Modeling Approximations for an IEEE 802.11 WLAN under Poisson MAC-Level Arrivals

Ioannis Koukoutsidis[1] and Vasilios A. Siris[1,2]

[1] FORTH-ICS, P.O. Box 1385, 71110 Heraklion, Crete, Greece
[2] Computer Science Department, University of Crete, P.O. Box 2208, 71409
Heraklion, Crete, Greece
{jkoukou,vsiris}@ics.forth.gr

Abstract. We examine two approximative models for the behavior of
an IEEE 802.11 WLAN under Poisson MAC-level arrivals. Both models
extend, in a simple manner, the analysis for saturated stations with the
decoupling approximation. The first follows a well-known approach con-
sidering a constant busy station probability at each idle-sensed slot, equal
to the load of the envisaged single-server queue. We extend the analysis
of this model to calculate more accurately the throughput seen by a sta-
tion in non-saturation conditions, by considering alternating ON/OFF
periods. The second model uses this ON/OFF process structure for cal-
culating attempt and collision probabilities, and subsequently perfor-
mance measures, based on regenerative process theory. By comparison
with simulation results, this model is shown to be more precise for low
load conditions. The accuracy of the modeling approximations is also
studied for a range of values of the minimum contention window, which
is the most influential protocol parameter.

Key words: WLAN, 802.11, mathematical modeling, non-saturation

1 Introduction

Tractable mathematical models of the IEEE 802.11 WLAN protocol operation
with saturated stations (i.e., that always have queued packets to send) consider
a *decoupling approximation*, in which the (re)-transmission processes of the dif-
ferent stations are mutually independent, and the probability of an attempt at
each idle-sensed slot is the same throughout the time evolution of the system.
The corresponding attempt and collision probabilities are then easily retrieved
by the solution of a nonlinear system of equations, as shown in [1–3].

The consideration of saturated wireless stations is more appropriate for data
traffic, as compared to voice traffic (where a codec transmits a fixed payload
periodically), or more generally real-time traffic. Therein it is important to model
the behavior of an unsaturated network in the whole range from low load to near-
saturation conditions, under a certain arrival pattern. The performance of the

network would certainly be different, but also the protocol parameters for which performance is optimal might change.

In this scope, a fair amount of research has been devoted to the modeling of the 802.11 protocol in more realistic, non-saturation conditions, e.g., [4–7]. The bulk of research papers consider Poisson MAC-level arrivals – both for mathematical simplicity and general applicability – and also use the decoupling approximation, assuming additionally a constant busy-station probability at each idle-sensed slot. This is taken equal to the load of the station regarded as a single-server queue. In [5, 6] the load is set as an additional parameter to the problem, that should be measured or estimated. However, a better model should avoid this need and include the characteristics of the traffic arrival distribution in its parameters. This is done in [7] where the authors, while studying voice traffic, extend the set of equations used in saturation conditions to also solve for the load. Then performance measures such as the mean channel access delay and throughput of a station are extracted.

In this paper we revisit the analytical modeling of 802.11 in non-saturation conditions, under Poisson arrivals. We view the aforementioned load model as the predominant modeling approach, and further explore its accuracy. A critical remark is that previous research has considered throughput *as seen by the system* and not *as seen by a station*. We show that these two quantities may differ substantially in non-saturation. We then propose an approximative calculation of the throughput seen by a station, by considering alternating ON/OFF periods. Further, we construct a new model for calculating attempt and collision probabilities, and subsequently performance measures, based on a regenerative process with such periods. This model is shown to be more accurate for low load conditions, which are especially important when considering real-time traffic.

We shall present and compare these modeling approaches, further examining their accuracy in a range of values of the minimum contention window, which is the most influential protocol parameter. We study a homogeneous network with the same parameters for all stations, but the models are readily extendable to multiple service classes, as suggested in the 802.11e standard.

2 Preliminaries

2.1 Protocol Description

The IEEE 802.11 protocol gives specifications for the interconnection of telecommunications equipment in a WLAN, using a CSMA/CA medium sharing mechanism, and exponential backoff upon collisions [8]. Main elements of the mechanism are the method of random slot selection and the backoff algorithm.

