Effective Capacity Theory for Modeling Systems with Time-Varying Servers, with an Application to IEEE 802.11 WLANs

Emmanouil N. Kafetzakis*

National Center for Scientific Research "Demokritos"
Institute of Informatics & Telecommunications
Department of Informatics & Telecommunications
National & Kapodistrian University of Athens
mkafetz@di.uoa.gr

Abstract. Many network applications rely on stochastic QoS guarantees. With respect to loss-related performance, the Effective Bandwidth/Capacity theory has proved useful for calculating loss probabilities in queues with complex input- and server-processes and for formulating simple admission control tests to ensure associated QoS guarantees. This success has motivated the application of the theory for delay-related QoS too. However, up to now this application has been justified only heuristically for queues with variable service rate. The thesis fills this gap by establishing rigorously that the Effective Bandwidth/Capacity theory may be used for the asymptotically correct calculation and enforcement of delay tail-probabilities in systems with variable rate servers too. Subsequently, the thesis applies the general results to IEEE 802.11 WLANs, by representing each IEEE 802.11 station as an On/Off server and employing the Effective Capacity function for this model. Comparison of analytical results with simulation validates the effectiveness of the On/Off IEEE 802.11 model for loss- and delay-related QoS. Finally, the thesis uses the Effective Capacity of an IEEE 802.11 station in yet another way, namely as a design tool. Indeed, the specific form of the IEEE 802.11 Effective Capacity function highlights the role of certain parameters of the IEEE 802.11 backoff window distributions. These parameters, when appropriately tailored, allow better delay-related (and loss-related) performance, while maintaining the standard saturation throughput of IEEE 802.11 WLANs.

Keywords: admission control, Effective Capacity, IEEE 802.11, quality of service, server modeling, tail-probabilities

1 Introduction

Many demanding network applications rely on stochastic Quality of Service (QoS) guarantees. With respect to loss-related performance, the asymptotic theory based on the notions of Effective Bandwidth and Effective Capacity has proved successful for

* Dissertation Advisors: Kimon Kontovasilis, Research Director N.C.S.R. "Demokritos" and Ioannis Stavrakakis, Professor N.K.U.A.

calculating low loss probabilities in queueing systems with complex time-varying input and server processes and for formulating simple admission control tests to enforce associated QoS guarantees (see, e.g., [12, 4, 5]).

This success has motivated the application of the theory to the calculation and enforcement of delay-related QoS too. However, up to now this application has only been justified on the basis of heuristic arguments when the queue is served at a variable rate (see, e.g., [1, 14]). The thesis fills this gap, by formally establishing that the Effective Bandwidth/Capacity theory may be applied for asymptotically correct calculation and/ or enforcement of delay tail-probabilities in systems with variable rate servers too. In particular, the heuristically suggested linkage between the exponential decay rates of the buffer content and delay probability tails through the server's Effective Capacity function is formally shown to apply [7].

Due to the prevalence of wireless networking, systems with time-varying servers are becoming all the more important. Indeed, a wireless station can be regarded as a time-varying data server, due to rate fluctuations at the physical or at the medium access control layer. In this context, the thesis proceeds with an application of the general results to IEEE 802.11 WLANs. In doing so, the thesis first establishes that an IEEE 802.11 mobile station can be regarded as a Semi-Markovian data server of the On/Off type, with known distributions for the On and Off periods, and subsequently derives the Effective Capacity function of this On/Off server [8]. The general results can then be used for computing buffer overflow and delay violation probabilities in WLANs, and for employing simple traffic control policies to enforce related QoS guarantees.

Finally, the thesis illustrates the usage of the Effective Capacity function of the IEEE 802.11 stations as a design tool: Towards this end, the form of the said function highlights certain parameters of the backoff window distributions, which, if appropriately tailored, may lead to higher Effective Capacity values, hence to better delay-related (or loss-related) performance.

The rest of this thesis summary is organized as follows: Section 2 firstly reviews preexisting Eff. Bandwidth/Capacity results about buffer content tail-probabilities and then justifies the applicability of Eff. Bandwidth/Capacity theory in connection with delay tail-probabilities. Section 3 briefly presents the analytical model used to characterize the IEEE 802.11 Distributed Coordination Function (DCF) as an On/Off server. This model is used to derive the Eff. Capacity of the IEEE 802.11 protocol. Subsequently, it discusses computational and algorithmic issues related to the application of the general theory of Section 2 with the particular Eff. Capacity function of this On/Off model. Section 4 describes how the Eff. Capacity function of IEEE 802.11 stations may be used for an informed choice of parameters for the backoff window distributions, towards better performance. Section 5 provides validation of the IEEE 802.11 model, through comparison of the analytical results with simulations. Finally, the thesis summary is concluded in Section 6.

