
Department of Applied Mathematics
Faculty of EEMCS



University of Twente
The Netherlands

P.O. Box 217
7500 AE Enschede
The Netherlands

Phone: +31-53-4893400

Fax: +31-53-4893114

Email: memo@math.utwente.nl
www.math.utwente.nl/publications

Memorandum No. 1767

**An analytical packet/flow-level modelling
approach for wireless LANs
with Quality-of-Service support**

S.K. CHEUNG, J.L. VAN DEN BERG¹

R.J. BOUCHERIE, R. LITJENS¹ AND F. ROIJERS¹

June, 2005

ISSN 0169-2690

¹TNO Information and Communication Technology, PO Box 5050, 2600 GB Delft, The Netherlands

An analytical packet/flow-level modelling approach for wireless LANs with Quality-of-Service support

S.K. Cheung^{1,*}, J.L. van den Berg^{2,1}, Richard J. Boucherie¹, R. Litjens², and F. Roijers²

¹ University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
Email: {S.K.Cheung, R.J.Boucherie}@utwente.nl

² TNO ICT, P.O. Box 5050, 2600 GB Delft, The Netherlands
Email: {J.L.vandenBerg, R.Litjens, F.Roijers}@telecom.tno.nl

Abstract. We present an analytical packet/flow-level modelling approach for the performance analysis of IEEE 802.11e WLAN, where we explicitly take into account QoS differentiation mechanisms based on minimum contention window size values and Arbitration InterFrame Space (AIFS) values, as included in the Enhanced Distributed Channel Access (EDCA) protocol of the 802.11e standard. We first enhance the packet-level approach previously used for best-effort WLANs to include traffic classes with different QoS requirements. The packet-level model approach yields service weights that discriminate among traffic classes. From these observations, the packet/flow-level model for 802.11e is the *generalized* discriminatory processor-sharing (GDPS) queueing model where the state-dependent system capacity is distributed among active traffic classes according to state-dependent priority weights. Extensive simulations show that the discriminatory processor-sharing model closely represents the flow behavior of 802.11e.

Keywords: IEEE 802.11e Wireless LAN, EDCA, Quality-of-Service, flow-level performance, file transfer times, generalized discriminatory processor-sharing (GDPS).

AMS Subject Classifications: primary: 90B18, 90B22; secondary 60K25.

1 INTRODUCTION

Wireless Local Area Networks (WLANs) fulfill the need for an additional public wireless access solution in hot spots (e.g. train stations, airports, etc.), besides the access provided by mobile cellular networks such as GSM/GPRS and UMTS. WLANs provide an interesting possibility to offer additional low-cost capacity and higher bandwidths to end-users without sacrificing the inherently scarce and expensive capacity of cellular networks. However, critical factors for successful introduction are security and performance, which applies in particular to the deployment of WLANs in the public environment.

WLAN performance is largely determined by the maximum data rate at the physical layer and the MAC layer protocols defined by the IEEE 802.11 standards [7]. An extension

* This work has been partially funded by the Dutch Ministry of Economic Affairs under the research program 'Technologische Samenwerking ICT-doorbraakprojecten', project TSIT1025 BEYOND 3G.

of the most widely employed DCF protocol (Distributed Coordination Function) is the EDCA protocol (Enhanced Distributed Channel Access). Both the DCF and the EDCA protocols are random access schemes based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). EDCA is aimed to provide QoS differentiation between various traffic classes, whereas DCF only supports best-effort services.

For best-effort WLANs (802.11b), several accurate analytical performance models have been developed (along the lines of Bianchi [2]) in order to study the system's saturated throughput as a function of the number of persistently active users. Foh and Zuckerman [6] and Litjens et. al. [9] considered 802.11b with the practical situation of dynamically varying number of active users due to the random initiation and completion of flow transfers. They obtain accurate approximations for the mean flow transfer time.

Analytical flow-level performance studies for WLANs with QoS support and with dynamically varying number of active users are not available. Performance studies of 802.11e WLANs are mainly based on simulation (e.g. [8, 10, 11]). Relatively few papers present an analytical approach, generally considering a fixed number of persistent users; see e.g. Zhao et. al. [12]. The simulation studies usually consider general scenarios (sometimes also capturing the impact of higher layer protocols like TCP), but without random user behavior.

In the present paper, an analytical performance evaluation of 802.11e WLANs is given along the lines of the analysis in [9]. From the flow-level point-of-view, the 802.11e WLAN is considered as a queueing system with Poisson arrivals and *generalized* discriminatory processor-sharing (GDPS) service discipline with state-dependent service capacity and state-dependent weights. This queueing model reflects the EDCA MAC design principle of distributing the available state-dependent transmission capacity among active traffic classes according to certain priority weights. In our modelling approach, the class weights and the system capacity depend on the number of active users and are obtained from a packet-level model that describes the MAC behavior of EDCA in detail in the situation with a fixed number of persistent users.

In contrast to *egalitarian* processor-sharing queues (as applied to best-effort WLANs in [9]), *discriminatory* processor-sharing (DPS) queues are significantly more difficult to analyze (see e.g. [1, 4]). Therefore, in this paper we use a simple analytical and efficient decomposition method for approximating the mean file transfer times in *generalized* discriminatory PS models, as introduced in Cheung et. al. [4]. The accuracy of our analytical approximations is validated by simulation of WLAN systems.

This paper is organized as follows. In Section 2, the EDCA MAC protocol is described in more detail. Section 3 describes our analytical modelling approach, which is presented in more detail in Section 4. The accuracy of our analytical model is validated in Section 5 by extensive simulation results. Finally, the principal conclusions of our investigation as well as some topics for further research are outlined in Section 6.