The number of idle-sensed slots that a station must wait prior to a packet transmission attempt is selected uniformly in the interval $[0, 2^{(k \wedge m)} \cdot CW_{min} - 1]$, termed as the *contention window*; k is the current backoff stage, initialized to zero for the transmission of a new packet and incremented by one each time a collision occurs, up to a maximum value m. There is no collision detection in the wireless

system and a successful transmission is understood by the proper reception of an acknowledgement (ACK) by the sender. All stations in range of each other are assumed to be synchronized with a perfect carrier sense mechanism so that they all perceive the same sequence of idle and busy slots.

2.2 Analysis in Saturation Conditions

The analysis in saturation conditions is the basis of our investigation, and shall be extended for non-saturation. Consider a number of contending stations N. We assume there does not exist a limit on the number of attempts by a station, after which a packet would be discarded, or that this limit is sufficiently high so that its effect is negligible. Denote the attempt and collision probabilities of a station by p, c, respectively, which are the same for all stations in a homogeneous network. Since the collision probability is dependent on the attempt rate and vice versa, appropriate expressions can construct a system of equations to solve for these values. From a stochastic analysis (either a Markov chain analysis [1], or a renewal theory analysis [2]), the attempt rate can be expressed as a function of the collision probability and the backoff parameters CW_{min}, m, as:

$$p = \frac{2(1-2c)}{(CW_{min}-1)(1-2c) + CW_{min}c(1-(2c)^m)} \; . \tag{1}$$

The collision probability at an attempt by a station is then given by

$$c = 1 - (1-p)^{N-1} \; . \tag{2}$$

We next proceed to derive the main performance measures of interest, which are the mean channel access delay and throughput. The channel access delay is defined as the delay a frame experiences from the time it arrives at the head of the transmission queue until it is successfully transmitted. The following derivation is simpler compared to other approaches, e.g., [7].

The probability of a successful transmission attempt by a station at an idle-sensed slot is $P^{succ} = p(1-p)^{N-1}$, while the probability of a collision is $P^{coll} = p(1-(1-p)^{N-1})$. The probability of no attempt is $1-p$. When the number of elapsed idle-sensed slots until (and including) a success is ω, the mean number of experienced collisions is $(\omega-1)P^{coll}$. Therefore the mean number of collisions experienced by a station until a successful transmission is

$$\sum_{\omega=1}^{\infty} (\omega-1)P^{coll}P^{succ}(1-P^{succ})^{\omega-1} = \frac{P^{coll}(1-P^{succ})}{P^{succ}} \; .$$

Similarly, the mean number of "no-attempt slots" until a success is $(1-p)(1-P^{succ})/P^{succ}$. Therefore the mean channel access delay is given by

$$\mathrm{E}[D^{acc}] = T^{succ} + \frac{P^{coll}(1-P^{succ})}{P^{succ}}T^{coll} + \frac{(1-p)(1-P^{succ})}{P^{succ}}\mathrm{E}[S^{(N-1)}] \; , \tag{3}$$

where T^{succ}, T^{coll} are the durations of a successful transmission (including the reception of the acknowledgement), and a transmission which resulted in a collision, respectively. $S^{(N-1)}$ is the duration of a slot as perceived by a station which does not attempt in that slot; it is an i.i.d r.v standing for the duration of a generic slot (i.e., either an idle slot, a successful transmission or a collision), when $N - 1$ stations are considered in the system. Hence

$$\mathrm{E}[S^{(N-1)}] = (1-p)^{N-1}T^{slot} + (N-1)p(1-p)^{N-2}T^{succ} \\ + [1 - (1-p)^{N-1} - (N-1)p(1-p)^{N-2}]T^{coll} \; . \tag{4}$$

The throughput *seen by a station* equals the rate of successfully transmitted MAC-level information per unit of time. We calculate it by considering that each end of a successful transmission is a renewal epoch for the system. Assuming a fixed MAC packet size σ and denoting the throughput by γ, we have

$$\gamma = \frac{\sigma}{\mathrm{E}[D^{acc}]} \; . \tag{5}$$

This is opposed to $\gamma = \sigma P^{succ}/E[S^{(N)}]$, mostly derived as a measure in the literature (e.g., [1, 5–7]), which can only be stated as the individual throughput *seen by the system*.

