Achieving Service Differentiation in 802.11e WLANs using p-Persistent CSMA Throughput/Delay Analysis

Richard MacKenzie† and Timothy O'Farrell‡

† School of Electronic & Electrical Engineering, University of Leeds
‡ Institute of Advanced Telecommunications, Swansea University

Abstract: There is a growing need for quality of service (QoS) provisioning over 802.11 wireless local area networks. The 802.11e amendment addresses this need by allowing for traffic to be given different priority assignments. This can allow for QoS to be provided by service differentiation. In this paper we provide a simple throughput/delay analysis for p-persistent CSMA which can be used as an analysis for 802.11e. We show how this simple analysis can allow for system optimisation and also show how effective service differentiation can be achieved.

1 Introduction.

Due to the widespread usage of 802.11 wireless local area networks (WLANs) [1] there is a growing necessity to provide quality of service (QoS). The 802.11e amendment to the standard specifies enhanced distributed channel access. This allows for each station to contain four parallel queues, where each queue can have different priority channel access. The underlying mechanism for 802.11 contention based access is carrier sense multiple access (CSMA) with collision avoidance (CSMA/CA). CSMA was first introduced in [2]. One of the CSMA schemes proposed was p-persistent CSMA. In [3] a model for p-persistent CSMA was introduced for a system with a finite number of stations. In [4] it is shown that the performance of 802.11 CSMA/CA can be modelled as p-persistent CSMA and that a fixed sized contention window with an appropriate value can allow for 802.11 performance to be optimised. Later [5] adds the concept of service differentiation, where different stations have different priorities depending on their relative importance. In [6] the analysis in [3] is modified to produce a p-persistent CSMA analysis as an alternative to that in [5]. In this paper we extend on the throughput analysis in [6] by adding a delay analysis and looking at the throughput/delay performance characteristics.

2 p-Persistent CSMA Model

This model extends on that in [3] by adding the functionality for multiple traffic types. Stations are considered to be of the same traffic type if they have the same priority. The system will contain \( M \) stations where \( M = M_1 + M_2 + ... + M_{d_{\text{max}}} \), where \( M_d \) represents the number of stations of traffic type \( d \). The traffic types are numbered from 1 up to \( d_{\text{max}} \). All stations are assumed to be in line of sight of one another and the channel is assumed to be perfect. This means that transmission failures will only occur due to collisions caused by two or more stations transmitting at the same time.

Every station in this model is saturated. Where the non-saturated analysis in [3] had three possible states this is now reduced to two. These two states are a contention period (\( R \)) followed by a transmission period (\( T \)). These periods occur alternately in a regenerative cycle. Stations can only begin a transmission during \( R \). \( R \) is slotted meaning transmissions can only commence on the boundary of a timeslot. At each timeslot boundary a station of traffic type \( d \) will begin a transmission with probability \( p_d \) or defer with probability \( (1-p_d) \). The start of a transmission signifies the end of the contention period (\( R \)) and the start of a transmission period (\( T \)). If only one transmission occurs to start \( T \), then that transmission period will be successful. However, if more than one transmission occurs to start \( T \), then those transmissions will collide with one another resulting in failed transmission attempts.

2.1 Throughput Analysis

The throughput analysis is summarised from that given in [6]. While \( d \) can represent any particular traffic type \( c \) is used to represent the traffic type whose performance we are calculating. The throughput for all traffic type \( c \) stations, \( S_c \), is calculated as shown in (1).
\[ S_c = \frac{E[U_c]}{E[R] + E[T]} \]  

(E[U_c]) represents the expected time spent on successful type \( c \) transmissions during each transmission period while \( E[R] \) and \( E[T] \) represent the expected time for each contention period and transmission period respectively. In this analysis, \( T \), and therefore \( E[T] \), is considered to be of unity length. This is the case for both successful and unsuccessful transmission periods. This unity length \( T \) means we are calculating normalised throughput. The timeslot length, \( a \), is therefore a fraction which reflects its size relative to \( T \). \( E[R] \) can therefore be calculated as shown in (2) by considering how long all stations are likely to defer before a transmission is likely to occur. \( k \) represents the timeslot boundaries during \( R \), which are numbered upwards from 0.

\[ E[R] = a \sum_{k=1}^{\infty} \prod_{d=1}^{d_{\text{max}}} (1 - p_d)^{kM_d} \]  

(2)

In order to calculate \( E[U_c] \) we need to know the probability of a successful transmission at each timeslot boundary from \( k=0 \). So for each timeslot boundary, \( K \), we calculate the probability that \( R \) is at least \( K \) timeslots in length and that only one type \( c \) station begins a transmission, \( P(1tx_c) \), and that no transmissions begin for stations of traffic types where \( d \neq c \), \( P(0tx_d) \). Therefore \( E[U_c] \) can be calculated as shown in (3)

\[ E[U_c] = b \sum_{k=0}^{\infty} \left( P(1tx_c) \prod_{d \neq c} P(0tx_d) \right) \]  

(3)

where

\[ P(1tx_c) = M_c p_c (1 - p_c)^k \left(1 - p_c\right)^{k+1} \]  

(4)

\[ P(0tx_d) = \left(1 - p_d\right)^{k+1} \]  

(5)

We introduce \( b \) into the equation to represent the fraction of a transmission period that is actually useful data, allowing for overheads such as packet headers to be removed from the calculation of \( S_c \).