2 ENHANCED DISTRIBUTED CHANNEL ACCESS

The EDCA protocol defines several traffic classes, indexed by $i = 1, \dots, C$. When a user from traffic class i wants to transmit a data packet in the BASIC access mode, it first senses the medium to determine whether or not the channel is already in use by another user. If the channel is idle, and remains idle for a contiguous time period called $AIFS_i$

(AIFS value for class i) the user has to wait a random number of time slots before it is permitted to send the packet. This random back-off procedure is intended to reduce the probability of multiple users sending at the same time resulting in a collision.

The discrete back-off counter is uniformly sampled from $\{0, \dots, cw_{r,i} - 1\}$, where $cw_{r,i}$ is the contention window size for a traffic class i user at the r -th re-attempt to send the packet. As long as the channel remains idle after an $AIFS_i$ period, a class i user decrements its back-off counter by 1 for each slot time. When the back-off counter of a particular user reaches zero, the user transmits the packet. If the packet is received correctly, the destination responds by sending an acknowledgment (ACK) to the source. In case of multiple packets are transmitted concurrently, a collision occurs. If a user does not receive an ACK, it assumes that the packet was lost and it will retransmit the packet. At the next re-attempt to send the packet, a new back-off counter is sampled from a contention window with a doubled size. After a successful packet transmission, the contention window size is reset to its minimum value $cw_{0,i} = cw_{\min,i} + 1$.

In case of best-effort WLANs, all users use the same values for e.g. $cw_{\min,i}$ and $AIFS_i$. In the latter case, all $AIFS_i$ values are equal to the DIFS (Distributed InterFrame Space) value. A distributed access approach and QoS differentiation can be achieved with EDCA by using different parameter values for different traffic classes. Additional tunable parameter values are for e.g. $TXOP_{\text{limit}}$ (transmission opportunity limit), packet size, and the maximum contention window size. With a $TXOP_{\text{limit}}$ a user may send multiple packets as long as the last packet is completely transmitted before the $TXOP_{\text{limit}}$ time duration has expired. In general, QoS differentiation between various traffic classes is mainly achieved by contention window size and AIFS-based differentiation mechanisms.

3 MODELLING APPROACH

We consider a single basic service set with users from C traffic classes contending for shared 802.11e WLAN radio access, and each traffic class has its own set of tunable parameter values (e.g. $cw_{\min,i}$ and $AIFS_i$). We follow an integrated packet/flow-level modelling approach similar to that used in [9]. The first stage of the modelling approach is an enhanced packet-level model that describes the MAC behavior of QoS mechanisms (EDCA in 802.11e) in detail. The second stage is a flow-level model that describes the behavior of 802.11e, when the number of active users varies dynamically in time.

For the first stage, when $\mathbf{n} = (n_1, \dots, n_C)$ is the number of persistent users in the system (with n_j fixed users of class j), the resulting output of the packet-level model is the expected aggregate throughput $R_i(\mathbf{n})$ for class i , and $R(\mathbf{n}) := \sum_{i=1}^C R_i(\mathbf{n})$ is the expected aggregate system throughput. From a single user's perspective, a class i user receives an expected throughput of $R_{\text{flow},i}(\mathbf{n}) = R_i(\mathbf{n})/n_i$, and for $n_i > 0, n_j > 0$, the relative received throughput between a single class i and class j user is defined as

$$w_{ij}(\mathbf{n}) := \frac{R_{\text{flow},i}(\mathbf{n})}{R_{\text{flow},j}(\mathbf{n})} = \frac{R_i(\mathbf{n})/n_i}{R_j(\mathbf{n})/n_j}.$$

QoS differentiation in 802.11e WLANs is achieved by establishing that $w_{ij}(\mathbf{n})$ generally differs from 1. The *egalitarian* processor-sharing queueing model, as accurately used in [9] for 802.11b file transfer time analysis, is not suitable for the extended 802.11e version

with QoS support. However, the essential principle of distributing the state-dependent bandwidth in a processor-sharing fashion still remains for 802.11e WLANs. In the latter case, the capacity is shared in a *discriminating* fashion between the classes, as intended by the design of EDCA, and shared in an *egalitarian* fashion for users within the same class. On top of the discriminating feature of 802.11e WLANs is that the priority effect $w_{ij}(\mathbf{n})$ is also dependent on the number of active users, as observed from the packet-level analysis (and obviously $w_{ij}(\mathbf{n})$ also depend on the type of QoS differentiation mechanisms).

Based on these observations, we propose a *generalized* discriminatory processor-sharing (GDPS) model with both state-dependent service capacity and state-dependent weights. An attractive feature of our modelling approach is that the general form of the flow-level model is independent of the packet-level model, in the sense that only the expected saturated class throughputs $R_i(\mathbf{n})$ from the packet-level model are required as input for the GDPS flow-level model. In particular, it is independent of the type of QoS mechanisms.

4 PERFORMANCE ANALYSIS

In this section, we first give our straightforward packet-level extension from [9], when only differentiation in the contention window sizes can be applied. Differentiation with TXOP_{limit} and packet sizes are easily incorporated. AIFS-based extensions have been well studied in the literature. E.g., Zhao et. al. [12] proposed an extended Markov chain analysis which has been accurately validated. We omit the details due to space constraints. In principle, any accurate analytical packet-level model can be used as input of our non-persistent flow-level model.