3 Modeling Approximations in Non-Saturation

Our setting consists of contending stations transmitting to an access point, whose only purpose is to send back ACKs. We consider the same arrival pattern for all stations. No characterizations are made about the stability of the system, as the memory-full nature of the stochastic model makes this extremely difficult. Instead we assume that there is a range of arrival rates for which the system will be stable, in the sense that the queue size of every station remains finite. This assumption is in accordance with our simulation results, as well as numerous other experimental results in the literature which confirm stability for a range of arrival rates (e.g., [4, 5]).

3.1 Load Model

The first model assumes a constant instead of a time-dependent busy station probability at each idle-sensed slot. This is taken equal to the load of an envisaged single server queue, which generally models the arrival and departure sequence at the station.[3]

Denoting the load by ϱ, the probability of a collision at an attempt by a station becomes

$$c = 1 - (1 - \varrho p)^{N-1} \; , \tag{6}$$

---

[3] It is worth noting that this modeling would be correct if stations were fed by a stationary (not only possessing stationary increments) arrival process and their queues were stable.

where

$$\varrho = \lambda \mathrm{E}[D^{acc}] \tag{7}$$

for an arrival rate λ. The probability of an attempt p (given the station has a packet to transmit) depends only on the collisions, and hence is given again by (1). The mean channel access time can then be derived by following the analysis in Sect. 2.2, where in place of p one uses the unconditional probability of an attempt, ϱp. Thus we have a system of equations to solve for p, c, ϱ.

To calculate the throughput experienced by a station, we would then have to consider a sequence of alternating ON/OFF periods, where the OFF period is geometrically distributed with parameter ϱ (and may take the value 0). Treating this as a regenerative process, we would have $\gamma = \sigma/(\mathrm{E}[D^{acc}] + (1 - \varrho)/\varrho \cdot \mathrm{E}[S^{(N-1)}])$. This turns out to be extremely inaccurate, yielding throughput results of about 1 order of magnitude greater. To tackle this, we shall make use of the throughput calculation in the more involved ON/OFF model of the next subsection.

3.2 Regenerative ON/OFF Model

Consider a probabilistic ON/OFF model, where durations of the ON and OFF periods are geometrically distributed. After a frame transmission in the ON period, a station "rests" with probability r_{ON} or has another frame to send and contends again with probability $1 - r_{\mathrm{ON}}$. The OFF period consists of a random number of generic slots, i.e., idle slots, successful transmission or collision times. At the end of each such slot, a station may keep resting with probability r_{OFF}, or have a frame to send and become busy with probability $1 - r_{\mathrm{OFF}}$.
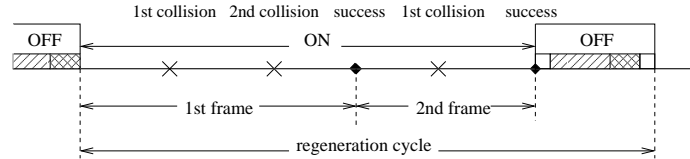

Fig. 1. Sample path evolution of the system with ON and OFF periods. The end of an OFF period is a regeneration epoch for the system

Fig. 1 illustrates a sample path of the system evolution. Clearly we have described a regenerative process and the time between the ends of successive OFF periods is a regeneration cycle.

Then, as in the renewal case [2], we can derive time averages. The attempt probability for a station will be

$$p = \frac{\mathrm{E}[R]}{\mathrm{E}[X]} ,$$

where $\mathrm{E}[R]$ is the mean number of attempts in the regeneration cycle and $\mathrm{E}[X]$ the mean number of *attempt opportunities* in this cycle. It is obvious that the mean number of attempts until a frame is transmitted has the same expression as in the saturation case. For a collision probability c, we have

$$\mathrm{E}[R] = \frac{1}{r_{\mathrm{ON}}(1-c)} \ .$$

Now denote the mean OFF period by θ, measured in attempt opportunities. The mean number of attempt opportunities in a regeneration cycle is

