Modelling Delay on IEEE 802.11 MAC Protocol for Unicast and Broadcast Non-saturated Traffic

Rodolfo Oliveira
FCT, Universidade Nova de Lisboa
P-2829-516 Caparica
Portugal
rado@fct.unl.pt

Luis Bernardo
FCT, Universidade Nova de Lisboa
P-2829-516 Caparica
Portugal
lflb@fct.unl.pt

Paulo Pinto
FCT, Universidade Nova de Lisboa
P-2829-516 Caparica
Portugal
pfp@fct.unl.pt

Abstract –The existing models for IEEE 802.11 DCF networks only consider unicast frames, ignoring the existence of broadcast traffic. In a real scenario, the stations are most of the times non-saturated and unicast and broadcast frames exist. These specific characteristics influence the service time which consequently affects the queue behaviour. In this paper, we model the total frame's delay for IEEE 802.11 DCF networks in presence of both unicast and broadcast traffic. Our theoretical analysis proposes a model able to compute the time between the instant when a new frame is inserted on the transmission queue and the instant when its transmission finishes. The time needed to serve a frame (service time) is formally deduced from the IEEE 802.11 transmission procedure, conceptually using the view of one network's node. We use an M/M/1/K queue model in order to model each frame's queueing delay. We validate the total frame's delay using several simulations and present some results. These are analysed for different scenarios of broadcast/unicast network loads and different number of nodes.

I. INTRODUCTION

This paper focus on IEEE 802.11 modelling and assumes a single-hop DCF (Distributed Coordination Function) [6,7] network. [1] provides a bound for the frame transmission probability and the network throughput in the condition of unicast traffic saturation. In [2] we proposed an extension to [1] that considers both unicast and broadcast traffic but only in saturated conditions.

Several non-saturated network models for finite unicast load were presented recently. For non-saturated networks it is not assumed that a station has always a frame to transmit. [3] presents two queueing models for IEEE 802.11 LANs based on a system centric and a user centric view of the network. The model considers unicast traffic in basic mode. [3] assumes that on average all the other users send one frame successfully between two successful transmissions of a chosen user. This is acceptable for the conditions assumed in [3] but it is not valid in the presence of both unicast and broadcast traffic. For broadcast frames the backoff contention window size is never increased while for unicast traffic it can be, which origins higher transmission probability for broadcast frames than for unicast (specially when the network load is high). [5] proposes a queueing model using the average backoff window size as a contention time measurement. This approach does not consider the case when a frame arrives on an empty queue and a pos-transmission backoff is still running. In this case, once this backoff is exhausted the frame is transmitted with the consequence of reducing the overall backoff window size. Furthermore the model is not extensible to broadcast traffic, where a fixed backoff window size is used. [4] proposes a model for multi-rate traffic which studies the throughput and fairness for saturated and non-saturated traffic.

In order to integrate broadcast and unicast traffic in non-saturated networks we extended our work in [2] which considers some of the remarks pointed above. [2] only considers saturated traffic and here we generalize it to treat saturated and non-saturated traffic. Our theoretical analysis proposes a model that considers the *pre* and *pos* backoff periods only existent in non-saturated traffic. It is also able to compute the time between the instant when a new frame is inserted on the transmission queue and the instant when its transmission finishes. As far as we know, this is the first study that analyses the influence of broadcast transmissions in the unicast performance considering non-saturated traffic.

The next section gives a brief overview of IEEE 802.11 features to model. Section III introduces the Markov chain model used to obtain the service time. The entire delay sensed by a frame before being effectively transmitted is stated in section IV. Our proposals are validated using a network simulator and the results are illustrated in section V. Finally, some conclusions are drawn in section VI, where further work is also discussed.

II. 802.11 BRIEF OVERVIEW

Each time a new frame is generated at the network layer, it is inserted in the transmission queue where it waits to be served by the MAC protocol. This is the **queue waiting time**. The time between the instant the MAC protocol starts to serve the frame until the time it is ready to serve another one is named **virtual service time** in this paper (the term virtual is used because of the influence of the pos-transmission backoff which extends the real service time). **Total time** is the sum of these two times.