### 2.2 Delay Analysis

For the delay analysis we show how to calculate the average service time for a packet of traffic type \( c \), \( E[D_{S_c}] \). That is the time from when a packet becomes head of line in a station’s queue so that it can start contending for transmission until the time that a successful transmission period is completed for that packet. First we must calculate the probability that a particular type \( c \) station is successful during each transmission period. This can be calculated as shown in (6).

\[ P_{S_c} = \frac{E[U_c]}{bM_c} \]  

(6)

We use \( b \) here as \( E[U_c]/b \) represents the number of successful type \( c \) transmissions during a transmission period. The number of cycles (\( E[R] + E[T] \)) that a type \( c \) packet contends for the channel before it is successfully transmitted forms a geometric distribution with mean value of \( 1/P_{S_c} \). Therefore the mean service time for a packet can be calculated as shown in (7).

\[ E[D_{S_c}] = \frac{1}{P_{S_c}} (E[R] + E[T]) \]  

(7)

### 3 Applying p-Persistent CSMA to 802.11e

Although we only have one traffic type at any station, we convert the p-persistent analysis to 802.11e and not 802.11 for three main reasons. Firstly the backoff mechanism is slightly different in 802.11e and matches the p-persistent analysis more accurately. Second, in 802.11e the contention windows are
allowed to be changed. Third, in future versions of this model we wish to consider multiple traffic types at each station as allowed by 802.11e. As all stations are saturated applying the p-persistent CSMA analysis to 802.11e is reasonably straight forward as saturated 802.11e stations are always in backoff. For each traffic type we need to assign each type \(d\) 802.11e station with a fixed sized contention window value, \(CW_d\), so that the behaviour during the contention period, \(R\), is similar to that in p-persistent CSMA with p-persistence value \(p_d\). This relationship, originally calculated in [4], is \(p_d = (2/CW_d + 2)\). As \(CW_d\) must be an integer value we use \(CW_d = \text{ceiling}((2/p_d)-2}\). In this work we are interested in the throughput of the media access control (MAC) layer Service data units (MSDU) so we set \(b\) in order to remove the MAC and physical layer overheads, which includes interframe spaces and the time for MAC layer packet acknowledgements, from the \(S_c\) calculations.

4 Test Procedure

In these tests we wish to show how to optimise an 802.11e system using this p-persistent CSMA analysis. In each test we have two different traffic types. We aim to achieve service differentiation by maintaining target throughput ratios (\(S\)ratios) which are defined as \(S_1:S_2\). We also want to identify the relationship between the throughput and delay performance. In all tests we use 1500 byte MSDUs. The 802.11 physical layer (PHY) used is 24Mb/s extended rate PHY - orthogonal frequency division multiplexing (ERP-OFDM). The timeslot length is 9µs. All 802.11e tests are performed using the ns2 simulator [7].

5 Results

The first set of results shows how system performance can be optimised by selection of the appropriate \(p\) values. The system has 6 stations per traffic type and a target throughput ratio of 2:1 is maintained. Fig. 1 shows the throughput performance against \(p_1\) while Fig. 2 shows the corresponding delay results. When \(p_1\) is very low, the throughput is low and the delay is high. This is due to the fact that each contention period, \(R\), is made extremely large as transmissions seldom occur. As \(p_1\) increases, the throughput increases while the delay decreases. However, as \(p_1\) increases beyond a certain point, the throughput begins to fall while the delay begins to increase. This deterioration in performance is due to the fact that fewer transmission periods, \(T\), are successful as more and more transmission periods contain collisions. Therefore there is an optimum point, which in this scenario is for \(p_1 \approx 0.018\), where \(R\) is reasonably short without causing the collision probability in \(T\) to be too high. We can also see that the 802.11e performance does closely match that of the p-persistent analysis.

Fig. 1  Throughput per traffic type with 2:1 target throughput ratio

Fig. 2  Service Time per traffic type with 2:1 target throughput ratio

In this next set of tests we look at the performance for systems with sizes varying from 1 up to 10 stations per traffic type. For each system size we try to achieve target throughput ratios (\(S\)ratios) of 1:1, 2:1 and 3:1. The achieved throughput ratios (\(S_1:S_2\)) are shown in Fig. 3 while the achieved delay ratios (\(D_1:D_2\)) are shown in Fig. 4. As you can see, the 802.11e performance closely matches the p-
persistent CSMA analysis in successfully achieving the target Sratios. The delay ratio achieved is the reverse of the target Sratio. From inspecting (7) which contains the fraction \((1/\lambda_{c})\) this is what we should expect.

Fig. 3 Throughput ratios achieved for networks of varying size

Fig. 4 Delay ratios achieved for networks of varying size

6 Conclusion

In this paper we have added to the authors’ existing p-persistent CSMA throughput analysis by adding an associated delay analysis. We have shown that the throughput and delay performance of 802.11e systems can be optimised using this simple p-persistent analysis and also shown the inverse relationship between the throughput and delay. Further work on this model is being carried out by the authors. This includes adding features to the p-persistent CSMA analysis to account for the other features in 802.11e such as interframe space differentiation, internal collision resolution and transmit opportunities. The authors are also developing a non-saturated model for this analysis.

Acknowledgments.

This work is jointly funded by British Telecommunications PLC (BT) and the Engineering and Physical Sciences Research Council (EPSRC) with BT/EPSRC agreement code CT1080038286.

References.


