A Two-level Cross-talked Admission Control Mechanism for QoS Guarantee in 802.11e EDCA

NIU Zhisheng (牛志升), LIU Jing (刘静)
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Abstract: In this paper, a two-level cross-talked admission control mechanism is proposed to guarantee QoS (Quality of Service) requirements for multimedia applications over WLANs (Wireless Local Area Network). Firstly, an EDCA (Enhanced Distributed Channel Access) analytic model is used to compute the maximum number of admitted users according to the QoS requirements and the packet arrival characters. Then, some channel resource is reserved for handoff calls according to the maximum number of admitted users and the call level traffic model. The channel utilization ratio is also measured to denote the current system traffic load. The maximum number of admitted users and channel utilization ratio are used for admission control for applications with QoS requirements in call level, while they are also used for rate control of best effort applications in packet level by using the p-NACK scheme. Thus, QoS requirements is statistically guaranteed while the system is efficiently utilized. Simulation results validate the efficiency of our mechanism in QoS guarantee and bandwidth utilization.

Key words: WLAN; EDCA; QoS; admission control; rate control

Received:2007-**.**
**To whom correspondence should be addressed. E-mail: niuzhs@tsinghua.edu.cn Tel: 86-10-62773634
Introduction

IEEE 802.11 based wireless LAN has been massively deployed in public and residential places, because of their low cost, simplicity of installation and high data rates. Meanwhile, multimedia applications are increasing tremendously. Multimedia applications such as voice and video are expected to be provided in WLAN. Unlike traditional data applications, those applications have QoS requirements on delay, packet loss rate and other parameters.

The IEEE 802.11 Task Group specified a distributed access mechanism called EDCA to support service differentiation in MAC layer by assigning different access parameters to users of different classes[1]. However, service differentiation can not guarantee the QoS requirements of each traffic class because of the shared and contention-based nature of the 802.11 channel which is originally designed for best-effort applications. With the increase of admitted stations, the collision probability also increases, then the QoS parameters such as packet loss rate and access delay of every stations does not satisfied the QoS requirements. Thus, Admission Control is necessary to support applications with QoS requirements. IEEE 802.11e also recommends the use of admission control for QoS guarantee, but specifies no concrete admission control mechanisms.

Some existing works focus on the call admission control mechanisms in 802.11 WLAN. An probe flow based admission control scheme is proposed in [2], in which system does the admission decision according to the end-to-end delay detected by the probe flow. A measure based call admission control scheme is proposed in [4], in which the access point (AP) detects the channel busy ratio, then does admission decision. These two schemes do CAC decision according to the system’s current conditions, but during the whole call hold time, the condition changes, so the QoS still can not be guaranteed. Author in reference [3] summarized several model based CAC mechanisms. These model based CAC mechanisms are mainly based on the system throughput rather than the QoS parameters of each users.

In this paper, we propose a two-level admission control mechanism for QoS guarantee in 802.11e system under infrastructure mode. This mechanism first uses a performance analytic model proposed in [7] to achieve the system capacity under certain QoS requirements. The channel utilization ratio is also monitored as a parameter of system’s current traffic load. Admission decisions for QoS application are performed according to these parameters. p-NACK mechanism is also proposed for rate control of best effort applications according to these parameters. Simulation shows the performance of the proposed admission control mechanisms under different traffic load.
1. System Analysis

In this paper, a micro cellular network consisting of homogeneous cells is considered, where each cell behaves identically from stochastic point of view. Therefore, we consider the stochastic behavior of the whole network by taking into account an arbitrary cell without loss of generality. IEEE 802.11e EDCA is adopted by the considered cell. This cell has one AP, and it works in infrastructure mode without RTS and CTS. Stations in the cell are divided into \( N \) classes. The number of stations of class \( i \) is denoted by \( n_i \). Vector \( n = \{n_1, n_2, \cdots, n_N\} \) is used to denote the user numbers of all classes.