When the MAC layer receives a request to transmit a frame, it starts running the algorithm illustrated in Fig.1 (the execution starts from the shadowed decision box). The standard [6] defines one initial backoff (pre-backoff) before the frame transmission, and another one when the transmission finishes (pos-backoff). For consecutive transmissions, the pos-backoff of the *k-1* frame is applied as a pre-backoff of the *k* frame replacing the real pre-backoff. Thus for consecutive frame transmissions only one backoff contention is applied during the virtual service time, and the station will be only at the states 1, 2 and 3. For a new frame arriving on an empty queue, if the last pos-backoff has expired, the virtual service time will be composed by pre-backoff, transmission and pos-backoff (MAC can be in states 4, 5, 2 or 3).

This work was supported in part by FCT/MCES (Portuguese Science and Technology Foundation) under the project SIGAPANO POSC/EIA/62199/2004 and the grant SFRH/BD/25284/2005.

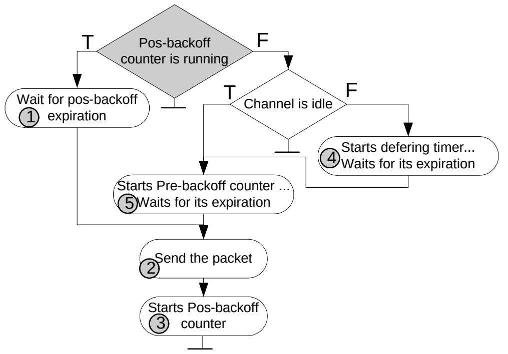

Figure 1. IEEE 802.11 MAC algorithm to transmit a frame.

For a unicast transmission, when a backoff is started the MAC selects uniformly a backoff value from the interval $(0, W_i - 1)$, where W_i is the current backoff stage contention window size. The backoff counter (BC) is decremented each time the channel is detected to be idle for an interval of a slot time. I.e., when busy slots are sensed, the BC is frozen. It is only decremented again after the channel is idle for a DIFS time interval (equals to 2.5 slots). A 2-way handshaking (basic mode) or a 4-way handshaking (RTS/CTS) can be used, depending on the frame's size to transmit. The backoff mechanism uses multiple backoff stages for a unicast transmission. When the BC expires at a given stage, the station tries a new transmission. As data transmission is acknowledged, a station retransmits the frame if the acknowledge is not received. When a retransmission is made, a station doubles the backoff window size, except for the last stage of backoff. If the retransmission at the last backoff stage is unsuccessful, the frame is discarded.

For a broadcast transmission, only a single stage of backoff is used. When the BC expires, the data is sent without being acknowledged, which can lead to unrecoverable collisions if one or more stations start transmitting on the same slot.

III. MODEL ASSUMPTIONS

The proposed model assumes a single-hop network where there are no hidden terminals present. It models the basic unicast and the broadcast transmission schemes, but the model is extensible to unicast transmissions using RTS/CTS handshaking. Non-saturated traffic is assumed. Each station transmits a broadcast or unicast frame according to a uniform probability density function, with probabilities p_b or $p_u = 1 - p_b$, respectively.

The Markov chain illustrated in Fig. 2 models the possible transmission states of each station. Without having frames to transmit a station is initially at state χ_{start}, and returns there after the pos-backoff, when the queue is empty. p_{QE} and p_{QNE} represent the probabilities of having at a given instant an empty or non-empty transmission queue.

The markov states represented by $\chi_{i,j} : 1 \le i \le m, 0 \le j \le W_i - 1$, denote the BC states used in a transmission. j expresses the BC state in the i^{th} backoff stage. When the BC reaches a state $\chi_{i,0} : i = 1$ a broadcast or unicast transmission can be started. The states $\chi_{i,0} : i > 1$ are only used for unicast transmissions. It is assumed that each frame collides with constant and independent probability p_c, and is successful with probability $p_s = 1 - p_c$. A represents the total flow of chains starting a new transmission. If we assume m backoff stages for unicast, then according to the standard [6], we have

$$\begin{cases} W_i = 2^{i-1} W_1, & 0 < i < m \\ W_i = W_{i-1}\,, & i = m \end{cases} \quad . \tag{1}$$