$$\mathrm{E}[X] = \theta + \frac{\mathrm{E}[\bar{X}]}{r_{\mathrm{ON}}} \ ,$$

with $\bar{X}$ being the number of elapsed backoff slots until a frame is transmitted, which also has the same expression as in saturation. Recovering the appropriate expression and substituting $\theta = 1/(1-r_{\mathrm{OFF}})$, we finally find

$$p = \frac{2(1-2c)}{2\frac{r_{\mathrm{ON}}}{1-r_{\mathrm{OFF}}}(1-c)(1-2c) + (CW_{min}-1)(1-2c) + CW_{min}c(1-(2c)^m)} \quad (8)$$

The inherent deficiency of the model is the assumption of a constant busy station probability after each successful transmission of a frame, while in reality it increases for each consecutive transmission. Here we take r_{ON} to be the probability that no packet arrives in the interval $[0, D^{acc})$. The error in this approximation is smaller in low loads, where consecutive frame transmissions are less frequent.

The distribution of D^{acc} is difficult to derive, and does not result in a simple expression. Instead we replace it with a constant value, equal to its mean. Considering Poisson arrivals with rate λ, we have

$$r_{\mathrm{ON}} = e^{-\lambda \mathrm{E}[D^{acc}]} \ . \quad (9)$$

The value of r_{OFF} is the probability that no packet arrives in $[0, S^{(N-1)})$. It can easily be calculated exactly, since $S^{(N-1)}$ takes discrete values:

$$\begin{aligned}
r_{\mathrm{OFF}} =&(1-p)^{N-1}e^{-\lambda T^{slot}} + (N-1)p(1-p)^{N-2}e^{-\lambda T^{succ}} \\
&+ [1-(1-p)^{N-1}-(N-1)p(1-p)^{N-2}]e^{-\lambda T^{coll}} \ .
\end{aligned} \quad (10)$$

Equations (8), (9), (10), along with $c = 1 - (1-p)^{N-1}$ can be used to solve for the unknowns p, c, r_{ON}, r_{OFF}.

Based on the derived attempt and collision probabilities, the mean channel access delay is calculated following the same approach as in Sect. 2.2. The throughput is calculated as the mean MAC-level information transmitted in a regeneration cycle, over the mean duration of this cycle:

$$\gamma = \frac{\sigma/r_{\mathrm{ON}}}{\mathrm{E}[D^{acc}]/r_{\mathrm{ON}} + \mathrm{E}[S^{(N-1)}]/(1-r_{\mathrm{OFF}})} \ . \quad (11)$$

We shall employ the same formula for calculating the throughput in the load model of Sect. 3.1, using the derived quantities $\mathrm{E}[D^{acc}]$, $\mathrm{E}[S^{(N-1)}]$ from the model, and replacing p with ϱp.

Remark 1. It should be clear that the ON/OFF periods here are just a work-around and do not refer to the arrival pattern. We briefly explain how such a pattern can be modeled in the Appendix.

4 Results

We note that it is very difficult to show existence and uniqueness of the solution of the nonlinear system of equations in the above models, despite their relatively simple form. The mapping f in the constructed multidimensional equation $x = f(x)$ does not possess attractive monotonicity and concativity properties. We solve numerically for the unknowns using MATLAB's `fsolve` function.

Test cases are taken for *basic channel access*, noting that the method of RTS/CTS messages [8] that mitigates the hidden terminal problem becomes less efficient for unsaturated stations. According to protocol timings, the duration the medium is busy because of a successful transmission is

$$T^{succ} = DIFS + \delta + T_{PLCP} + \frac{\sigma}{R} + \delta + T_{PLCP} + SIFS + \frac{ack}{R_b} \, , \qquad (12)$$

while the duration of a collision is

$$T^{coll} = DIFS + \delta + T_{PLCP} + \frac{\sigma}{R} \, . \qquad (13)$$

In the above formulae, $DIFS$ (Distributed Inter-Frame Space) is the idle-sensed time that must elapse prior to a transmission attempt, δ the propagation delay, T_{PLCP} the time to transmit the PLCP preamble and header (adjoined by the physical layer) and ack the size of a MAC level acknowledgement. ACK packets are always transmitted at a (usually lower) basic service rate R_b, whereas the T_{PLCP} is fixed for each physical layer configuration. In addition, a $SIFS$ (Short Inter-Frame Space) interval is used between the transmission of a frame and the sending of an acknowledgement, to allow the MAC layer to receive the packet and subsequently the transceiver to turn around.