In this part, in section 1.1, we use a performance analytic model to get the system capacity based on applications’ QoS requirements. Then in section 1.2, we analysis the call level traffic model and reserve source for handoff users according to the system capacity and user’s call level QoS requirements. In section 1.3, we analysis the QoS parameter’s relation with the channel busy ratio, and choose proper threshold for CAC and rate control in packet level.

1.1 Analysis for maximum admitted users

EDCA still lacks the ability to allocate specific bandwidth for their users with heterogeneous QoS requirements. Thus, We consider delay and packet loss are two important parameters for the QoS guarantee. Call admission control is proposed to guarantee that current admitted user will not degrade the QoS of existing users. To achieve this goal, we defined the range \( N = \{n\} \).

Each vector \( n \) in \( N \) make the system performance satisfy

\[
\begin{align*}
P(d_i > D_{m_i}) &< \varepsilon. \\
P(p_l > P_{m_l}) &< \varepsilon.
\end{align*}
\]

(1)

Here \( d_i \) and \( p_l \) denotes the delay and the packet loss rate of a packet of class \( i \) station respectively. \( P_{m_l} \) and \( D_{m_l} \) denote the maximum packet loss rate and maximum access delay of a packet of class \( i \) station according to the QoS requirements.

Thus, \( N \) denotes the system capacity under specific QoS requirements. If current system’s \( n \) is in \( N \), then the delay and packet loss requirements is statistically guaranteed. Edge points of \( N \) denotes the maximum admitted user number in different conditions. In our previous researches[7], we have proposed an analytic model. This model achieves the relation between QoS parameters and user numbers under specified system conditions. Thus, \( N \) can be achieved with specific QoS requirements.

Here we briefly introduce the performance analytic model. We consider a station of access class \( i \). Packets queue in the sender buffer with the FIFO order. Buffer of such a station can
be modeled as a queueing system. The system’s arrival process is this station’s packet generate process, the service time is the sum of back-off delay and transmit delay. Experiments and former analysis show that the service time can be approximated as an exponential distribution with average $\mu_i^{-1}$ [4]. Thus, the access process of such a packet of a class $i$ station is a GI/M/1 queueing model, and the whole access delay can be achieved. The average backoff delay $d_i$ is required to get the $\mu_i^{-1}$. Two dimensional Markov process is used to analysis the backoff process, and then the transmission probability $\tau_i$ is achieved. Then $d_i$ and $p_{li}$ can be obtained with the following equations.

$$d_i = \sum_{l=0}^{\infty} P\{C_k = l\} \left( \sum_{k=0}^{CW_i-1} P\{B_{il} = k\} \sum_{t=0}^{k} d_{ilt} \right).$$ (2)

$$p_{li} = p_i^{R_i+1}.$$ (3)

Here the $C_k$ denotes the number of collisions that an AC $i$ station has involved, $B_{il}$ denotes the backoff counter of a class $i$ station that has involved in $l$ times of collision. $CW$ is the current contention window of class $i$. $R_i$ is the maximum retry time. $p_i$ denotes the collision probability when a packet of class $i$ is transmitted. $d_{ilt}$ denotes the average time duration of the time between $t-1$ and $t$th backoff counter decrements of a class $i$ packet after the $lth$ collision. $p_i$ and $d_{ilt}$ can be calculated out with the $\tau_i$. $\tau_i$ is calculated with $n$. Thus we get the relation between $d_i$, $p_{li}$ and $n$.

After getting $N$, according to the resource arrangement for each class, we can choose a $n^* = \{n_1^*, n_2^*, \ldots, n_N^*\}$ in $\mathbb{N}$ to perform admission decisions. To fully utilize the bandwidth of the WLAN, $n^*$ is always chosen in the edge of the range $\mathbb{N}$ to achieve higher system throughput.