In the second part of this section we briefly indicate some qualitative insights of the QoS mechanisms at the packet-level on the relative throughput measures $w_{ij}(\mathbf{n})$. This subsection is included to place our *generalized* discriminatory processor-sharing model with state-dependent weights $w_{ij}(\mathbf{n})$ in the right setting. Finally, in the last part of this section, we present our flow-level modelling approach based on the GDPS model in significant more detail.

4.1 Packet-level: Throughput analysis for persistent users

The throughput analysis for 802.11b at the packet-level and with a fixed number of persistent users (as used in [9] and originally developed by Bianchi [2]) essentially remains the same for 802.11e if no AIFS-based differentiation is considered, since the Markov chain for the back-off counting process is described only from an isolated user point-of-view.

For 802.11e, we assume that all traffic classes i , $i = 1, \dots, C$, have their own so-called packet error probability $\mathbf{P}_{e,i}$, independently of the other classes and independently of the number of collisions already involved. This is the same key assumption made in [2]. The influence of all other active users is captured by this packet error probability. Hence, the equilibrium distribution of the embedded jump chain for a class i user is similar to [9] and given by

$$\pi_i(r, b) = \frac{cw_{r,i} - b}{cw_{r,i}} \cdot \frac{2(1 - \mathbf{P}_{e,i}) \mathbf{P}_{e,i}^r}{\left(1 - \mathbf{P}_{e,i}^{r_{\max,i}+1}\right) + (1 - \mathbf{P}_{e,i}) \sum_{k=0}^{r_{\max,i}} cw_{k,i} \mathbf{P}_{e,i}^k},$$

where (r, b) denotes the back-off state ($0 \leq r \leq r_{\max,i}$ and $0 \leq b \leq cw_{r,i} - 1$), $r_{\max,i}$ denotes the maximum number of retries, and $cw_{r,i}$ is the contention window size of a class i user at the r -th re-attempt. Note that only the QoS parameters $cw_{\min,i}$ and $cw_{\max,i}$ appear in the equilibrium distribution of the Markov chain ($cw_{\min,i} := cw_{0,i} - 1$ and $cw_{\max,i} := cw_{r,i} - 1$ with $r = r_{\max,i}$).

The packet error probability $\mathbf{P}_{e,i}$ is readily expressed by

$$\mathbf{P}_{e,i} = 1 - (1 - \mathbf{P}_{tr,i}^*)^{n_i - 1} \prod_{k=1, k \neq i}^C (1 - \mathbf{P}_{tr,k}^*)^{n_k}, \text{ for all } i = 1, \dots, C,$$

where $\mathbf{P}_{tr,i}^*$ is the packet transfer probability (successful or not) for a class i user, i.e., $\mathbf{P}_{tr,i}^* = \sum_{r=0}^{r_{\max,i}} \pi_i(r, 0)$. It can be shown that a unique vector $(\mathbf{P}_{tr,i}^*, \mathbf{P}_{e,i})_{i=1}^C$ exists. The expected aggregate throughput $R_i(\mathbf{n}) \equiv R_i(n_1, \dots, n_C)$ for class i is given by (cf. [2, 9])

$$R_i(\mathbf{n}) = \frac{\mathbf{P}_{suc,i} \cdot \mathbb{E}P_i}{\mathbf{P}_{idle} \cdot \tau + \sum_{i=1}^C \mathbf{P}_{suc,i} \cdot \mathbf{T}_{suc,i} + \mathbf{P}_{col} \cdot \mathbf{T}_{col}},$$

where $\mathbf{P}_{idle} = \prod_{j=1}^C (1 - \mathbf{P}_{tr,j}^*)^{n_j}$ is the probability that the channel is idle at a randomly selected slot time, $\mathbf{P}_{suc,i} = n_i \mathbf{P}_{tr,i}^* (1 - \mathbf{P}_{tr,i}^*)^{n_i - 1} \prod_{j \neq i}^C (1 - \mathbf{P}_{tr,j}^*)^{n_j}$ is the probability that exactly one class i user transmits a packet at a random slot time, and the collision probability is given by $\mathbf{P}_{col} = 1 - \mathbf{P}_{idle} - \sum_{i=1}^C \mathbf{P}_{suc,i}$. Furthermore, τ is the slot time and the inter-event times $\mathbf{T}_{suc,i}$ (successful) and \mathbf{T}_{col} (collision) are given by

$$\begin{aligned} \mathbf{T}_{suc,i}^{BASIC} &= PHY + MAC + r_{WLAN}^{-1} \mathbb{E}P_i + \delta + SIFS + ACK + \delta + DIFS, \\ \mathbf{T}_{col}^{BASIC} &= PHY + MAC + r_{WLAN}^{-1} \mathbb{E}\bar{P} + \delta + DIFS. \end{aligned}$$

under BASIC access mode and where r_{WLAN} is the channel rate, PHY is the physical header (plus preamble), SIFS is the Short InterFrame Space time, δ is the propagation delay between sender and receiver (in seconds), $\mathbb{E}P_i$ is the expected net payload size for traffic class i (in kbits) and $\mathbb{E}\bar{P}$ is the expected net payload size of the largest packet involved in a collision. The MAC header and ACK size are converted to seconds. Differentiation with $TXOP_{\text{limit}}$ and packet size is easily incorporated in the inter-event times $\mathbf{T}_{suc,i}$ (\mathbf{T}_{col}) through the expected payload sizes.

The packet-level analysis for the RTS/CTS (ReadyToSend/ClearToSend) access mode is essentially the same; only the inter-event times $\mathbf{T}_{suc,i}^{RTS/CTS}$ and $\mathbf{T}_{col}^{RTS/CTS}$ are slightly different computed; see e.g. [2, 9]. An advantage of the RTS/CTS mode is that the time wasted by a collision is smaller as the RTS frame is significantly smaller than a data packet. A drawback is that more overhead is involved than BASIC access.