We consider an 802.11a implementation, for R, R_b equal to 6 Mbps and MAC packet size 160 bytes. Default values for MAC and physical layer parameters are taken from [9] and shown in Table 1.

We compare the numerical results with those of a discrete-event simulator written in C++, which closely follows the modeled setup. The load model is also referred to as Model 1, and the regenerative ON/OFF model as Model 2.

We first present in Fig. 2 a comparison of different throughput approximations in a range of Poisson arrival rates until saturation, based on Model 1. We have calculated the corresponding individual throughput "seen by the system", which is shown to be highly inaccurate for non-saturation conditions. On the

Table 1. 802.11a MAC and physical layer parameters

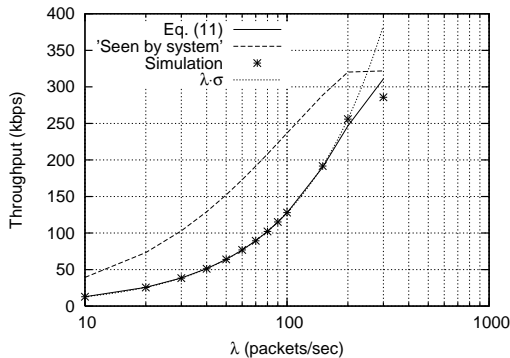| Parameter	Value
T_{slot}	9 μs
$SIFS$	16 μs
$DIFS$	34 μs
T_{PLCP}	20 μs
ack	14 bytes

Fig. 2. Comparison of different throughput approximations based on Model 1 in a case with $CW_{min} = 32$, $m = 5$ and $N = 10$ stations, for different Poisson arrival rates until saturation

other hand, what is also shown on the graph is that the approximation by $\lambda\sigma$ is extremely accurate for a very large range of loads, reflecting the simple fact that when collisions are fewer, the throughput is almost equal to the information arrival rate.

Results on the comparison of the two methods are presented in Table 2. For cases where the load obtained by the numerical evaluation is smaller than 1 and thus the system appears stable, we calculate the associated queueing delay. The total packet transmission delay is, for an M/G/1 queue,

$$\mathrm{E}[D^{total}] = \mathrm{E}[D^{acc}] + \frac{\lambda\mathrm{E}[(D^{acc})^2]}{2(1-\varrho)} \ . \tag{14}$$

For the calculation of the second moment of the channel access delay, see e.g. [7]. If the load turns out greater than 1, results are shown for saturation.

Inaccuracies manifest themselves more explicitly in the estimation of mean delay, whose units are on a finer scale. We see that Model 1 overestimates delay for low or moderate arrival rates, and underestimates it for high loads. This reveals the nature of the average approximation. On the other hand, Model 2 is more accurate for lower loads but performs significantly poorer for higher load values, and fails to predict instability cases well.

Table 2. Comparison of the two models

(a) $CW_{min} = 32$, $m = 5$, $N = 5$

λ (pkts/s)	Load	E[D^{acc}] (ms) Analysis Model 1	Model 2	Simulation	E[D^{total}] (ms) Analysis Model 1	Model 2	Simulation	Throughput (kbps) Analysis Model 1	Model 2	Simulation
10^2	0.04	0.537	0.485	0.441	0.553	0.498	0.452	127.74	127.79	127.80
$2\cdot10^2$	0.11	0.644	0.519	0.542	0.699	0.552	0.591	253.56	254.39	255.58
$3\cdot10^2$	0.22	0.811	0.576	0.734	0.968	0.643	0.934	372.62	377.94	383.70
$4\cdot10^2$	0.46	1.121	0.680	1.152	1.678	0.829	2.460	468.79	493.70	511.94
$5\cdot10^2$	1.09	2.010	0.899	2.185	∞	1.342	∞	636.74	585.92	585.40
$6\cdot10^2$	1.31	2.010	2.010	2.186	∞	∞	∞	636.74	636.74	585.40