### 1.2 Handoff and traffic model

When the number $n_i^*$ of admitted users in class $i$ is decided, the system for users of class $i$ can be treated as a service system with fixed number of channels as shown in figure 1. To guarantee the users’ QoS requirement on call level, we reserved specific channel number for handoff users. The reserved channel number of class $i$ is denote as $n_{r_i}$. It is calculated out by using queueing theory, and it guarantees that the handoff call dropping probability is below $\varepsilon$.

To achieve $n_{r_i}$, we set up a call level traffic model for the above service system. In this model, calls that are being served in a cell will release their resources in two conditions: the calls are finished in current cell and the mobile terminals move into a neighboring cell with the calls unfinished. The time of station of $i$ dwelling in current cell is defined as $T_{w_i}$, and the total
duration time of a call without being forced into termination is defined as $T_{c_i}$. We assume that $T_{w_i}$ and $T_{c_i}$ are exponentially distributed and independent from each other. The average value of $T_{w_i}$ and $T_{c_i}$ are $\mu_{w_i}^{-1}$ and $\mu_{c_i}^{-1}$ respectively. Thus, the channel holding time of a call $T_{h_i}$ is also exponentially distributed with the means $\mu_{h_i}^{-1}$ of $\mu_{h_i} = \mu_{w_i} + \mu_{c_i}$. Assuming a uniform density of terminals throughout the area and an arbitrary direction of movements with respect to the cell boundary, $\mu_{w_i}$ can be given by

$$\mu_{w_i} = \frac{E(v)L}{\pi A}. \quad (4)$$

where $E(v)$ is the average speed of terminals in the cell, $L$ is the length of the perimeter of the cell and $A$ is the area of the cell.

Handoff is taken into account in our system. In this paper, we assume that the arrival processes of new calls are Poisson distributed with rates $\lambda_{n_i}$, while $\lambda_{h_i}$ denotes the arrival rate of handoff calls. Based on an equilibrium homogeneous mobility pattern, the mean number of incoming users into a cell is equal to that of outgoing ones from the cell. Thus, we have

$$\lambda_{h_i} = E(n_i)\mu_{w_i}. \quad (5)$$

Here, $E(n_i)$ is the average number of the class $i$ stations. With the above arrival, handoff and service parameters, we can achieve $n_{r_i}$ for each class $i$.

### 1.3 Channel utilization ratio

Section 1.1 derived the maximum admitted user number in call level timescale. In practical scenarios, the load of the access channel is time varying. We adopt channel utilization ratio to show the channel’s traffic load in a short time scale [4]. The channel utilization ratio $R_b$ is the ratio of the time the channel is busy to the total time. The busy time includes both successful transmissions and collisions. This ratio can be obtained by the hardware of current
802.11 network adapter cards. The $R_b$ can also be achieved with the former analytic model with equation 6.

$$R_b = \frac{p_c T_c + p_s T_s - AIFS(p_c + p_s)}{p_c T_c + p_s T_s + p_c \sigma}.$$  \hspace{1cm} (6)

Here, $p_s$, $p_c$ and $p_e$ denote the probability that a slot including successful transmission, collision or empty respectively. $T_s$, $T_c$ and $\sigma$ denotes the average time for successful transmission, collision or empty slots respectively.

Figure 2 illustrates the function of average backoff delay as a function of channel busy ratio achieved by the analytic model. Two kinds of applications are provided, one is data application, the other is voice. The channel busy ratio increases with the increase of station numbers. From the results, we can see when the channel busy ratio is below 0.8, the delay of data and voice both increases slightly. When the channel busy ratio is between 0.8 and 0.9, the delay increased remarkably but can still support some realtime applications. When the channel busy ratio is beyond 0.9, the delay increased dramatically, and not satisfied the QoS requirement. The maximum throughput of the system is also achieved when the $R_b$ is between 0.8 and 0.9. When $R_b$ is beyond 0.9, the system comes into saturated conditions. So, we can choose a threshold between 0.8 and 0.9 to perform rate control and admission control. Thus, QoS requirements of the admitted users is guaranteed.
2. Admission Control Mechanism

