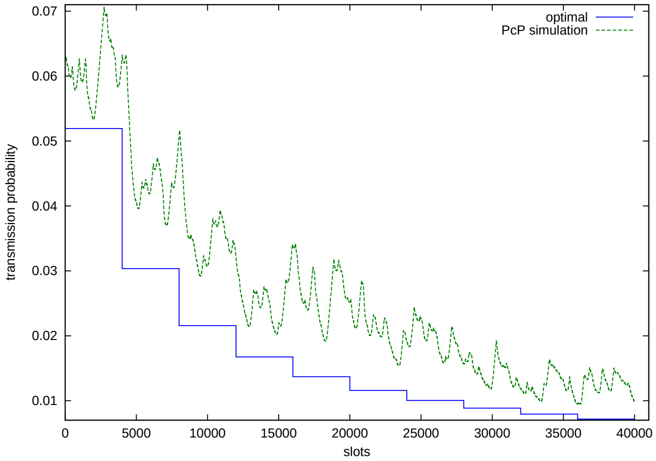

Fig. 4. The actual transmission probability τ is compared to the optimal transmission probability τ^{opt}. The number of active stations is increased from 2 to 11. A station is added every 4000 slots

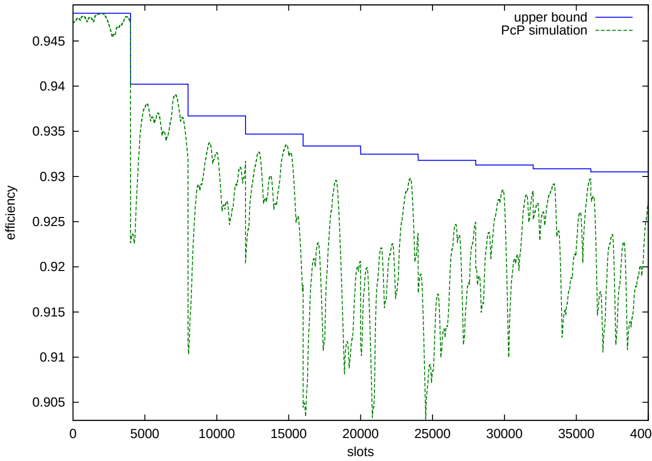

Fig. 5. The actual efficiency ϕ is compared to the optimal efficiency ϕ^{opt}. The number of active stations is increased from 2 to 11. A station is added every 4000 slots

Fig. 4 shows the transmission probability observed in the simulations compared to the optimum transmission probability. Again, it can be observed that the stations require some time to adapt to a scenario change. However, in the long term, the actual transmission probability approximately follows the optimal transmission probability.

Finally, in Fig. 5, we can observe the benefits of the proposed backoff scheme. The obtained efficiency closely sticks to the optimal efficiency for any number of stations.

In the previous example and figures, the dynamic behaviour of the algorithm has been explained by observing a simulation in which the number of active stations is variable and the control loop implemented in the backoff algorithm actuates to

---

[3] The simulations and the numerical computations were performed using octave. All the scripts are available upon request to the corresponding author.

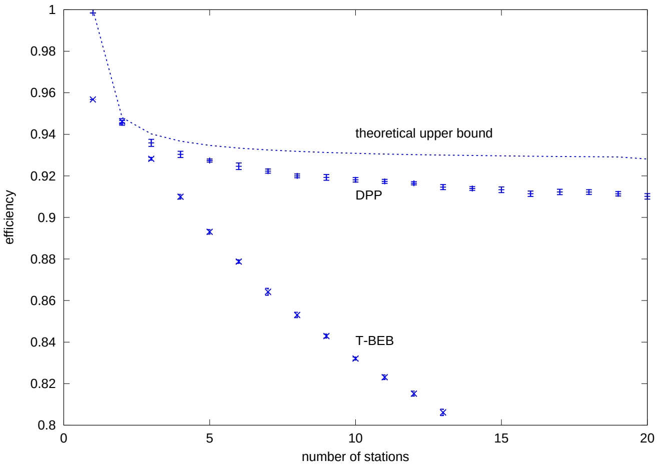

Fig. 6. Theoretical maximum (dashed line) compared to simulations results of DP-Persistent CSMA and T-BEB. The 95% confidence intervals are plotted.

adjust the probability of a collision slot to a fixed (optimal) value.

In order to asses with greater accuracy the performance delivered by DPP, simulations for a fixed number of stations have been performed. Each simulation comprises 80,000 slots and has been repeated 10 times with different random seeds. Fig. 6 shows the results and compares them to the theoretical maximum computed in Sec. III and depicted in Fig. 1. It can be observed that DPP performs close to the theoretical maximum in steady-state operation.

VI. IMPLICIT PRIORITIZATION

Current data networks carry heterogeneous traffic. Internet traffic can be classified in background, interactive and real-time traffic. Background traffic transfer large amounts of data with no stringent delay constraints. This traffic is carried by long-lived TCP flows that are permanently active. A good example of background traffic is peer-to-peer file sharing. This data is transferred without the active participation of any human being.