4.2 Packet-level: Qualitative analysis of QoS differentiation mechanisms

The packet-level model yields throughputs $R_i(\mathbf{n})$, that discriminate among active traffic classes and depend on the number of active users \mathbf{n} . Clearly, when cw_{\min} differentiation is applied with $cw_{\min,j} < cw_{\min,i}$ and the other QoS parameter values are kept equal for both classes, then it must hold that $w_{ij}(\mathbf{n}) < 1$. The same differentiating effect $w_{ij}(\mathbf{n}) < 1$ is also achieved when only $AIFS_j < AIFS_i$ is applied. However, the impact of the QoS differentiation parameters is generally different, and moreover, also heavily dependent on

the number of active users \mathbf{n} in the system. We illustrate this with two examples, as observed from the packet-level analysis.

Example 1. If $AIFS_j < AIFS_i$ is applied, then the high priority class j users always start and resume their back-off counting procedure sooner than the low priority class i users. When many high priority users are active in the system, a so-called *starvation effect* can occur, i.e., $w_{ij}(\mathbf{n}) \approx 0$. To this end, observe that after the end of the $AIFS_j$ time duration, that the back-off counters for all class i users always remain unchanged for at least a time duration of $AIFS_i - AIFS_j$, whereas every class j user resume to decrement their back-off counters after the end of the $AIFS_j$ time. In fact, a class j user can gain access to the medium before the end of the $AIFS_i$ time and hence leaving all back-off counters for class i users unchanged, while all class j users have decremented their back-off counters. Any successful packet transfer or any packet collision is beneficial for the high priority class j (in terms of the frequency of access to the medium for type j users). In the same scenario of $AIFS_j < AIFS_i$, but with few active users in the system, then this ratio $w_{ij}(\mathbf{n})$ is usually much larger than zero but obviously still less than 1.

Example 2. The relative weight $w_{ij}(\mathbf{n})$ can be even both less and greater than 1, dependent on \mathbf{n} . For example if $cw_{\min,i} < cw_{\min,j}$ in combination with $AIFS_i > AIFS_j$ is applied, then we have that cw_{\min} dominates the differentiation effect when the number of active users is small ($w_{ij}(\mathbf{n}) > 1$), and the smaller $AIFS_j$ value dominates the differentiation effect when the number of active users is large ($w_{ij}(\mathbf{n}) < 1$).

4.3 Flow-level: Transfer time analysis for non-persistent users

From the flow-level point-of-view, the number of active class i users in the system is not fixed, but varies dynamically in time due to initiation of file transfers and file transfer completions. We shall let N_i denote the (steady state) random variable of the number of active class i users in the 802.11e WLAN network. Since the packet-level model yields state-dependent capacity $R(\mathbf{n})$ and state-dependent weights $w_{ij}(\mathbf{n})$, the 802.11e WLAN network can be considered as a service center with a *generalized* DPS service discipline. When no service differentiation is applied under 802.11e, then the packet-level model will result in $w_{ij}(\mathbf{n}) = 1$ (for all i, j and for all \mathbf{n} such that $n_i > 0, n_j > 0$) and $R(\mathbf{n})$ only depends on the total number of active users $n_1 + \dots + n_C$. In the latter case, the *generalized* DPS model is equivalent to Cohen's GPS model [5], which is applied to the flow-level performance analysis for best-effort WLANs [9].

We assume that the traffic classes generate data flows according to independent Poisson processes with rate λ_i , $i = 1, \dots, C$, and a class i user request the download transfer of a data file whose size is generally distributed with mean $\mathbb{E}X_i$ (in kbits). Each file is segmented into packets of a given size (with a final packet containing the file's remainder) which are processed at the WLAN's MAC layer. The offered data load of class i is denoted by $\rho_i \equiv \lambda_i \mathbb{E}X_i / r_{WLAN}$, and the total offered load by $\rho := \sum_{i=1}^C \rho_i$. To ensure stability and provide *statistical guaranteed* QoS, we limit the number of contending data flows by $n_{\max,i}$ for each class separately.

Analytical expressions for (G)DPS models are generally not available in tractable form. Therefore, we use an analytical approximation/decomposition technique for *general*

processor-sharing queues, as proposed by Cheung et. al. [4]. In the remainder of this paper, we keep the demonstration of the solution technique simple, by considering the example of two traffic classes, i.e., $C = 2$. The approximation can be generalized for arbitrary number of traffic classes, see [4].

We apply the decomposition technique as follows. For specific traffic loads ρ_i and given the class capacities $R_1(n_1, n_2)$ and $R_2(n_1, n_2)$, we first approximate the equilibrium distribution of N_1 and N_2 . If class 2 users are persistent (permanent) in the system, then the distribution of N_1 (conditional on n_2 fixed class 2 users) is easily computed in closed-form formulas according to Cohen's GPS model [5], i.e., the conditional probabilities defined by $\alpha(n_1, n_2) = \mathbb{P}(N_1 = n_1 \mid n_2 \text{ persistent class 2 users})$ are given by

$$\alpha(0, n_2) = \left(1 + \sum_{n_1=1}^{n_{\max,1}} \left(\rho_1^{n_1} \prod_{k=1}^{n_1} \frac{r_{WLAN}}{R_1(k, n_2)} \right) \right)^{-1} =: G_1(n_2)^{-1}, \quad (1)$$

$$\alpha(n_1, n_2) = \left(\rho_1^{n_1} \prod_{k=1}^{n_1} \frac{r_{WLAN}}{R_1(k, n_2)} \right) / G_1(n_2), \text{ for } n_1 = 1, \dots, n_{\max,1}, \quad (2)$$