We propose a dynamic admission control mechanism in two levels, call admission control in call level timescale and rate control in packet level timescale. We denote users of higher priority classes with QoS requirements as QoS users. The call level admission control is proposed to avoid the new admitted QoS user to deteriorate the QoS of existing QoS users. While the packet level rate control is proposed to avoid the best effort users to deteriorate the QoS of QoS users. The proposed admission control scheme is implemented in QoS access point (AP) of 802.11e. Those QoS APs collect traffic and channel information. Before setting up a new call with QoS requirements, each station first transmits an ADDTS (add traffic stream) request message to the AP. We suppose in handoff scenarios, an ADDTS request is also sent to the new AP. The ADDTS request message contains the class information of the traffic stream (TS). When receiving a new ADDTS, the AP knows the class information of the new station, and matches it to class $i$.

Now we propose a two-level admission control algorithm to decide whether accept a coming call of class $i$ or not. The algorithm is shown in figure 3.

In this algorithm, when receive a ADDTS, we first judge the class $i$. If it’s belongs to the
lowest priority best effort application, it is accepted. Then rate control is applied on it in packet level timescale. If the call belongs to QoS users, for handoff QoS users, only one metrics $R_b$ is inspected for CAC decision, while for new QoS users, $R_b$ and $n_i$ are both inspected for CAC decision. $n_i$ stands for the average long time system load condition, while the $R_b$ stands for the traffic load condition of the system in a much short timescale. The threshold $R_i$ is chosen to denote a worse condition than $n_i^*$. Thus the handoff users have higher priority than the new users. Users of different classes also have different $R_i$ to achieve different priority.

Rate control in packet level is only carried out on best effort applications, here we assume the best effort is class 0. If $R_b < R_0$, the QoS AP does not send ACK to the best effort user with probability $1 - \frac{n_0^*}{n_0}$. Thus the efficient bandwidth for $n_0$ best effort users is equal to the sum of the required bandwidth of $n_0$ users. This rate control mechanism is called the $p$-NACK mechanism. This mechanism will enlarge the backoff window of best effort users if the system traffic load is too heavy. Thus, the efficient bandwidth of best effort users is restricted while their effect on QoS users is controlled.

3. Simulation Results

In the analysis, we consider a single cell providing VoIP and data combined service. A practical voice traffic model is presented as an example using the conversation model specified in the ITU P.59 recommendation[5]. The traffic and packet parameters of voice and data are listed in table 1.

<table>
<thead>
<tr>
<th>parameters</th>
<th>value</th>
<th>units</th>
</tr>
</thead>
<tbody>
<tr>
<td>packets interval $T_i$(voice)</td>
<td>20</td>
<td>ms</td>
</tr>
<tr>
<td>packet size(voice)</td>
<td>40</td>
<td>Bytes</td>
</tr>
<tr>
<td>mean packets size(data)</td>
<td>500</td>
<td>Bytes</td>
</tr>
<tr>
<td>average interval(data)</td>
<td>100</td>
<td>ms</td>
</tr>
<tr>
<td>$\mu_{\omega,1}(voice)$</td>
<td>120</td>
<td>s</td>
</tr>
<tr>
<td>$E(v)$</td>
<td>0.2</td>
<td>m/s</td>
</tr>
<tr>
<td>$L$</td>
<td>100</td>
<td>m</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>0.01</td>
<td>/</td>
</tr>
</tbody>
</table>

The QoS requirements of VoIP application include delay and packet loss rate. Suppose the end-to-end delay of voice is required to below 150ms and the packet loss rate is below 3% [6]. After reducing the queueing delay in the network(about 50ms), codec and packet process delay(20ms)
Table 2  Parameters of 802.11e