Interactive traffic is originated and consumed by users. It consists in small data burst such as a request for a webpage and the consequent response from the server. This are short-lived TCP interactions in which a relatively small amount of data needs to be transmitted in a reasonable amount of time. Reasonable is a lax definition and depends on the expectations from the users, and is probably in the order of one second. Users would prefer a shorter reaction time; therefore, for this kind of traffic, delay does matter.

The last kind of traffic is real-time traffic. Very small quantities of data are sent periodically to maintain a voice or video flow. For real-time flows delay is critical, and those packets that suffer excessive delay are useless at reception and are discarded.

It is a desired property of a network that allows the harmonious coexistence of different kinds of traffic. Ideally, real-time traffic would traverse the networks with the highest priority to reach the destination in tens of milliseconds. Interactive traffic comes second in the priority row, since there is a user waiting for an answer and that waiting time should be minimized. When neither real-time nor interactive traffic is transmitted, the network can be used to transmit background traffic.

From the previous argumentation it can be concluded that the priority of a data transfer maintains an inverse relationship with its duration. In the following, it will be explained that this is exactly the treatment that stations deserve under the DPP backoff mechanism.

It has to be noticed that every station enters the playground with a initial transmission probability $\tau_0 = 1/16$. In its commitment to lower the number of collisions to achieve the maximum efficiency, DPP lowers the transmission probability. The result is a large fraction of empty slots (about 90%) and transmission probabilities lower than τ_0 for a number of stations equal or larger than 3. With this scenario, a station becoming active after an inactivity period enjoys priority for a limited initial period of time.

Due to the slow nature of the EMA average and the τ adjustment mechanism explained in Sec. IV, it takes some time for the newcomer to lower its own transmission probability from the initial value τ_0 to the optimal value τ_{opt}. This time can be used to transmit with higher priority than the other stations that have been active for a long time. A station transmitting a burst of data will observe that the first packets of the burst enjoy priority, but that priority vanishes as times passes and its own transmission probability is slowly decreased. The result is that shorter burst will be transmitted with higher priority that longer bursts.

The behaviour of DPP can be summarized as assigning priority to stations that become active after an inactivity period. This priority fades away as the station continues active for a longer period. Fig. 7 shows a single station generating voice traffic competing against five peer-to-peer saturating stations. The voice station has a new packet to send one in every 100 slots, it competes for the channel until it has sent that packet and then leaves the contention. When the voice station rejoins the contention to send a new packet, it uses the initial transmission probability τ_0. The peer-to-peer stations are constantly contending for the channel and do not have the chance to reset their transmission probability to τ_0.

Even though DPP exhibits convenient prioritizing properties, it does not completely solve priority issues. There are two aspects in which DPP falls short of solving the problem. The first one involves uplink/downlink unfairness in infrastructure scenarios. All the stations transmit to the access point and the access point transmits to all stations. The latter easily becomes the bottleneck of the network and requires higher priority.

DPP does not solve the issue of stations transmitting heterogeneous traffic. A station that sends both real-time and background traffic would be continuously active and would not benefit from the early priority commented in this section.

Nevertheless, DPP offers advantageous implicit prioritizing properties when compared with IEEE 802.11.

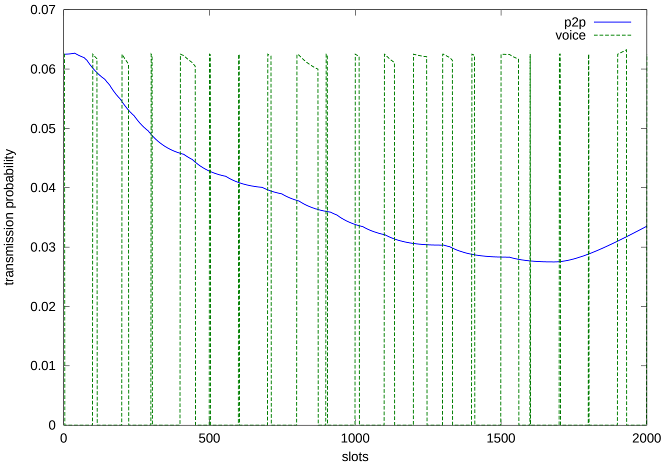

Fig. 7. A single station generating voice traffic competes against five peer-to-peer stations for the channel. The voice station periodically enters the contention with transmission probability τ_0 and leaves the contention once the voice packet has been transmitted.

VII. CONCLUSION

This paper studies the performance of backoff mechanisms in terms of efficiency, *i.e.* the fraction of time that is devoted to successful transmissions compared to the time wasted in empty slots and collisions. Optimal efficiency can be obtained by adjusting the transmission probability τ of the stations. It is shown that the optimal transmission probability τ_{opt} depends on the packet length and the number of active stations. It is also observed that the fraction of slots containing a collision P_c is almost constant when optimal transmission probability is used.