2 Effective Bandwidth and Effective Capacity Theory

Effective Bandwidth and Effective Capacity theory offers a linkage between input load, system capacity and QoS requirements and it was developed by a great number of contributions from various researchers. The theory was originally developed for queueing systems with constant server capacity.

When the server's capacity is time-varying independently from the input, the theory can be generalized, by defining an *Effective Capacity function* to capture the server's burstiness. Although this generalization has been studied for some years (see, e.g., [5, 6]) it did not attract much attention until recently [13, 14, 1] when the importance of wireless systems grew considerably. This is because most such systems feature a variable service rate and Effective Capacity is ideal for modeling such settings.

For a quick review of the Eff. Bandwidth/Capacity theory, consider at first a single-server queue, fed by a traffic stream that produces an amount of data $V(t)$ within a time-window $(-t, 0]$ and let c be the constant service rate. According to the Eff. Bandwidth theory, provided that $V(t)$ has stationary increments and satisfies some additional mild technical conditions (see, e.g., [7]), the probability that the queue size Q exceeds a certain threshold b has at all times an exponential upper bound of rate $\theta > 0$. Specifically, the tail of the queue-length distribution satisfies

$$a_V(\theta) < c \Rightarrow \lim_{b\to\infty} b^{-1} \log \Pr\{Q > b\} \leq -\theta,$$

where

$$a_V(s) \triangleq \frac{1}{s} \lim_{t\to\infty} \frac{1}{t} \log \mathrm{E}\left[e^{sV(t)}\right], \quad s \in \mathbb{R}, \tag{1}$$

determines the Effective Bandwidth function.

Now consider a single-server queue with time-varying capacity, where the capacity fluctuations are independent from the input. Let the input traffic be as previously and denote by $C(t)$ the amount of data that can be processed within the time-window $(-t, 0]$. Assuming the same technical conditions for the input and output processes (in particular stationary increments), the probability that the queue size Q exceeds a certain threshold b has at all times an exponential upper bound of rate $\theta > 0$, viz.,

$$a_V(\theta) < a_C(-\theta) \Rightarrow \lim_{b\to\infty} b^{-1} \log \Pr\{Q > b\} \leq -\theta, \tag{2}$$

where $a_V(\theta)$ stands for the Effective Bandwidth function in (1) and

$$a_C(s) \triangleq \frac{1}{s} \lim_{t\to\infty} \frac{1}{t} \log \mathrm{E}\left[e^{sC(t)}\right], \quad s \in \mathbb{R}, \tag{3}$$

determines the Effective Capacity function.

Given that *strict* monotonicity holds for at least one of the Eff. Bandwidth and Eff. Capacity functions,

$$a_V(\theta) \leq a_C(-\theta) \Leftrightarrow \lim_{b\to\infty} b^{-1} \log \Pr\{Q > b\} \leq -\theta. \tag{4}$$

Equivalence (4) always holds in the context of IEEE 802.11 WLAN, since the IEEE 802.11 Eff. Capacity is always a strictly increasing function.

Assume now that the queue has a finite size q, and that we want to provide stochastic QoS by limiting the overflow probability to a value $\leq e^{-\epsilon}$. By using the tail percentile $\Pr\{Q > q\}$ of the respective infinite queue as a proxy for the overflow probability, (2) implies that (asymptotically) $\theta = \epsilon/q$ and the input traffic must satisfy

$$a_V(\epsilon/q) < a_C(-\epsilon/q).$$

This inequality directly suggests an Admission Control (AC) scheme by bounding the input traffic.

Finally, let $\theta^* \triangleq \sup\{\theta \in \mathbb{R} \,|\, a_V(\theta) \leq a_C(-\theta)\}$. If there exist $\theta_o > 0$ such that $a_V(\theta_o) < a_C(-\theta_o)$ then the asymptotic exponential decay rate of the tail-probabilities of Q equals to θ^* [7], i.e.,

$$\lim_{b\to\infty} b^{-1} \log \Pr\{Q > b\} = -\theta^*. \tag{5}$$

The conceptual simplicity of the Eff. Bandwidth/Capacity theory makes it an attractive choice for coping with delay-related QoS as well. For First-Come-First-Served (FCFS) queueing systems with a constant service rate c this is directly possible, because delay probabilities of the form $\Pr\{D > d\}$ are equal to the queue length probabilities $\Pr\{Q > cd\}$. However, this simple equivalence does not hold when the service rate is time-varying.

