RTS/CTS Based Endpoint Admission Control for VoIP Over 802.11e

John Fitzpatrick, Seán Murphy, and John Murphy

Performance Engineering Lab,
UCD School of Computer Science and Informatics,
University College Dublin, Belfield, Dublin 4
John.fitzpatrick@ucd.ie

Abstract. In this paper an endpoint-based admission control mechanism for VoIP over WLAN is proposed. The mechanism operates by first measuring the utilisation of the channel and comparing this to a pre-determined threshold, if the channel utilisation exceeds this threshold, the call is rejected. One important aspect of the mechanism is that it can operate in the presence of hidden terminals. This is done by using the Request-to-Send/Clear-to-Send WLAN signalling to determine channel utilisation. The scheme is designed to be flexible; it can operate with heterogeneous data rates, with varying traffic types and in the presence of legacy 802.11 nodes. The scheme was developed and evaluated using the Qualnet network simulator. An empirical approach was used to determine appropriate admission thresholds. Then, simulations were performed to demonstrate the successful operation of the scheme.

Keywords: Call Admission Control, VoIP, 802.11e, RTS/CTS.

1 Introduction

WLAN and VoIP are two technologies that have seen unprecedented growth in recent times. The use of VoIP via WLAN access is still quite uncommon, however. The release of multimode terminals supporting applications such as Skype or Googletalk heralds a significant change for use of VoIP over wireless access, but there are still some issues to be addressed in the case of WLAN.

Unlike cellular systems, WiFi was not designed to transport delay-sensitive data. The legacy IEEE 802.11 MAC [1] uses a contention based mechanism that provides best effort throughput for all traffic. This mechanism is inadequate for supporting delay-sensitive, realtime applications such as VoIP; a one-way delay of no more than 150ms is required to deliver PSTN quality, although delays of up to 400ms can be tolerated [2]. Loss thresholds are somewhat codec dependent, but, in general, losses in excess of 3-4% are problematic.

To address this problem, the IEEE developed the 802.11e standard [3] to introduce QoS support to 802.11. A new contention mechanism called Enhanced Distributed Channel Access (EDCA) was developed as part of the standard. It provides QoS support by prioritising channel access for real time applications.

While EDCA can protect real time applications in the presence of best effort traffic, the presence of too many users generating high priority traffic can result in

A. Helmy et al. (Eds.): MMNS 2006, LNCS 4267, pp. 223 – 234, 2006.

unacceptable QoS for priority applications. For this reason, call admission control mechanisms are needed to prevent existing VoIP calls from being degraded due to the addition of further calls. If one more VoIP call than can be supported by the network is added, the quality of all existing calls becomes unacceptable.

The call admission control scheme proposed in this paper is an endpoint centric approach which requires no modifications to the WLAN infrastructure. More specifically, it does not require modifications to APs or routers in the network. The mechanism operates exclusively in the terminal and hence only an upgrade to software in the mobile terminal is needed to realise the scheme.

The admission control mechanism operates by measuring the utilisation of the radio channel. This is then compared with a pre-determined threshold. If the channel utilisation is below the threshold, the call is admitted; else the call is rejected.

To make accurate AC decisions the collective channel utilisation of all nodes must be correctly estimated. This can be problematic in the presence of hidden nodes[1]. To overcome this problem, it is assumed here that the RTS/CTS mechanism is used for transmission of all packets. While this does incur an extra overhead, the advantage of using this approach is that each node can detect activity generated by hidden nodes.

This paper is structured as follows. Section 2 provides a brief overview of some related work in the area of admission control for WLAN. Section 3 discusses WLAN technologies relevant to the proposed scheme. In section 4 the design and implementation of the proposed scheme is discussed. Section 5 describes the simulation setup and is followed by a discussion of the results obtained. This paper is then concluded in section 6.

2 Related Work

A large amount of the related work in the area of call admission control (CAC) over WLAN is access point (AP) centric. Most approaches require network infrastructure modifications such as the introduction of non-standard APs. Standardised solutions based on 802.11e have been considered but these fail to provide admission control for legacy 802.11a/b/g equipment.

2.1 Distributed Admission Control (DAC)

DAC, originally included in earlier drafts of 802.11e but later removed, was proposed by the IEEE 802.11e working group to protect active QoS streams such as VoIP and video in the EDCA mode of operation. The AP announces a transmission budget as part the beacon frame; the transmission budget specifies the amount of extra transmission time available for each access category (AC) over the next beacon interval. To calculate the transmission budget the AP measures the amount of utilised transmission time for each AC per beacon interval. The utilised transmission time is then subtracted from the transmission limit to obtain the transmission budget. This is done individually for each AC. From the transmission budget announced by the AP,

[1] The hidden node problem refers to a situation in which two or more nodes are in range of the access point (AP), but out of radio range of one another. Consequently the transmissions of multiple nodes may interfere due to simultaneous transmission, leading to a collision.

each station sets its own transmission budget for each AC based on the successfully used transmission time during the previous beacon interval. When the budget for an AC is depleted, no new streams will be allowed and all existing flows are prevented from increasing their transmission time. However, as previously stated, it fails to provide admission control for legacy equipment.