The efficiency of T-BEB is compared to the optimum to show that there is room for improvement. Then an algorithm called DPP is proposed. This algorithm dynamically adjusts the transmission probability τ to achieve optimal collision probabiliy P_c which is known and constant. As opposed to backoff mechanisms proposed in previous art, DPP does not need to estimate the number of contending stations. Additionally, DPP outperforms BEB and achieves near-optimal efficiency.

DPP is a completely distributed backoff scheme in which the stations monitor the channel to estimate the collision probability and dynamically adjust their transmission probability in the quest for optimal efficiency. Both the estimation and the parameter adjustment takes some time. This results in stations awaking from an inactivity period having higher priority than those that have been active for a longer period of time. This proves beneficial since reduces the delay of real-time and interactive applications while maintains near-optimal throughput for background traffic.

ACKNOWLEDGMENT

We would like to acknowledge the anonymous reviewers for their insightful comments.

REFERENCES

[1] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, ANSI/IEEE Std. Std 802.11, 1999 Edition (Revised 2003).

[2] N. Abramson, “The ALOHA System–Another Alternative for Computer Communications,” *Cluster Computing*, vol. 5, pp. 187–201, 1970.

[3] *Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD), Access Method and Physical Layer Specifications*, IEEE Std. Std 802.3, 2000 Edition.

[4] J. Jubin and J. Tornow, “The DARPA packet radio network protocols,” *Proceedings of the IEEE*, vol. 75, no. 1, pp. 21–32, 1987.

[5] J. Goodman, A. Greenberg, N. Madras, and P. March, “Stability of binary exponential backoff,” *Journal of the ACM (JACM)*, vol. 35, no. 3, pp. 579–602, 1988.

[6] B. Kwak, N. Song, and L. Miller, “Analysis of the Stability and Performance of Exponential Backoff,” *Proceedings of IEEE WCNC*, pp. 1754–1761, 2003.

[7] G. Bianchi, “Performance analysis of the IEEE 802.11 distributed coordinationfunction,” *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 3, pp. 535–547, 2000.

[8] S. Wu, Y. Tseng, C. Lin, and J. Sheu, “A Multi-channel MAC Protocol with Power Control for Multi-hop Mobile Ad Hoc Networks,” *The Computer Journal*, vol. 45, no. 1, pp. 101–110, 2002.

[9] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications for low-rate wireless personal area networks*, IEEE Std. 802.15.4, 2003 Edition (Revised 2006).

[10] F. Cali, M. Conti, E. Gregori, and P. Aleph, “Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit,” *Networking, IEEE/ACM Transactions on*, vol. 8, no. 6, pp. 785–799, 2000.

[11] G. Bianchi and I. Tinnirello, “Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network,” *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE*, vol. 2.

[12] A. Lopez-Toledo, T. Vercauteren, and X. Wang, “Adaptive Optimization of IEEE 802.11 DCF Based on Bayesian Estimation of the Number of Competing Terminals,” *IEEE TRANSACTIONS ON MOBILE COMPUTING*, vol. 5, no. 9, p. 1283, 2006.

[13] H. Anouar and C. Bonnet, “Optimal Constant-Window Backoff Scheme for IEEE 802.11 DCF in Single-Hop Wireless Networks Under Finite Load Conditions,” *Wireless Personal Communications*, vol. 43, no. 4, pp. 1583–1602, Dec. 2007.

[14] J. Capetanakis, “Tree algorithms for packet broadcast channels,” *Information Theory, IEEE Transactions on*, vol. 25, no. 5, pp. 505–515, 1979.

[15] X. Wang, Y. Yu, and G. Giannakis, “Design and analysis of cross-layer tree algorithms for wireless random access,” *Wireless Communications, IEEE Transactions on*, vol. 7, no. 3, pp. 909–919, March 2008.

[16] L. Zhao, J. Zhang, K. Yang, and H. Zhang, “Using Incompletely Cooperative Game Theory in Mobile Ad Hoc Networks,” *Communications, 2007. ICC’07. IEEE International Conference on*, pp. 3401–3406, 2007.

[17] L. Zhao, J. Zhang, and H. Zhang, “Using Incompletely Cooperative Game Theory in Wireless Mesh Networks,” *Network, IEEE*, vol. 22, no. 1, pp. 39–44, 2008.

[18] “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications; Amendment: Medium Access Control(MAC) Quality of Service Enhancements,” no. IEEE Std 802.11e, 2005.

[19] T. Vercauteren, A. Lopez-Toledo, and X. Wang, “Batch and Sequential Bayesian Estimators of the Number of Active Terminals in an IEEE 802.11 Network,” *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 55, no. 2, pp. 437–450, 2007.

[20] C. Na, J. Chen, and T. Rappaport, “Measured traffic statistics and throughput of IEEE 802.11 b public WLAN hotspots with three different applications,” *IEEE Trans. Wireless Commun*, vol. 5, no. 11, pp. 3296–3305, 2006.