The thesis (see the relevant results in [7]) formally establishes that the Eff. Bandwidth/Capacity theory may be applied for the asymptotically correct calculation and enforcement of delay tail-probabilities in the general setting with variable service rate. Specifically,

$$a_V(\theta) < a_C(-\theta) \Rightarrow \lim_{d\to\infty} d^{-1} \log \Pr\{D > d\} \leq -\theta a_C(-\theta). \tag{6}$$

As with the queue content, we now consider admission control for ensuring delay-related QoS guarantees. Let $u_C(s) \triangleq \lim_{t\to\infty} t^{-1} \log \mathrm{E}\left[e^{sC(t)}\right]$, $s \in \mathbb{R}$, be the asymptotic cumulant generator of $C(t)$. By the definition of the Eff. Capacity function in (3),

$$a_C(s) = u_C(s)/s, \quad s \in \mathbb{R}. \tag{7}$$

In order to ensure that the decay rate of the delay tail-probabilities is bounded below by some $\xi = \theta a_C(-\theta) = -u_C(-\theta) > 0$, the admission control $a_V(\theta) < a_C(-\theta)$ in (6) must be tested for

$$\theta(\xi) = -u_C^{-1}(-\xi). \tag{8}$$

The value of the parameter ξ to employ in the tests is determined in a way analogous to the one used for loss-related QoS requirements. This time the QoS specification dictates that the delay should not exceed some given threshold τ with probability higher than $e^{-\epsilon}$. Provided that both τ and ϵ are large maintaining a finite ratio, the QoS specification leads to $-\epsilon/\tau \geq \tau^{-1} \log \Pr\{D > \tau\} \approx \lim_{d\to\infty} d^{-1} \log \Pr\{D > d\}$, so $\xi = \epsilon/\tau$ should be used in the admission control tests.

Finally, the thesis establishes the linkage between the decay rate θ^* of the buffer content tail-probabilities and the decay rate ξ^* of the delay tail-probabilities through the server's Effective Capacity function, viz.,

$$\xi^* = \theta^* a_C(-\theta^*).$$

3 Effective Capacity of IEEE 802.11 WLAN

A simple, but accurate analytical model for the saturation throughput computation of the IEEE 802.11 protocol was provided in [2]. The analysis focuses on the saturation condition, where every station has always a packet to send. The analysis also assumes that the number of stations n under contention is known and constant and the probability of a collision seen by a packet being transmitted on the channel, named conditional collision probability p, is constant and independent from the number of retransmission suffered. In order to compute the station's throughput, its behaviour is studied through a Markov chain model. This model yields the probability that a station transmits in a random time-slot, referred to as transmission probability τ.

In [3], a more general model, permitting the usage of arbitrary backoff window distributions, is proposed, generalizing and supplementing [2]. This model takes into account more details of the IEEE 802.11 protocol. By employing the more general analysis of [3] and assuming an infinitive number of retries, one obtains

$$\tau = \left[1 + (1-p)\left(\frac{\overline{W}_0}{1-B_0} - 1 + \sum_{i=1}^{m-1} p^i \overline{W}_i + \frac{p^m \overline{W}_m}{1-p}\right)\right]^{-1} \tag{9}$$

where $\overline{W}_i,\ i = 0, \ldots, m$ is the mean backoff window at the i^{th} stage, m is the backoff stage beyond which the upper window margin does not grow anymore and B_0 is the probability that a backoff window drawn at the 0^{th} stage is zero. The expression $\overline{W}_0/(1-B_0) - 1$ is a modified mean backoff window at 0^{th} stage, when the nonzero backoff window drawn is examined after being initially decremented by one, for synchronization purposes.

Assuming that the Markov chains of the mobile stations are independent, one also obtains that

$$1 - p = (1-\tau)^{n-1}, \tag{10}$$

because a packet will not suffer a collision exactly when all other stations do not attempt to transmit when the station emitting the packet does so. Equations (9) and (10) can be solved uniquely for p and τ.

As already noted, the preceding analysis assumes that all stations are saturated. This thesis employs the values of p and τ obtained from (9) and (10), to calculate the Eff. Capacity of an IEEE 802.11 station, effectively assuming that all other stations are saturated. This approximation is on the safe side (i.e., 'conservative') since assuming the other stations saturated corresponds to the worst case, and has the merit that in this way the Eff. Capacity of a station can be computed without regard to input traffic details. As will be discussed later, the saturation assumption can be waived and the model be applicable to all network load settings.

Due to the CSMA/CA access algorithm, the system can be modeled as a Semi-Markov server model featuring four states [8], as depicted in Fig. 1. State `bc` corresponds to the backoff procedure when the backoff counter is nonzero. State `ov` models overhead time before and after the transmission. It can be proved that the overhead time before and after transmission is possible to be merged in one state. State `tr` corresponds to the transmission (active) period and State `dc` models the idle slot needed for the initial decrement of the backoff window so that other stations realize that a successful transmission is over.