with $\sum_{n_1=0}^{n_{\max,1}} \alpha(n_1, n_2) = 1$ for all $n_2 = 0, 1, \dots, n_{\max,2}$. Analogously, the conditional probabilities defined by $\beta(n_2, n_1) = \mathbb{P}(N_2 = n_2 \mid n_1 \text{ persistent class 1 users})$ are readily computed as

$$\beta(0, n_1) = \left(1 + \sum_{n_2=1}^{n_{\max,2}} \left(\rho_2^{n_2} \prod_{k=1}^{n_2} \frac{r_{WLAN}}{R_2(n_1, k)} \right) \right)^{-1} =: G_2(n_1)^{-1}, \quad (3)$$

$$\beta(n_2, n_1) = \left(\rho_2^{n_2} \prod_{k=1}^{n_2} \frac{r_{WLAN}}{R_2(n_1, k)} \right) / G_2(n_1), \text{ for } n_2 = 1, \dots, n_{\max,2}, \quad (4)$$

with $\sum_{n_2=0}^{n_{\max,2}} \beta(n_2, n_1) = 1$ for all $n_1 = 0, 1, \dots, n_{\max,1}$.

The approximation of the unconditional and marginal distribution $\mathbb{P}(N_1 = n_1)$ is obtained by factorizing the conditional distribution $\alpha(n_1, n_2)$ over $\mathbb{P}(N_2 = n_2)$, and similarly $\mathbb{P}(N_2 = n_2)$ is obtained by factorizing $\beta(n_2, n_1)$ over $\mathbb{P}(N_1 = n_1)$, i.e., solve the linear system (cf. [4])

$$\mathbb{P}(N_1 = i) = \sum_{k=0}^{n_{\max,2}} \alpha(i, k) \mathbb{P}(N_2 = k), \quad \text{for } i = 0, 1, \dots, n_{\max,1}, \quad (5)$$

$$\mathbb{P}(N_2 = j) = \sum_{k=0}^{n_{\max,1}} \beta(j, k) \mathbb{P}(N_1 = k), \quad \text{for } j = 0, 1, \dots, n_{\max,2}. \quad (6)$$

This system can be solved efficiently and it can be shown that the solution is unique up to a multiplicative constant. A sufficient condition is that $R_i(\mathbf{n}) > 0$ whenever $n_i > 0$. We stress that Eq. (5)-(6) is an approximation, since $\alpha(n_1, n_2) \neq \mathbb{P}(N_1 = n_1 \mid N_2 = n_2)$ and $\beta(n_2, n_1) \neq \mathbb{P}(N_2 = n_2 \mid N_1 = n_1)$, i.e., the coefficients are not equal to the conditional steady state probabilities of the GDPS model when users from both classes are non-persistent. The latter distributions for (G)DPS models have not yet been obtained and simple expressions seem not to exist in contrast to egalitarian PS models.

From the approximation for $\mathbb{P}(N_1 = n_1)$ and $\mathbb{P}(N_2 = n_2)$, the expectations $\mathbb{E}N_1$ and $\mathbb{E}N_2$ are easily computed, and by applying the well-known Little's formula, we approximate the mean file transfer time $\mathbb{E}T_i$ for class i with $\widehat{\mathbb{E}T}_i$, given by the simple formula

$$\widehat{\mathbb{E}T}_i = \frac{\mathbb{E}N_i}{\lambda_i(1 - \mathbb{P}(N_i = n_{\max,i}))}, \quad i = 1, 2. \quad (7)$$

It is proven that the approximation for the queue length distributions, as obtained from Eq. (1)-(6), yields an exact result for *egalitarian* PS models with state-dependent system capacity that only depends on the total number of users in the system (see [4]). In addition, it is numerically demonstrated and discussed that the approximation for *discriminatory* PS models is accurate for the high priority class for a wide region of traffic loads. For the low priority class the approximation breaks down if the relative weight ratio is extremely large (or small) in combination with heavy traffic. For moderate traffic loads the approximation is accurate. For more than $C = 2$ traffic classes, the similar approximation method also works well; again except for the cases with extremely asymmetric weights and heavy traffic (see [4]).

5 NUMERICAL RESULTS

In this section we present numerical results obtained from our analysis and compare them with WLAN simulation results. A detailed representation of the EDCA MAC layer has been implemented in a C/C++ program. Sufficient independent replications were run to obtain 95% confidence intervals with a relative precision no worse than 5%. The default parameter settings (without QoS differentiation) for the MAC layer and the physical layer are given in Table 1.

We present results for cw_{\min} and AIFS-based differentiation only, and for BASIC access mode. We first present the throughput results for persistent users (Stage 1). The considered QoS scenarios are labelled by $cw_{\min} = (cw_{\min,1}, cw_{\min,2})$ and $AIFS = (AIFS_1, AIFS_2)$. The simulation and analytical results for the aggregate system throughput $R(n_1, n_2)$ and the class throughputs $R_i(n_1, n_2)$, $i = 1, 2$, are shown only for $n_1 = n_2$ with total number $n_1 + n_2$, since it is difficult to visualize the difference between 2-dimensional functions (if mapped on \mathbb{R}^3). Graphs for $n_1 \neq n_2$ are similar. In the second part of this section, we present the mean file transfer time results (Stage 2) for the corresponding non-persistent flow-level model.