(b) $CW_{min} = 32$, $m = 5$, $N = 10$

λ (pkts/s)	Load	E[D^{acc}] (ms) Analysis Model 1	Model 2	Simulation	E[D^{total}] (ms) Analysis Model 1	Model 2	Simulation	Throughput (kbps) Analysis Model 1	Model 2	Simulation
10^2	0.06	0.678	0.527	0.576	0.706	0.543	0.604	127.59	127.75	127.88
$2\cdot10^2$	0.33	1.382	0.728	1.674	1.705	0.801	3.163	246.48	252.94	255.80
$3\cdot10^2$	1.34	4.119	2.305	4.479	∞	5.443	∞	310.78	320.59	285.74
$4\cdot10^2$	1.79	4.119	4.119	4.479	∞	∞	∞	310.78	310.78	285.79

The two models are generally more precise in the estimation of the throughput. It should be stressed that – in contrast to delay measurements – the throughput measurement scale of kbps is a coarse scale when protocol timings are in μs. Notice, however, that the throughput expression used in Model 1 yields a greater error for highly loaded but not saturated stations.

We further proceed to modify the minimum contention window CW_{min}, which is the most influential protocol parameter. The maximum backoff stage has a much smaller influence, especially in non-saturation. We take numerical results for both models in a wide range of CW_{min} values. In Fig. 3 we plot simulated and best approximative analytical values for the mean channel access delay, as a more sensitive metric to changes of CW_{min}, compared to the mean total delay or throughput. Filled points are used to discern cases for which Model 2 was more accurate; as shown in the figure, this occurrs for a range of low loads (note that for fixed other parameters the load increases or decreases with CW_{min} proportionally to E[D^{acc}]). Inaccuracies are more pronounced in both models as CW_{min} increases, but except for exorbitantly large CW_{min} values, the system's behavior is accurately followed.

Qualitative aspects of behavior are well captured by both models. We saw in Table 2 that both models demonstrate the occurrence of instability. Instability may also occur for very large values of CW_{min}, because of the increased waiting time prior to an attempt. This is confirmed by the numerical and simulation results in Fig. 3, for $\lambda = 3\cdot10^2$ and CW_{min} equal to 512 and 1024 (the corresponding values coincide with the ones of the saturation case). The most
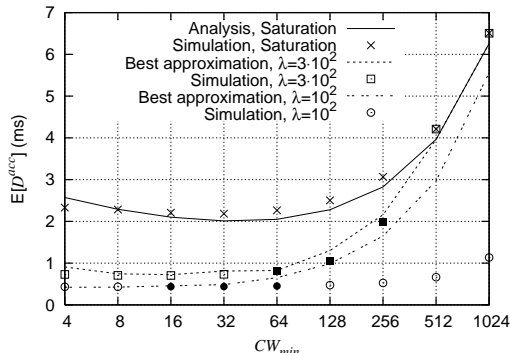
Fig. 3. Mean channel access delay for different CW_{min} values, and load conditions (N=5 stations, $m = 5$). Filled points refer to cases for which Model 2 was more accurate

important qualitative aspect concerns the decrease of the optimal CW_{min} value as the arrival rate decreases. The optimal CW_{min} equals 32 in the saturation case, for $N = 5$ stations, while for $\lambda = 3 \cdot 10^2$ packets/sec it is equal to 8 and for $\lambda = 10^2$ it falls down to 4, as shown in Fig. 3.

5 Concluding Remarks

We have examined modeling approximations for the 802.11 protocol in non-saturation conditions, with respect to the accuracy of performance evaluation measures. A regenerative ON/OFF structure was used to calculate more accurately the actual throughput experienced by a station. This can be used as a simple extension to the load model, or as a stand-alone model in low load conditions. These modelings can also be extended to the case of multiple service classes suggested in 802.11e, based on modifications of the saturation analysis shown e.g. in [3,7].