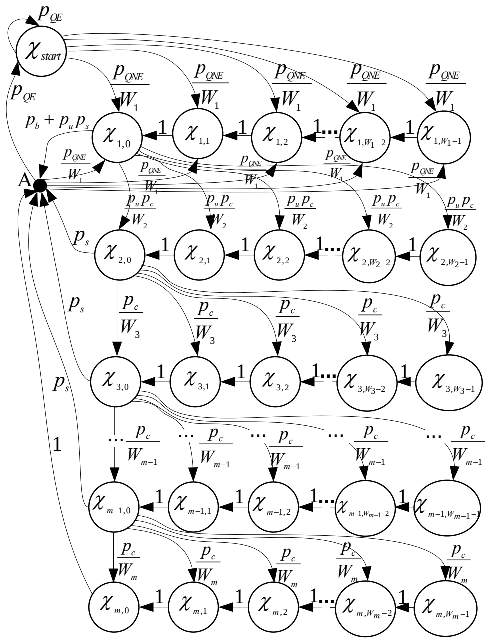

Figure 2. Markov chain model for a single station

IV. TOTAL DELAY

A. Virtual Service Time Component

From the model illustrated in Fig. 2, we know that

$$\chi_{start} = \frac{p_{QE}}{p_{QNE}} A\,, \tag{2}$$

and

$$\begin{cases} \chi_{i,W_i-k} = \dfrac{k}{W_i} A, & i = 1 \\ \chi_{i,W_i-k} = \dfrac{k}{W_i} p_u p_c^{\,i-1} A, & 2 \le i \le m \end{cases} \quad . \tag{3}$$

Using the normalization condition

$$\sum_{i=1}^{m} \sum_{j=0}^{W_i - 1} \chi_{i,j} + \chi_{start} = 1\,, \tag{4}$$

the parameter A is defined as

$$A = \left[\frac{w+1}{2} + \frac{p_{QE}}{p_{QNE}} + p_u \sum_{i=2}^{m} \frac{W_i + 1}{2} p_c^{\;i-1} \right]^{-1}, \qquad (5)$$

being the probabilities of starting a new broadcast or unicast transmission respectively given by:

$$\begin{cases} \chi_b = p_b \chi_{1,0} = p_b A \\ \chi_u = p_u \chi_{1,0} + \sum_{i=2}^{m} \chi_{i,0} = p_u A \dfrac{1-p_c^m}{p_s} \end{cases}. \qquad (6)$$

Generalizing, the probability of starting a new transmission is:

$$\chi = \chi_u + \chi_b = \left(p_b + p_u \frac{1-p_c^m}{p_s} \right) A . \qquad (7)$$

Assuming a network composed by n stations, the probability of a successful transmission p_s is related with the probability of $n-1$ stations not beginning a transmission on the same slot:

$$p_s = (1-\chi)^{n-1}. \qquad (8)$$

Each node's queue has a finite capacity. Let's consider that the maximum number of frames in the node's system is K. One of the frames is being served by the MAC which implies a maximum queue length of $K-1$. An arriving frame enters the queue if it finds fewer than K frames in the node's system and is lost otherwise. Considering the frame's interarrival time exponential with parameter λ, and as hypothesis, the queue's service time also exponential, an M/M/1/K queue model can be used. The K parameter plays an important role in this model, because the model considers the frame's arrival rate as function of the queue capacity

$$\lambda_k = \begin{cases} \lambda, & 0 \le k < K \\ 0, & k \ge K \end{cases}. \qquad (9)$$

If $k \ge K$, the model will assume a null frame generation which results in an effective average frame's arrival rate λ' lower than the initially specified λ.

Using the M/M/1/K queueing model, the probability of having zero elements in the queue is given by [9]:

$$p_{QE} = \frac{1-\rho}{1-\rho^{K+1}}, \qquad (10)$$

Representing the average virtual service time by T_S, the parameter ρ is defined as

$$\rho = \lambda T_S . \qquad (11)$$

T_s is the time needed by the MAC protocol to transmit one frame, including a probability of containing a pre backoff period if the pos backoff of the previous frame has finished (see (15) and (16)). This time is composed by one or more contention periods and one or more frame transmission tries.

The contention period depends on the average time waiting on each backoff state, which is given by

$$T_\chi = \sum_{k=1}^{n} \binom{n-1}{k-1} p_{QNE}^{k-1} p_{QE}^{n-k} \Big[(1-\chi)^{k-1}\sigma + p_{b_s}(k) T_{b_s} + p_{u_s}(k) T_{u_s} + p_{b_c}(k) T_{b_c} + p_{u_c}(k) T_{u_c} + p_{m_c}(k) T_{m_c} \Big] . \qquad (12)$$

Equation (12) assumes that each station's empty queue probability is independent from the other ones. In (12) σ is the duration of an idle slot time. T_{b_s} and T_{u_s} are the average contention times sensed at the station due to a successful transmission from another station of a broadcast or a unicast frame, respectively. T_{b_c}, T_{u_c} and T_{m_c} are respectively, the contention time felt by a station due to a collision with broadcast transmissions, unicast transmissions, or both transmissions. p_{b_s} and p_{u_s} are defined as the probabilities of (any station except ours) having a successfully transmission of a broadcast or unicast frame, respectively. p_{b_c} and p_{u_c} represent the probabilities of only broadcast or only unicast frames colliding. Finally the probability of having a collision with only mixed broadcast and multicast transmissions is given by p_{m_c}. All these parameters appearing in (12) are defined in appendix A.