Overheads and transmissions always occur in pairs thus transitions from State `ov` to State `tr` occur with probability one. After the transmission is over, State `ov` is visited

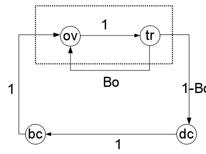

Fig. 1. Semi-Markov chain for IEEE 802.11 MAC.

again with probability B_0 (probability that backoff counter at the 0^{th} stage is zero) so the station transmits a packet successfully, one more time. With probability $1 - B_0$ the station enters the backoff procedure (State `bc`) after the backoff counter has been decremented by one (State `dc`). The service rate in Stage `tr` is equal to the nominal bit rate $\hat{r}$ of the IEEE 802.11 channel, while in all other states the service rate is zero.

Sojourn time distributions in every state are characterized by the respective moment generators. The moment generator of the sojourn time in State `tr` (i.e., payload transmission time) is

$$\gamma_{\text{tr}}(\omega) \triangleq \mathrm{E}\left[e^{P/\hat{r}}\right],$$

where P denotes the payload size. Note that for constant payload size the transmission time is deterministic. The moment generators of time distributions for States `ov` and `dc` are given by

$$\gamma_{\text{ov}}(\omega) \triangleq e^{\omega t_{\text{ov}}}, \qquad \gamma_{\text{dc}}(\omega) \triangleq e^{\omega t_{\text{slot}}}, \tag{11}$$

where

$$t_{\text{ov}} \triangleq \big(RTS + CTS + PHY_{hdr} + ACK\big)/r_{\text{signal}} + MAC_{hdr}/\hat{r} + 3SIFS + DIFS,$$

and

$$t_{\text{coll}} \triangleq RTS/r_{\text{signal}} + EIFS + t_{\text{slot}}$$

are the deterministic overhead time of the transmission and the collision duration, respectively. The quantities r_{signal}, t_{slot}, RTS, CTS, $SIFS$, $DIFS$, $EIFS$, ACK, MAC_{hdr}, PHY_{hdr} denote respectively the transmission rate used for signaling operations, the slot time of the system, the RTS packet size, the CTS packet size, the SIFS time, the DIFS time, the EIFS time, the ACK packet size, the MAC header and the PHY header. The values of all these parameters are determined by the IEEE 802.11 standard [11].

Finally, the backoff time distribution is characterized by the moment generator

$$\begin{aligned}\gamma_{\text{bc}}(\omega) = &\frac{g_0\big(\gamma_s(\omega)\big) - B_0}{\gamma_s(\omega)(1 - B_0)}\left[\sum_{l=0}^{m-1}\Big((1-p)p^l e^{l\omega t_{\text{coll}}}\prod_{j=1}^{l} g_j\big(\gamma_s(\omega)\big)\Big)\right.\\ &\left.+\frac{(1-p)(pe^{\omega t_{\text{coll}}})^m \prod_{j=1}^{m} g_j\big(\gamma_s(\omega)\big)}{1 - pg_m\big(\gamma_s(\omega)\big)e^{\omega t_{\text{coll}}}}\right],\end{aligned} \tag{12}$$

where $g_i(z)$ stands for the probability generator function of the backoff window distribution associated with the i^{th} stage (beyond stage m, the backoff windows maintain the same probability generator $g_m(z)$) and $\gamma_s(\theta)$ denotes the moment generator of the time needed for the reduction by one of the backoff counter.

$$\gamma_s(\omega) = P_{\text{coll}} e^{\omega t_{\text{coll}}} + P_{\text{empty}} e^{\omega t_{\text{slot}}} + P_{\text{succ}} \frac{(1-B_0)\gamma_{\text{tr}}(\omega)}{1-B_0\gamma_{\text{tr}}(\omega)} e^{\omega t_{\text{slot}}}, \tag{13}$$

where

$$P_{\text{succ}} = (n-1)\tau(1-\tau)^{n-2}, \quad P_{\text{empty}} = (1-\tau)^{n-1}, \quad P_{\text{coll}} = 1 - P_{\text{succ}} - P_{\text{empty}}, \tag{14}$$

are the probabilities of a successful transmission, an empty slot and a collision respectively, observed by a station backing-off (which observes $n-1$ other stations).

Note that the generator function $g_i(z)$ depends on the distribution of the backoff window at the i^{th} stage. By definition, $g_0(0) = B_0$ and $g_i'(1) = \overline{W}_i$. For the uniform backoff window distribution, described in the standard,

$$g_i(z) \triangleq \sum_{l=0}^{w_i-1} \frac{1}{w_i} z^l = \frac{1}{w_i} \frac{z^{w_i}-1}{z-1}, \qquad w_i = 2^{\min\{i,m\}} w_0, \quad i \geq 0, \tag{15}$$

where $w_0 - 1$ is the maximum value of the backoff counter at the backoff stage zero.