2.2 VMAC and VS

The Virtual MAC (VMAC) and virtual source (VS) proposed in [5, 6] are a set of algorithms that emulate the MAC and source applications in order to estimate achievable QoS. VMAC and VS operate in parallel with the real MAC and monitor the load on the radio channel to estimate service level quality metrics such as delay and loss. The VS consists of a virtual application and a virtual queue. The virtual application produces data packets just as a real application would, the packets are then placed in the interface queue. The VMAC performs scheduling, CSMA/CA and random back off just as a real MAC would. However no packet is transmitted, the VMAC estimates the probability of a collision if the packet was sent. A collision is detected when another node chooses the same time slot to transmit. In this case the VMAC enters a back off procedure as a real MAC would. The VMAC algorithm measures delay and loss on the uplink of the client. However, the limiting factor for a VoIP call is usually on the downlink as the AP which has to contend for the medium many more times than each client and therefore becomes a bottleneck. Another issue with this approach is that VMAC can only monitor the radio channel of nodes that are within radio range of each other and hence cannot give a true estimation of achievable QoS that would be delivered by the network due to the presence of hidden terminals.

3 Channel Access Mechanisms

This section gives an overview of DCF and EDCA channel access mechanisms as they are relevant to the proposed CAC scheme. Also, the RTS/CTS mechanism that allows the proposed scheme to operate in the presence of hidden terminals is also discussed.

3.1 Distributed Coordination Function (DCF)

The IEEE 802.11 MAC uses a contention based channel access mechanism called DCF. Although DCF is much written about and is well understood, this section is included for completeness.

DCF employs carrier sense multiple access with collision avoidance (CSMA/CA) to enable distributed channel sharing. Each DCF station with a packet to transmit must sense the wireless channel and determine it to be idle for a period equal to one distributed interframe space (DIFS) before proceeding with transmission. If the channel is determined to be busy, the station will wait until the channel becomes idle for a period defined by the DIFS. After detecting the channel to be idle for DIFS, the station waits a random backoff interval before attempting to transmit a frame.

DCF implements a binary exponential backoff to select the backoff interval. The backoff interval is uniformly selected over the interval [0, CW-1], where CW is the current contention window size in units of timeslots. CW has an initial value of $CW_{\min}$ for the first transmission attempt. After each unsuccessful transmission attempt the CW is doubled to a maximum value of $CW_{\max}$. On receiving acknowledgement of a successfully received frame the CW is reset to $CW_{\min}$. The backoff counter is decremented once every timeslot after the medium is idle for a DIFS period. If the medium becomes busy during a backoff interval the backoff counter is paused and resumes again after the medium becomes idle for a DIFS. A transmission is attempted when the backoff counter reaches zero.

In DCF each successfully received frame must be acknowledged by the receiving station. Upon successful reception of a frame the receiving station transmits an ACK frame after a period equal to the short interframe space (SIFS). The transmitting station assumes the frame to be lost if an ACK frame is not received within a specified ACK timeout, it then reschedules the frame to be retransmitted. As SIFS is smaller than DIFS, ACK frames will access the medium earlier than data frames, meaning they have priority. This means that other stations are prevented from accessing the medium during this short period before the ACK can be sent, as they must wait for at least DIFS before attempting any transmissions.

3.2 RTS/CTS

The CSMA/CA mechanism implemented by DCF suffers from the hidden node problem. To combat the hidden node problem, the request-to-send/clear-to-send (RTS/CTS) mechanism was developed. The RTS/CTS is a four way handshake mechanism designed to reduce the number of frame collisions. Before a data frame can be transmitted a successful handshake must take place between the sending and receiving nodes using RTS/CTS control frames. In general RTS/CTS is only used for large packets due to the signalling overhead the mechanism introduces. For small packets it is more efficient to perform a retransmission if a collision occurs.

Using RTS/CTS, the sending node transmits an RTS control frame containing its source address as well as the intended destination address and required channel time for the transmission of the data frame. Having successfully received the RTS frame, the destination node transmits a CTS frame after a SIFS. The CTS frame contains only the sending node address and the required channel time. This serves two purposes, it informs the sender to proceed with transmission of the data frame and it also notifies other nodes of the impending transmission duration so they will not attempt a transmission, essentially reserving the channel for the specified period to allow for the transmission of the data frame and corresponding ACK.