<table>
<thead>
<tr>
<th>parameters</th>
<th>value</th>
<th>units</th>
</tr>
</thead>
<tbody>
<tr>
<td>slot time</td>
<td>20</td>
<td>µs</td>
</tr>
<tr>
<td>SIFS</td>
<td>10</td>
<td>µs</td>
</tr>
<tr>
<td>PHY header</td>
<td>192/96</td>
<td>µs</td>
</tr>
<tr>
<td>ACK</td>
<td>14</td>
<td>Bytes</td>
</tr>
<tr>
<td>Mandatory data rate</td>
<td>1</td>
<td>Mbps</td>
</tr>
<tr>
<td>Data rate</td>
<td>11</td>
<td>Mbps</td>
</tr>
<tr>
<td>$CW_1^{min}$ (voice)</td>
<td>16</td>
<td>/</td>
</tr>
<tr>
<td>$CW_2^{min}$ (data)</td>
<td>32</td>
<td>/</td>
</tr>
<tr>
<td>$A_1$ (voice)</td>
<td>1</td>
<td>/</td>
</tr>
<tr>
<td>$A_2$ (data)</td>
<td>2</td>
<td>/</td>
</tr>
<tr>
<td>$m_1$</td>
<td>2</td>
<td>/</td>
</tr>
<tr>
<td>$m_2$</td>
<td>6</td>
<td>/</td>
</tr>
<tr>
<td>R</td>
<td>6</td>
<td>/</td>
</tr>
</tbody>
</table>

Figure 4.  Delay of voice and data.

and receive jitter buffer delay (average 60ms), we suppose the average access delay is below 20ms and the average back-off delay below 10ms, thus 99% back-off delay is below 30ms.

Using the above parameters, simulation is proposed to show the performance of our CAC mechanism. In the simulation scenario, voice and data stations’ arrival rate $\lambda_{n_i}$ gradually increase, it means that the system traffic load gradually becomes heavier. Thus, the simulation results in figure 4-6 show the performance of the CAC mechanism under different traffic load. And here $n_1^*$ for voice user is chosen as 20, the chosen $n_{r1}$ is changed with the user arrival rate, and in this simulation, it changes from 1 to 2. And the call drop rate during the whole simulation is 0.6
From the results of figure 4 and 5, we can see our CAC mechanism can statistically guarantee the QoS of the real-time VoIP stations. It keeps the backoff delay of VoIP stations always below 10ms and the packet loss rate below 2%, at the same time the delay and packet loss of data stations without QoS requirements gradually increase with the traffic load. From figure 6, we compare the two-level CAC mechanism with a mechanism only using call level admission control. Using the compared CAC mechanism, when the traffic load increases, the admitted voice users number decreases to keep the QoS of voice user. With our mechanism, the rate control avoids data load’s effect on the voice users, and the whole throughput of all voice users
keeps slightly fluctuating around a fix number according to the chosen \( n_i^* \). The results also show that the data users utilize remained bandwidth resource by using our mechanism, so the whole throughput of all user also keeps fluctuating around some number near the system’s maximum throughput.

4. Conclusion

In this paper, we propose a two-level cross-talked admission control mechanism to guarantee the QoS requirements of realtime applications. This mechanism needs not modify the existing MAC protocol. In this mechanism, admission control is proposed in both call level and packet level. The call level admission control is proposed to avoid the new admitted QoS user deteriorating the QoS of existing QoS users. While the packet level rate control is proposed to avoid the best effort users deteriorating the QoS of QoS users. The admission control schemes in two level have cross-talk and adopt information of each other, including packet arrival characters, call level traffic model and channel utilization ratio. Thus, the QoS requirements of realtime multimedia users can be statistically guaranteed while the system bandwidth is efficiently utilized. Simulation results show that by using our admission control scheme, the QoS parameters of VoIP application such as access delay and packet loss rate can be well guaranteed while the best effort data users can utilize the remained bandwidth resource. Meanwhile, the reserved resource guarantees the handoff users’ call drop rate, and the rate control scheme guarantees that the increase of data user number has no impact on the QoS of VoIP users.

References