Given the model just described, a straightforward extension of the Eff. Bandwidth theory for Semi-Markovian models [9] yields the Eff. Capacity function. In fact, it is possible to show that, in terms of the Eff. Capacity, the model is equivalent with an On/Off server model [8] with an On period characterized by the moment generator

$$\gamma_{\text{on}}(\omega) = \gamma_{\text{tr}}(\omega) = \mathrm{E}\left[e^{\omega P/\hat{r}}\right] \tag{16}$$

and an Off period whose moment generator is equal to

$$\gamma_{\text{off}}(\omega) = \gamma_{\text{ov}}(\omega)\Big(B_0 + (1-B_0)\gamma_{\text{bc}}(\omega)\gamma_{\text{dc}}(\omega)\Big) \tag{17}$$

where $\gamma_{\text{on}}(\cdot), \gamma_{\text{ov}}(\cdot), \gamma_{\text{bc}}(\cdot), \gamma_{\text{dc}}(\cdot)$ as in (16), (11), and (12).

Using this alternative On/Off representation the Eff. Capacity function is given by (7) where $u_C(s)$ is the unique negative solution of

$$f\big(s, u_C(s)\big) = 0, \qquad f(s,u) \triangleq \log \gamma_{\text{on}}(\hat{r}s - u) + \log \gamma_{\text{off}}(-u) = 0. \tag{18}$$

The formulation of (9) assumes saturation condition. The dependence of $\gamma_{\text{off}}(\cdot)$ on the saturation assumption is only through the conditional collision probability p, used in (12) and the probabilities P_{succ}, P_{empty} and P_{coll} employed by (13). Under non-saturation condition these parameters retain their meaning, but take different values. Thus, if each mobile station assesses these probabilities by direct measurement, rather than computing them through (9), (10) and (14), the model works well in all settings, lightly loaded ones included.

For the construction of a loss-related traffic control mechanism is not necessary to numerically solve (18). Using the admission control condition $a_V(\theta) \leq a_C(-\theta), \theta \geq 0$

and the monotonicity of the related functions, we can prove that in order to accept a traffic stream the following inequality must be satisfied:

$$\log \gamma_{\text{on}}\big(-\hat{r}\theta^\star + \theta^\star a_V(\theta^\star)\big) + \log \gamma_{\text{off}}\big(\theta^\star a_V(\theta^\star)\big) \le 0,$$

where $\theta^\star = \epsilon/q$, according to (2). This condition simplifies greatly the computational aspects of the loss-related AC scheme.

Calculations for the delay-related AC test are also simple. According to the results of Section 2, given the QoS specification ξ, one must first determine $\theta(\xi)$ in (8) and then check if the left-hand side inequality in (6) holds. Since $u_C\big(-\theta(\xi)\big) = -\xi$, (18) suggests that $\theta(\xi)$ is the unique solution in θ of $f(-\theta, -\xi)$, thus

$$\theta(\xi) = \xi/\hat{r} - (\log \gamma_{\text{on}})^{-1}\big(-\log \gamma_{\text{off}}(\xi)\big)/\hat{r}. \tag{19}$$

This requires only a single evaluation of the function $\gamma_{\text{off}}(\cdot)$ at the argument ξ, keeping the computational complexity low. Moreover, when the payload of the transmitted packets has a constant value P, (16) yields $(\log \gamma_{\text{on}})^{-1}(x) = \hat{r}x/P$, so (19) simplifies further to the closed form solution $\theta(\xi) = \xi/\hat{r} + \log \gamma_{\text{off}}(\xi)/P$. Note that, as long the conditions[1] in the WLAN remain unchanged, a single evaluation of $\theta(\xi)$ suffices to enable an arbitrary number of admission control tests (6), each of them being invoked whenever the mobile station is about to engage a new traffic flow.

4 Tuning the backoff window distributions for improved Effective Capacity

We now investigate appropriate choices of the backoff window distributions employed by the IEEE 802.11 MAC protocol, so as to obtain an Eff. Capacity function greater than the one corresponding to the standard distributions. A greater Eff. Capacity function signifies improved performance.