References

1. Bianchi, G.: Performance analysis of the IEEE 802.11 distributed coordination function. IEEE J. Select. Areas Commun. **18**(3) (March 2000) 535–547
2. Kumar, A., Altman, E., Miorandi, D., Goyal, M.: New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs. In: Proc. IEEE Infocom 2005, Miami, FL, USA (March 2005) 1550–1561
3. Ramaiyan, V., Kumar, A., Altman, E.: Fixed point analysis of single cell IEEE 802.11e WLANs: Uniqueness, multistability, and throughput differentiation. In: Proc. ACM Sigmetrics 2005, Banff, Canada (June 2005) 109–120
4. Clifford, P., Duffy, K., Foy, J., Leith, D., Malone, D.: Modeling 802.11e for data traffic parameter design. In: Proc. IEEE WiOpt 2006, Boston, MA, USA (April 2006) 28–37

5. Engelstad, P., Østerbø, O.: Non-saturation and saturation analysis of IEEE 802.11e EDCA with starvation prediction. In: Proc. ACM/IEEE MSWiM 2005, Montreal, Quebec, Canada (October 2005) 224–233
6. Ergen, M., Varaiya, P.: Throughput analysis and admission control for IEEE 802.11a. Mobile Networks and Applications **10**(5) (2005) 705–716
7. Hegde, N., Proutière, A., Roberts, J.: Evaluating the voice capacity of 802.11 WLAN under distributed control. In: Proc. IEEE LANMAN 2005, Chania, Greece (September 2005)
8. IEEE Computer Society: IEEE Std 802.11. (1999)
9. IEEE Computer Society: IEEE Std 802.11a-1999 (R2003). (2003)
10. Wolff, R.: Stochastic Modeling and the Theory of Queues. Prentice-Hall, Englewood Cliffs, NJ (1989)

Appendix: ON/OFF arrivals

Let X_{ON}, X_{OFF} denote the durations of the ON and OFF periods respectively. We will calculate the mean number of frames sent successively in an ON period, m_{ON}, and the mean number of attempt opportunities in an OFF period, m_{OFF}, only relying on expectations of X_{ON}, X_{OFF}.

If M packets are sent in an ON period, the total time to send these is $X_{\mathrm{ON}} = \sum_{j=1}^{M} D_j^{acc}$. Employing Wald's theorem, we have

$$m_{\mathrm{ON}} = \frac{\mathrm{E}[X_{\mathrm{ON}}]}{\mathrm{E}[D^{acc}]} \ . \tag{15}$$

Now suppose L generic slots occur in an OFF period, the total duration of which is $T_{\mathrm{OFF}} = \sum_{j=1}^{L} S_j^{(N-1)}$. Then $T_{\mathrm{OFF}} = X_{\mathrm{OFF}} + R_L$, where R_L is the residual time of the last generic slot which begun before the end of the OFF period. If we assume that the end of the OFF period occurs randomly within the last generic slot, we approximately have $\mathrm{E}[R_L] = \frac{\mathrm{E}[(S^{(N-1)})^2]}{2\mathrm{E}[S^{(N-1)}]}$ (see e.g., [10]).

Then from $T_{\mathrm{OFF}} = X_{\mathrm{OFF}} + R_L$, taking expectations in both sides, dividing by $\mathrm{E}[S^{(N-1)}]$, and employing Wald's theorem, we have

$$m_{\mathrm{OFF}} = \frac{\mathrm{E}[X_{\mathrm{OFF}}]}{\mathrm{E}[S^{(N-1)}]} + \frac{\mathrm{E}[(S^{(N-1)})^2]}{2(\mathrm{E}[S^{(N-1)}])^2} \ . \tag{16}$$

It is noted that in the case where X_{OFF} is considerably larger than the time of a generic slot the residual term can be removed without significant error.

Following the notations in Sect. 3.2, the attempt probability for a station writes

$$p = \frac{\mathrm{E}[R]}{\mathrm{E}[X]} = \frac{m_{\mathrm{ON}} \cdot \frac{1}{(1-c)}}{m_{\mathrm{ON}} \cdot \mathrm{E}[\bar{X}] + m_{\mathrm{OFF}}} \ . \tag{17}$$

As always, the collision probability is $c = 1 - (1-p)^{(N-1)}$, so that a system of equations is constructed to solve for p, c.