Table 1. Default parameter settings for the MAC layer and the physical layer

parameter	value	parameter	value	parameter (all i)	value
PHY	192 bits	r_{WLAN}	$1 \cdot 10^3$ kbits/s	$n_{\max,i}$	25
MAC	272 bits	δ	$1 \mu s$	$cw_{\min,i}$	31
RTS	PHY + 160 bits	τ	$20 \mu s$	$r_{\max,i}$	3 (BASIC)
CTS	PHY + 112 bits	SIFS	$10 \mu s$	AIFS $_i$	2 (time slots)
ACK	PHY + 112 bits	DIFS $_{\text{default}}$	2 (time slots)		
packet size	12 kbits	DIFS $_{\text{default}}$	$50 \mu s$ (time)*		

* DIFS $_{\text{default}} = \text{SIFS} + 2\tau = 50 \mu s$ (converted from number of time slots to seconds $\times 10^{-6}$).

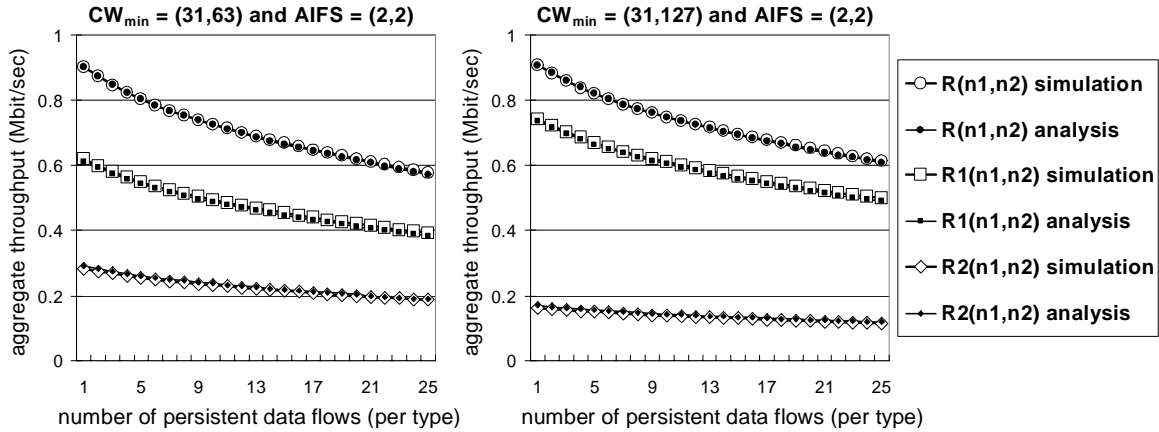


Fig. 1. Aggregate system and class throughputs as function of the number of persistent users $n_1 = n_2$, $n_i = 1, \dots, 25$; for $cw_{\min} = (31, 63)$ and $cw_{\min} = (31, 127)$ (no AIFS differentiation, i.e., $AIFS = (2, 2)$).

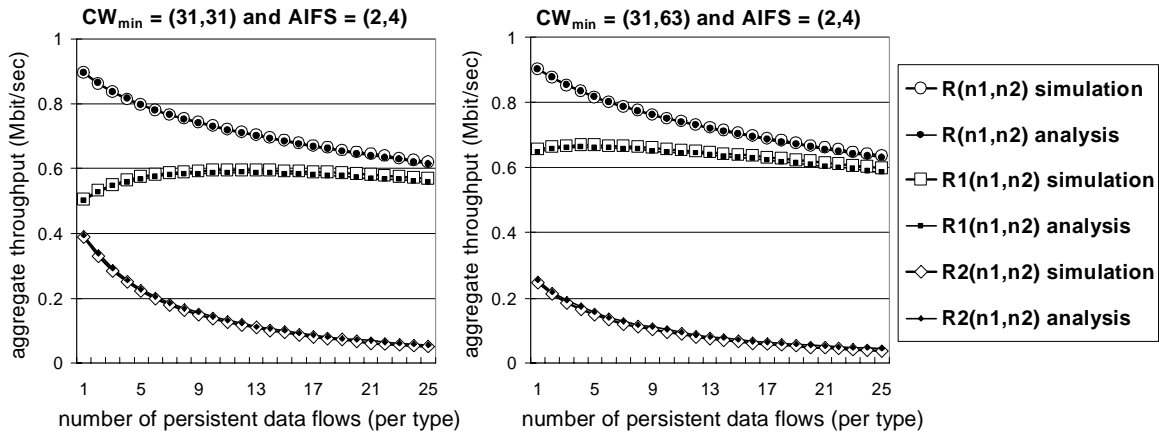


Fig. 2. Aggregate system and class throughputs as function of the number of persistent users $n_1 = n_2$, $n_i = 1, \dots, 25$; for $cw_{\min} = (31, 31)$, $AIFS = (2, 4)$ and $cw_{\min} = (31, 63)$, $AIFS = (2, 4)$.

5.1 Packet-level: Throughput results for persistent users (Stage 1)

Figure 1 presents the throughput results where only cw_{\min} -based QoS differentiation is applied; respectively for $cw_{\min} = (31, 63)$ and $cw_{\min} = (31, 127)$. Figure 2 presents results for the scenario $AIFS = (2, 4)$, and the scenario $cw_{\min} = (31, 63)$ with $AIFS = (2, 4)$. The analytical results for AIFS-based differentiation are based on the Markov chain model [12]. The numerical results accurately represent 802.11e packet-level behavior.

The cw_{\min} and AIFS parameters differ in several ways. The cw_{\min} is used to reduce collision probabilities as well as providing QoS support. The cw_{\min} -based differentiation mechanism approximately maintains the bandwidth ratio $cw_{\min,2}/cw_{\min,1}$ for large number of users. The QoS capabilities of the AIFS-based differentiation mechanism are particularly effective for busy systems. The class with largest AIFS value will eventually suffer from starvation as the system gets busier. However, in a system with realistic, non-persistent traffic load, these large numbers of simultaneously active users are mostly not achieved.