The mean rate of the On/Off model for an IEEE 802.11 mobile station is

$$\bar{r}_C \triangleq u_C'(0) = a_C(0) = \frac{\hat{r}\mathrm{E}\left[T_{\text{on}}\right]}{\mathrm{E}\left[T_{\text{on}}\right] + \mathrm{E}\left[T_{\text{off}}\right]}. \tag{20}$$

In view of (20), one might attempt to obtain a greater Eff. Capacity by reducing $\mathrm{E}\left[T_{\text{off}}\right]$. As seen in the thesis, this corresponds to reducing the mean window sizes $\mathrm{E}\left[W_i\right]$ at all backoff stages $i \ge 0$. However, when the saturation-based variant of the model is used, (9) indicates that a reduction of the mean window sizes affects the transmission probability τ and the conditional collision probability p, increases contention on the shared channel and may ultimately negate the intended effect, due to the impact of p on $\gamma_{\text{off}}(\cdot)$ through (12) and (13). The same phenomenon occurs also under non-saturated environments and affects the measured values of p, P_{succ}, P_{empty} and P_{coll} employed by the other variant of the model. Similarly, increasing the probability B_0 of sampling a null

[1] Number of active stations in the WLAN and (if the measurement-assisted variant of the model is used), loading conditions at other stations.

window at stage zero decreases $\gamma_{\text{off}}(\cdot)$ through (17), but also results in repeated successful transmissions in other competing stations, indirectly increasing $\gamma_{\text{off}}(\cdot)$ through the third term in (13) and (12).

For the reasons just described, in the following analysis it is assumed that any changes in the backoff window distributions leave the mean window sizes and B_0 invariant, thus also maintain the same value of the mean server rate $\bar{r}_C$. In order to examine the effect of higher order properties of the backoff window distributions, we use a Taylor series expansion of $u_C(\theta)$ around $\theta = 0$ and remember that $u_C(0) = 0$, to obtain

$$a_C(\theta) = u_C(\theta)/\theta = u'_C(0) + \frac{u''_C(0)\theta}{2} + O(\theta^2),$$

for small values of the parameter θ. Employing (18), differentiating twice, setting $\theta = 0$ and remembering that $(\log \gamma_i)'(0) = \mathrm{E}\left[T_i\right]$ and $(\log \gamma_i)''(0) = \mathrm{Var}\left[T_i\right]$ ($i = \text{on}, \text{off}$) leads to

$$u''_C(0) = \frac{\mathrm{Var}\left[T_{\text{on}}\right](\hat{r} - \bar{r}_C)^2 + \mathrm{Var}\left[T_{\text{off}}\right]\bar{r}_C^2}{\mathrm{E}\left[T_{\text{on}}\right] + \mathrm{E}\left[T_{\text{off}}\right]}. \tag{21}$$

Greater values of the Eff. Capacity function for negative arguments correspond to smaller values of $u''_C(0)$. Quantities in (21) relating to the On period do not depend on the backoff window distributions.

The mean value of the Off period can be obtained by differentiating (17) at $\omega = 0$. One gets

$$\mathrm{E}\left[T_{\text{off}}\right] = t_{\text{ov}} + (1 - B_0)(t_{\text{slot}} + \mathrm{E}\left[T_{\text{bc}}\right]), \tag{22}$$

where T_{bc} is the time spent in backoff mode, with moment generator as in (12) and mean equal to

$$\mathrm{E}\left[T_{\text{bc}}\right] = \frac{p}{1-p} t_{\text{coll}} + \mathrm{E}\left[T_s\right]\Big(\frac{\mathrm{E}\left[W_o\right]}{1 - B_0} - 1 + \sum_{l=1}^{\infty} p^l \mathrm{E}\left[W_l\right]\Big). \tag{23}$$

$\mathrm{E}\left[T_s\right] = \gamma'_s(0)$ is the mean of the time needed for the reduction by one of the IEEE 802.11 backoff counter. Equations (22) and (23) verify the earlier claim stating that, when the mean backoff window sizes and the probability B_0 remain invariant, the mean duration of the Off period and, by (20), $\bar{r}_C$ also remain invariant. Thus, in view of (21), the only way of reducing the value of $u''_C(0)$ is through a smaller value of $\mathrm{Var}\left[T_{\text{off}}\right]$. By differentiating twice (17) at $\omega = 0$ and collecting terms,

$$\mathrm{E}\left[T_{\text{off}}^2\right] = B_0 t_{\text{over}}^2 + (1 - B_0)\big((t_{\text{over}} + t_{\text{slot}} + \mathrm{E}\left[T_{\text{bc}}\right])^2 + \mathrm{Var}\left[T_{\text{bc}}\right]\big),$$

so, using also (22),

$$\mathrm{Var}\left[T_{\text{off}}\right] = \mathrm{E}\left[T_{\text{off}}^2\right] - \mathrm{E}\left[T_{\text{off}}\right]^2 = (1 - B_0)\big(B_0(t_{\text{slot}} + \mathrm{E}\left[T_{\text{bc}}\right])^2 + \mathrm{Var}\left[T_{\text{bc}}\right]\big)$$

and reducing $\mathrm{Var}\left[T_{\text{off}}\right]$ can only be achieved by reducing $\mathrm{Var}\left[T_{\text{bc}}\right]$. It is shown in the thesis that a reduction of $\mathrm{Var}\left[T_{\text{bc}}\right]$ may occur only through smaller variances for the backoff window sizes.