The average contention time due to a backoff procedure frame transmission is the sum of the average times contending on each backoff state weighed by the probability of being on each one. For the first backoff stage with W_1 states the average contention time is

$$T_{C_1} = \sum_{k=0}^{W_1-1} \frac{k}{W_1} T_\chi = \frac{W_1 - 1}{2} T_\chi . \qquad (13)$$

The time expressed in (13) exists for both unicast and broadcast frames transmission. But when unicast frame transmissions are considered, the average contention time increases due to the multiple backoff stages used at different transmission retries. The average contention time spent in the other backoff stages $(i \ge 2)$ is only applied to unicast transmissions, and it includes the frame transmission time T_{u_s} spent on each transmission retry:

$$T_{C_r} = \sum_{i=2}^{m} p_c^{i-1} \left(\frac{W_i - 1}{2} T_\chi + T_{u_s} \right). \qquad (14)$$

The average virtual service time for broadcast frame transmission is given by:

$$T_{S_b} = T_{pre\_b} + T_{b_s} + T_{pos\_b}\,, \quad (15)$$

where $T_{pre\_b} = p_{QE}\left[DIFS + T_{C_1} + p_c T_{defer}\right]$, T_{b_s} represents the time needed for frame transmission, and $T_{pos\_b} = T_{C_1}$.

For unicast frame's transmission the average virtual service time is similarly defined as

$$T_{Su} = T_{pre\_u} + T_{u_s} + T_{pos\_u}\,, \quad (16)$$

being $T_{pre\_u} = p_{QE}\left[DIFS + T_{C_1} + T_{C_r} + p_c T_{defer}\right]$ and $T_{pos\_u} = T_{C_1} + (1 - p_{QE})T_{C_r}$.

The average virtual service time is given by

$$T_S = p_b T_{S_b} + p_u T_{S_u}\,. \quad (17)$$

Finally (7), (11) and (17) form a system of three non-linear equations being T_S, ρ and χ unknown, which can be numerically solved.

B. *Frame queue waiting time*

Following Little's theorem, the average queue waiting time in steady state is given by

$$T_q = \frac{L - (1 - p_{QE})}{\lambda'}\,. \quad (18)$$

λ' represents the effective frame load, deducting the frame loss rate due to queue overload (9), and can be written as [9]

$$\lambda' = \left[1 - \frac{(1-\rho)\rho^K}{(1-\rho^{K+1})}\right]\lambda\,. \quad (19)$$

L represents the mean number of frames in the queue, and is given by [9]

$$L = \begin{cases} \dfrac{\rho\left[1-(K+1)\rho^K + K\rho^{K+1}\right]}{(1-\rho)(1-\rho^{K+1})}, & \text{if } \rho \neq 1 \\ \dfrac{K}{2}, & \text{if } \rho = 1 \end{cases}. \quad (20)$$

Finally, the mean total delay per frame is expressed by

$$T_d = T_S + T_q\,. \quad (21)$$

V. MODEL VALIDATION

We used the ns-2 [8] simulator version 2.28 to validate our model. Simulations were made considering all frames having the same size ($\mathrm{E}(P) = 960\mu s$), and the parameters presented in Table 1.

Fig. 3 illustrates the average virtual service time T_S validation for different λ values and using a constant queue length ($K = 20$). p_b is constant for all curves. The figure also presents the service time curve for saturation, which was obtained by the model presented in [2]. Increasing the number of nodes, the network's load also increases until reaching the saturation. For $\lambda = 67$, the figure presents the service time crossing two extreme conditions: low network load, and almost saturated traffic. For higher λ values, the model presents distinct error regions as a function of the network's load. For a low number of nodes and in the saturation's proximity the error is negligible. But for a number of nodes between these two regions the model is less accurate. This comment can be verified in Fig. 3. Considering the $\lambda = 67$ curve, for a number of nodes between 2 and 8 and between 22 and 30 the model's error is small, while for a number of nodes between 8 and 22 the error significantly increases.