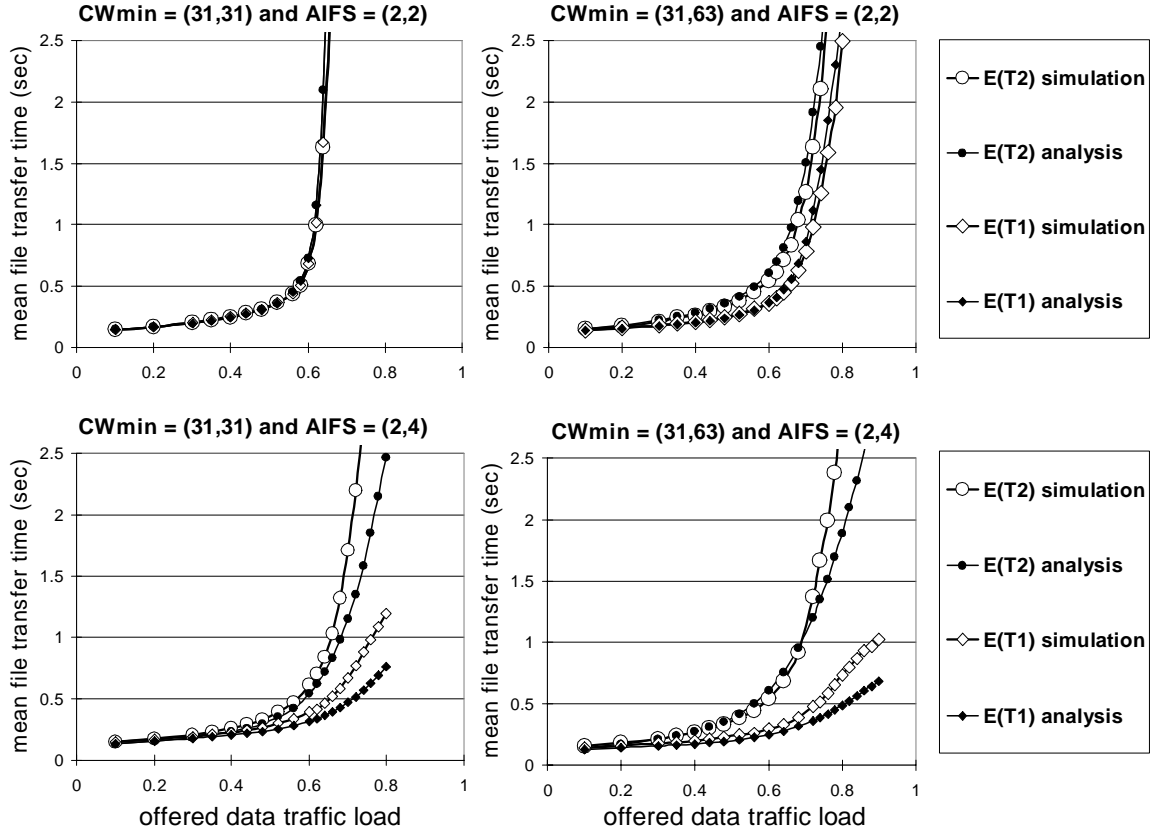


Fig. 3. Mean file transfer time $\mathbb{E}T_1$ and $\mathbb{E}T_2$ as a function of the offered data traffic load $\rho = \rho_1 + \rho_2$, with $2\rho_1 = \rho_2$, for the default scenario (no QoS differentiation) and 3 other QoS scenarios from Stage 1.

5.2 Flow-level: Transfer time results for non-persistent users (Stage 2)

We consider the mean file transfer time $\mathbb{E}T_j$ for traffic class j as a function of the offered traffic load $\rho = \rho_1 + \rho_2$, with $\rho_2/\rho_1 = 2$ (the low priority class contributes twice as much to the total offered system load, compared to the high priority class). Any pair of loads (ρ_1, ρ_2) can be chosen; however, to avoid 3-dimensional graphs we only depict the transfer time results for the ratio of loads $\rho_2/\rho_1 = 2$. The approximation result will be at ‘worst case’ if $\rho_2/\rho_1 = 1$; and under a constant total load $\rho = \rho_1 + \rho_2$, it will improve for increasing ratio ρ_2/ρ_1 , and also improve for decreasing ρ_2/ρ_1 . To this end, observe that if $\rho_2/\rho_1 \rightarrow \infty$, then the corresponding scenario is simply a single class (egalitarian) PS model with only active users from class 2. Analogously, if $\rho_2/\rho_1 \rightarrow 0$, then the corresponding scenario is a single class (egalitarian) PS model with only active users from class 1.

We limit the number of users for each class in the system with $n_{\max,i} = 25$. Figure 3 shows results for BASIC access mode and with exponentially distributed file sizes with mean of 120 kbits (10 packets of 12 kbits) for both classes.

When no service differentiation is applied, then clearly $\mathbb{E}T_1 = \mathbb{E}T_2$, and the approximation is exact for egalitarian PS models. Also, the analytical approximation accurately represents the WLAN simulation result. For the other three scenarios with QoS differentiation, the approximation yields accurate results for a realistic region of parameter

settings. The absolute approximation error is small for the high priority class 1 and relatively small for the low priority class 2, when the offered traffic load is moderate ($\rho < 0.7$). The region where the approximation breaks down is when the system is highly overloaded ($\rho \gg 0.7$). In the latter case, N_i is often close to its maximum value $n_{\max,i}$ and the number of blocked users tends to increase to infinity if the traffic load is above a certain threshold.