We now describe a specific way of adjusting the backoff window distributions, in order to reduce the variance in all backoff stages by a uniform percentage: The range of the standard uniform distributions in (15) is narrowed from $[0, w_i - 1]$ to $\big[(w_i - 1)/2 - 1/2 - \delta_i, (w_i - 1)/2 + 1/2 + \delta_i\big]$. The same mean value is maintained, equal to $(w_i - 1)/2$. The parameter δ_i (a positive integer) is chosen so that the variance of the modified distribution is a fraction $\alpha < 1$ of the standard distribution's variance, so

$$\frac{4(\delta_i + 1)^2 - 1}{12} = \mathrm{Var}\left[W_{i,\text{modified}}\right] = \alpha \, \mathrm{Var}\left[W_{i,\text{standard}}\right] = \alpha \frac{w_i^2 - 1}{12},$$

which yields $\delta_i = \frac{1}{2}\sqrt{\alpha(w_i^2 - 1) + 1} - 1, \quad i \geq 1$. This result is rounded to the nearest integral value.

Extra arrangements must be made for the modification at the 0^{th} backoff stage, because at this stage, besides the mean, one must also preserve B_0, the probability of sampling a null backoff window.

Although that best results are achieved by selecting the smallest possible variances for all backoff stages, this is not an advisable strategy. The reason is that the IEEE 802.11 model has been constructed on the assumption that the competing mobile stations operate independently. Specifically, it is assumed that the collision of an observed station with one or more competing stations does not affect the probability with which this station will collide at the end of the next backoff stage. However, if deterministic backoff windows are used, when two or more stations collide at the end of some backoff stage $i \geq 0$, they will draw the same backoff window at the stage $i + 1$ and collision at the end of all stages from that point onwards will be certain. It follows that the backoff window distributions should retain a sufficient degree of randomness.

5 Validation of the IEEE 802.11 model for tail-related QoS

The Eff. Capacity model has been validated against ns-2 [10] simulation results, under various forms of traffic load and number of competing terminals. For details, see [7, 8]. Here we limit ourselves in two results, in the interest of further highlighting the concepts already discussed. In both of the results the system parameter values correspond to Frequency Hopping Spread Spectrum (FHSS) PHY layer [11]. Also, in all cases, the payload size (see (16)) was chosen constant and equal to $P = 8184$ bits.

Fig. 2 illustrates the accuracy of the Eff. Capacity, by comparing curves of the function (dashed, dotted and solid lines), computed with the use of the saturation-based model, against simulation results (marks).

In the simulation runs used for producing Fig. 2, the values of the Eff. Capacity function were indirectly measured, by feeding a "tagged" IEEE 802.11 station with traffic of known profile, sampling the probability with which the station's buffer exceeded a given threshold and exploiting the linkage (see (5)) between the Eff. Bandwidth, the Eff. Capacity and the probability tail just mentioned. All terminals, besides the tagged one, were operating under saturation conditions. The match between theory and simulations validates the model and indicates its suitability for estimating tail-probabilities or, equivalently, for taking AC decisions in IEEE 802.11 WLANs.

Fig. 3 depicts curves of the queue tail-probabilities versus the tail threshold (in semilog scale) for a network with 10 stations, of which 9 are saturated. The queue of the unsaturated station has been observed under two kinds of traffic load, CBR and Poisson,

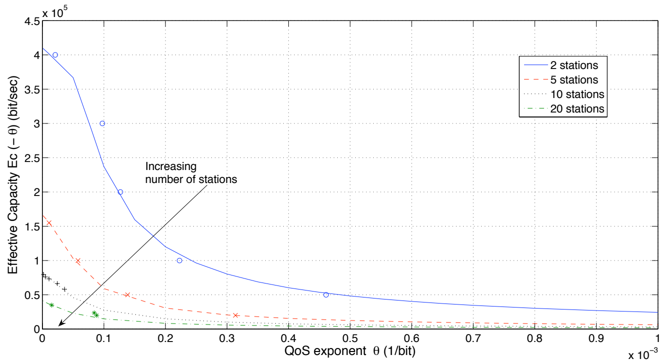

Fig. 2. Curves of $a_C(-\theta)$ vs. θ, for different values of the number of stations n.

both featuring the same mean rate of 79.84 kbps. The slope θ^* of the queue tail for the model-derived curve in each loading case was determined according to the theory (see (5)). As shown in the figure, the simulation-derived queue tails decay exponentially