TABLE I
VALIDATION PARAMETERS USED

SIFS	10 µs	Channel bit rate	1 Mbit/s
DIFS	50 µs	MAC+PHY header	416 bits
EIFS	364 µs	ACK	304 bits
Slot Time (*σ*)	20 µs	Propagation delay (*δ*)	2 µs
BC stages (*m*)	7	ACK_TIMEOUT	304 µs
W_1	32	Simulated Time	500 s

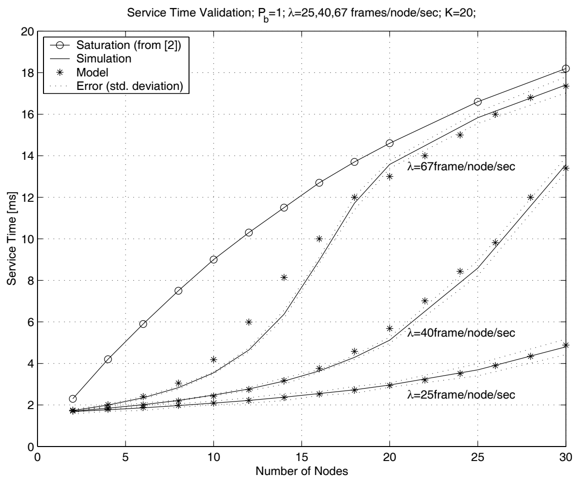

Figure 3. Average virtual service time T_s for different frame's rate generation.

The service time model was also validated for different probability values of p_b and constant λ, for low traffic. Fig. 4 shows the validation points. The error margin curves are only plotted for p_b values equal to 0.25 and 0.5. For the other p_b values all the validation points are contained in the simulation's error margin.

The validation data for the mean total delay T_d is presented in Fig. 5. For $p_b = 0$ frame drops exists for a number of nodes greater than 10, validating the M/M/1/K queue model both in saturation and non-saturation operating zones. Fig. 6 plots a surface of the numerical results of T_d for different loads of broadcast/unicast traffic and for different number of nodes. The results show a non-linear increase of the delay as more unicast frames are transmitted, for the same network load.

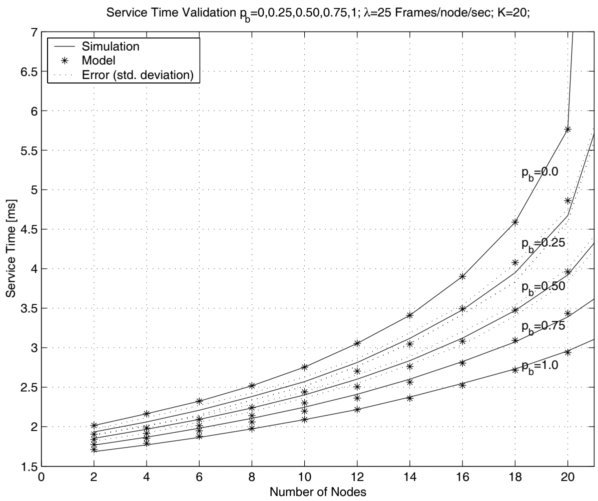

Figure 4. Average virtual service time T_s for different P_b values.

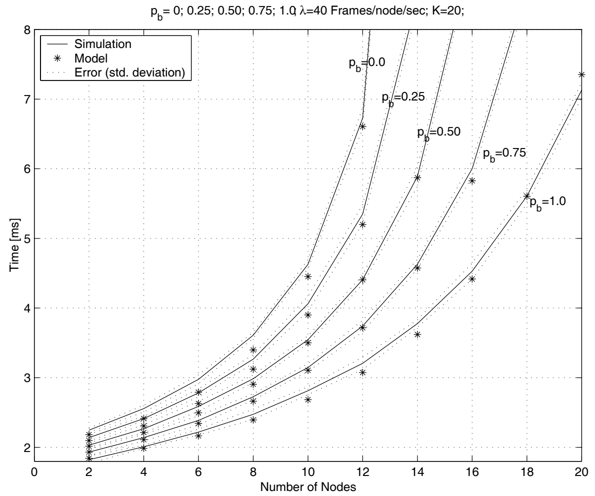

Figure 5. Total delay T_d considering different P_b values.