The analytical approximation for particularly the AIFS-based differentiation and extreme heavy traffic is as expected not accurate since $w_{21}(\mathbf{n}) \approx 0$, if the number of active users is constantly large. However, from a practical flow-level point-of-view, an overloaded WLAN system under BASIC access mode is not a realistic scenario setting. It is more efficient to use the RTS/CTS mode instead of the BASIC mode when the system is heavily loaded. From simulation results for RTS/CTS (not shown) we observed that the accuracy of the approximation is mostly slightly better than the results for BASIC mode.

Finally, we note that the analytical approximation method is insensitive to the service time (file size) distributions, i.e., the approximated mean file transfer times do *not* depend on the choice of service time distributions. The queue length distributions and mean flow transfer times in egalitarian PS models, such as Cohen's GPS [5] model, have the attractive property of insensitivity to the service time distributions. These egalitarian PS models belong to the so-called class of product-form networks. The discriminatory PS models (including GDPS) are non-product-form and do *not* have the insensitivity property, see e.g. [3]. However, the sensitivity only becomes significant when the priority weights are extremely asymmetric in combination with heavy traffic.

6 CONCLUSION AND FURTHER RESEARCH

We have presented an integrated packet/flow-level modelling approach for performance evaluation of IEEE 802.11e WLANs with dynamically random user behavior. The packet-level model describes the QoS mechanisms of the EDCA MAC layer in detail, and the flow-level model is based on the observation that the considered 802.11e system behaves approximately as a queueing system with a *generalized* discriminatory processor-sharing (GDPS) service discipline.

We used an analytical based decomposition method [4] for approximating mean file transfer times, since exact evaluation of (G)DPS models is not tractable. The 802.11e WLAN simulations show that the flow-level behavior of 802.11e is closely represented by a GDPS model with state-dependent service capacity and state-dependent service weights. The approximation is accurate for a wide region of realistic parameter settings. The scenarios where the approximation breaks down (particularly for the low priority class) is when the *starvation effect* becomes significant, i.e., when QoS scenarios are considered such that the relative priority weights are extremely small or extremely large in combination with heavy traffic (particularly from the high priority class users).

The (G)DPS modelling approach offers additional insights in the flow behavior of 802.11e. When a low priority user generates a data flow at a time instant when few high priority users are active, then the low priority user will have relatively small file transfer times. However, if more high priority users start to generate data flows, then the low priority's performance decreases severely, as if the service process is suddenly 'frozen'.

But on the other hand, due to the large share of bandwidth that the high priority users receive, the high priority users reside in the system for a relatively short period of time. When few high priority users are active, the low priority users still get a substantial share of the available bandwidth.

We conclude that the low priority's flow-level performance is characterized by a high variance, whereas the high priority's performance is much less variable. This shows that providing time-bounded QoS guarantees seems impossible, particularly for the low priority class. When statistical performance guarantees are given for the low priority class, then these are at a generally small confidence level (unless the system is light loaded). Topics for further research include taking more enhancements in the physical layer into account (e.g. capture effects) and to optimize the EDCA performance under certain guaranteed QoS. Also, finding a tractable evaluation of the mean file transfer times conditional on the user's initial file size for IEEE 802.11e WLAN and (G)DPS is a challenging task.

References

1. K. Avrachenkov, U. Ayesta, P. Brown, and R. Núñez-Queija. Discriminatory Processor Sharing Revisited. *Proceedings of IEEE INFOCOM 2005*, Miami, USA.
2. G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications* 18 (2000), 3: 535-547.
3. T. Bonald, and A. Proutière. Insensitivity in processor-sharing networks. *Performance Evaluation* 49 (2002), 193-209.
4. S.K. Cheung, J.L. van den Berg, and R.J. Boucherie. Decomposing the queue length distribution of processor-sharing models into queue lengths of permanent customer queues. *Proceedings of IFIP's Performance 2005*, Oct. 3-7, Juan-les-Pins, France.
5. J.W. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica* 12 (1979), 245-284.
6. C.H. Foh, and M. Zukerman. Performance analysis of the IEEE 802.11 MAC protocol, *Proceedings of European Wireless 2002*, Florence, Italy.
7. IEEE, "Standard for Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless MAC and PHY Specifications: MAC enhancements for Quality of Service (QoS)", P802.11e 4.3, May 2003.
8. J. Kim, and C. Kim. Performance analysis and evaluation of IEEE 802.11E EDCA, *Wireless Communications and Mobile Computing* 2004, 4: 55-74.
9. R. Litjens, F. Roijers, J.L. van den Berg, R.J. Boucherie, and M. Fleuren. Analysis of flow transfer times in IEEE 802.11 wireless LANs. *Annals of telecommunications* 59, no. 11-12, nov.-dec. 2004, Traffic engineering and routing.
10. T. Raimondi, and M. Davis. Design rules for a Class-based Differentiated Service QoS Scheme in IEEE 802.11E Wireless LANs, *MSWiM 2004*, Venice, Italy.
11. H. Truong, and G. Vannuccini. The IEEE 802.11E MAC for Quality Of Service in Wireless LANs. *Proceedings of SSGRR 2003*, L'Aquila, Italy.
12. J. Zhao, Z. Guo, Q. Zhang, and W. Zhu. Performance study of MAC for service differentiation in IEEE 802.11. *Proceedings of IEEE GLOBECOM 2002*, Taipei, Taiwan.