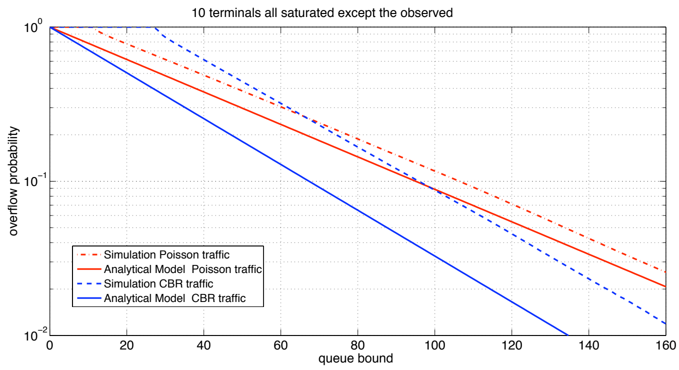

Fig. 3. Modeling and simulation results of queue tail-probabilities for 10 stations, one unsaturated, for two types of load.

and the decay rates agree well with the analytical results.

6 Conclusions

The thesis provided the formal justification for the use of the Eff. Bandwidth/Capacity theory in delay-related performance contexts. Specifically, it was established rigorously that the theory is capable of providing an asymptotically tight approximation to delay tail-probabilities. The thesis also formalized the association of the asymptotic exponential decay rate of the queue content probabilities with its counterpart for the delay probabilities, through the server's Eff. Capacity function. The asymptotic approximation to the delay tail-probabilities was complemented by associated admission control schemes that are useful for providing delay-related QoS guarantees. The general results were applied to the important setting of IEEE 802.11 WLANs, by modeling each IEEE 802.11 station as an On/Off server and then using the Eff. Capacity function corresponding to this model. Computational and algorithmic details relating to the application of the general theory with the particular Eff. Capacity function of this On/Off model were also discussed. Comparison of the analytical results with simulation validated the effectiveness of the On/Off IEEE 802.11 model in providing tail-related QoS guarantees. Finally, the particular form of the Eff. Capacity function of an IEEE 802.11 station suggested an appropriate modification of the backoff window distributions for reduced variance, without affecting the mean backoff window sizes. This modifications results in a greater Eff. Capacity function, hence better performance.

References

1. A. Abdrabou and W. Zhuang. "Stochastic Delay Guarantees and Statistical Call Admission Control for IEEE 802.11 Single-Hop Ad Hoc Networks". *IEEE Trans. Wireless Commun.*, 7(10):3972–3981, October 2008.
2. G. Bianchi. "Performance Analysis of the IEEE 802.11 Distributed Coordination Function". *IEEE JSAC*, 18(3):535–547, 2000.
3. G. Bianchi and I. Tinnirello. "Remarks on IEEE 802.11 DCF Performance Analysis". *IEEE Commun. Lett.*, 9(8):765–767, 2005.
4. C. Chang. "Stability, Queue Length, and Delay of Deterministic and Stochastic Queueing Networks". *IEEE Trans. Autom. Control*, 39(5):913–931, 1994.
5. C. Chang and J. Thomas. "Effective Bandwidth in High-Speed Digital Networks". *IEEE JSAC*, 13(6):1091–1100, 1995.
6. C. Chang and T. Zajic. "Effective bandwiths of departure processes from queues with time varying capacities". In *Proc. IEEE INFOCOM*, pages 1001–1009, 1995.
7. E. Kafetzakis, K. Kontovasilis, and I. Stavrakakis. "Effective Capacity-based Stochastic Delay Guarantees for Systems with Time-Varying Servers, with an Application to IEEE 802.11 WLANs". *Elsevier Performance Evaluation*.
8. E. Kafetzakis, K. Kontovasilis, and I. Stavrakakis. "A Novel Effective Capacity-Based Framework for Providing Statistical QoS Guarantees in IEEE 802.11 WLANs". *submitted to Elsevier Computer Communications*, 2010.
9. K. Kontovasilis and N. Mitrou. "Effective bandwidths for a class of non Markovian fluid sources". In *Proc. ACM SIGCOMM, Computer Communication Review*, volume 27, pages 263–274, 1997.
10. The ns-2 network simulator. www.isi.edu/nsnam/ns, 1998.
11. "IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Nov. 1997. P802.11".
12. G. Vecianna, G. Kesidis, and J. Walrand. "Resource Management in Wide-Area ATM Networks Using Effective Bandwidths". *IEEE JSAC*, 13(6):1081–1090, 1995.
13. D. Wu and R. Negi. "Effective Capacity: A Wireless Link Model for Support of Quality of Service". *IEEE Trans. Wireless Commun.*, 2(4):630–643, 2003.
14. X. Zhang, J. Tang, H. Chen, S. Ci, and M. Guizani. "Cross-Layer-Based Modeling for Quality of Service Guarantees in Mobile Wireless Networks". *IEEE Commun. Mag.*, 44(1):100–106, 2006.