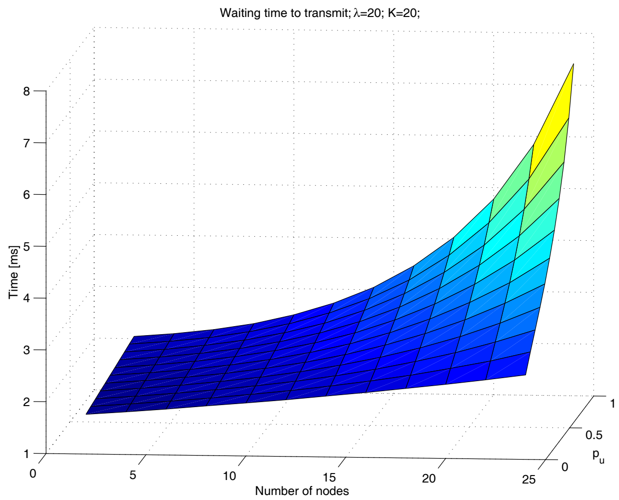

Figure 6. Model's numerical solutions for total delay.

VI. Conclusions

In this paper, we model the total frame's delay for IEEE 802.11 DCF networks in presence of both unicast and broadcast traffic. Observing Fig. 6, we conclude that the average frame's delay assume bigger values for unicast frames than for broadcast, as expected.

Almost all the validation points presented are contained in the simulation error margins, which confirm the model accuracy. Our initial hypothesis in which we approximate the service time as being exponential, leads to results that positively supports this approximation. The choice of a simple M/M/1/K queue model models the queue behaviour in presence of both broadcast and unicast traffic for low traffic.

Future work will aim to complement the present model with further properties evaluation and to use the model to perform some cross-layer optimizations at upper network layers.

Appendix A

Assuming k stations, and being $\mathrm{E}(P)$ the average time to put the entire frame in the channel (effective transmission time), the parameters in (12) are defined as:

$$T_{b_s} = \mathrm{E}(P) + DIFS + \delta$$

$$T_{u_s} = \mathrm{E}(P) + SIFS + DIFS + ACK + 2\delta$$

$$T_{b_c} = T_{u_c} = T_{m_c} = \mathrm{E}(P) + DIFS + EIFS + \delta$$

$$p_{b_s}(k) = (k-1)\chi_b(1-\chi)^{k-2}$$

$$p_{u_s}(k) = (k-1)\chi_u(1-\chi)^{k-2}$$

$$p_{b_c}(k) = (1-\chi_u)^{k-1} - (1-\chi)^{k-1} - p_{b_s}(k)$$

$$p_{u_c}(k) = (1-\chi_b)^{k-1} - (1-\chi)^{k-1} - p_{u_s}(k)$$

$$p_{m_c}(k) = 1 - (1-\chi)^{k-1} - p_{b_s}(k) - p_{u_s}(k) - p_{b_c}(k) - p_{u_c}(k)$$

References

[1] G. Bianchi, "Performance analysis of the IEEE 802.11 Distributed Coordination Function", in IEEE JSAC., vol. 18, no. 3, pp. 535–547, March 2000.

[2] R. Oliveira, L. Bernardo, P. Pinto, "Performance analysis of the IEEE 802.11 Distributed Coordination Function with unicast and broadcast traffic", in Proceedings of the IEEE PIMRC, Helsinki, Finland, September 2006.

[3] K- Medepalli, F. A. Tobagi, "System centric and user centric queueing models for IEEE 802.11 based Wireless LANs", Proceedings of IEEE Broadnets 2005, Boston, MA, October 2005.

[4] G. Cantieni, Q. Ni, C. Barakat, T. Turletti, "Performance Analysis under Finite Load and Improvements for Multirate 802.11", Computer Communications Journal, special issue on Performance Issues of Wireless LANs, PANs, and Ad Hoc Networks, vol. 28, no. 10, pp. 1095-1109, June 2005

[5] O. Tickoo, B. Sikdar, "A Queuing Model for Finite Load IEEE 802.11 Random Access MAC", IEEE ICC, Paris, France, June 2004

[6] ANSI/IEEE 802.11 standard, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications", 1999.

[7] IEEE 802.11b, "Supplement to 802.11-1999,Wireless LAN MAC and PHY specifications: Higher speed Physical Layer (PHY) extension in the 2.4 GHz band", 1999.

[8] Network simulator (version 2.28). Retrieved from http://www.isi.edu/nsnam/ns/

[9] G. Bolch, S. Greiner, H. de Meer, K. S. Trivedi, Queueing Networks and Markov Chains - Modeling and Performance Evaluation with Computer Science Applications, 2nd edition, John Wiley & Sons, 2006.