3.3 Enhanced Distributed Channel Access (EDCA)

The legacy 802.11 MAC provides distributed channel access shared equally among all users and traffic types, which works well for best effort traffic. However, the legacy DCF mechanisms cannot provide the QoS needed by real time delay sensitive applications such as VoIP. EDCA is a contention based channel access mechanism described in 802.11e, it provides an enhancement to DCF allowing for QoS support based on prioritized access to the medium. Prioritization is achieved through eight prioritizations (0-7) that further map to four access categories (AC) (0-3), with each access category obtaining differentiated channel access. Each packet is passed to the

MAC layer with a specific priority value specified in the IP packet header. The EDCA mechanism then maps the frame to a particular AC based upon the priority value. The mappings are shown in table 1.

Table 1. Priority & Access Category Mappings

Traffic Type	Priority	Access Category (AC)
Best Effort	0	0
Best Effort	1	0
Best Effort	2	0
Video	3	1
Video	4	2
Video	5	2
Voice	6	3
Voice	7	3

An AC within a station independently contends for the medium using a backoff process similar to DCF as previously discussed. Each AC has different parameters associated with the backoff mechanism, which give the different channel access prioritizations.

The AC dependent EDCA parameters are $AIFS[AC]$, $CW_{\min}[AC]$, and $CW_{\max}[AC]$, which replace the DCF parameters $DIFS$, $CW_{\min}$, and $CW_{\max}$ respectively. $AIFS[AC]$ (Arbitration InterFrame Space) is used instead of $DIFS$ as the minimum time the channel must be sensed idle before a transmission can proceed.

As each AC maintains independent backoff counters, an issue known as a *virtual collision* can arise, in which two or more ACs within a single station finish backoff during the same timeslot. When this occurs the AC with the highest priority is allowed to transmit, and the other ACs double their CW values and recontend for channel access.

4 RTS/CTS Based Admission Control

In this section the proposed AC scheme is described. The essence of the proposed scheme is to use transmission duration information contained in the header of RTS/CTS control frames to estimate the current utilisation of the wireless channel. The channel utilisation estimate can then be used to make admission control decisions for VoIP calls. The use of RTS/CTS mitigates any hidden node problems and allows each station to passively monitor network activity.

4.1 Call Admission Decision

The call admission decision is performed by the application, utilising lower layer metrics. The decision to admit or reject the call is based upon whether the channel can support the additional call without adversely affecting calls already in progress.

Each call added to the network increases the channel utilisation by an amount dependent on the station's transmission rate. When the capacity of the channel is exceeded in the downlink direction, the end-to-end delay experienced by the VoIP calls rises dramatically, as shown in [8]. The increase in end-to-end delay greatly reduces the VoIP quality experienced by all users present in the network.

By using 802.11e real time traffic takes prioritisation over best effort. As was shown in [7], as the volume of high priority real time traffic increases; the throughput of low priority best effort traffic approaches zero. This allows the proposed scheme to operate in the presence of best effort traffic.

The admission decision uses predetermined channel utilisation threshold values, the determination of which will be discussed in the results section. If the channel utilisation as measured by the node when the call attempt is being made is greater than the threshold value, the call is rejected, else it is admitted. The threshold values chosen are the maximum channel utilisation values at which one extra VoIP call can be added to the network without excessively increasing the delay so as to reduce the QoS of the admitted and existing call. Each threshold value is chosen so as to allow one extra call to be added without overloading the network, such that when a node performs an admission control decision using the threshold values and the call is admitted it will be supported.

4.2 RTS/CTS Control Frames

Having successfully contended for the medium a station must transmit an RTS control frame to the intended recipient of the impending data frame. The transmission duration field of the control frame is calculated based upon the current physical layer transmission data rate, the data frame size and the basic data rate at which all control frames will be transmitted. Upon successful reception of an RTS frame the recipient responds with a CTS frame. The CTS header duration field contains the transmission duration specified in the RTS frame, less the time taken to transmit the RTS frame a SIFS period. In other words each header duration field of each control frame specifies how much longer the channel will be occupied for. This allows other stations to backoff until after the transmission duration has finished.

The proposed mechanism is intended for use in an infrastructure mode WLAN, hence all traffic is transmitted via the AP. Each station in the network monitors all RTS/CTS control frames originating from the AP. RTS frames originating from the AP will have a source address matching that of the AP. As CTS frames contain only a destination address, any CTS frames received whose destination address does not match that of the AP are assumed to be destined for other nodes in the same BSS and as such must have been transmitted by the AP (where infrastructure mode of operation is assumed).

Each stations MAC layer maintains a channel utilisation timer that is incremented using the transmission duration from the control frames sent by the AP. When an RTS frame is received from the AP, the channel utilisation timer is simply updated with the value specified in the duration field of the RTS frame. However, when a CTS frame is received from the AP the incremented time is calculated as the duration specified in the CTS header with the addition of the transmission duration of the RTS packet and a SIFS. This corrects for the smaller transmission duration specified by CTS frames than by RTS frames.

Every user definable period T, the channel utilisation is calculated as a percentage of total channel time that equals the period T over which the measurement is being performed. The utilisation timer is then reset for the next measurement period. AP beacon frames and other management traffic that do not use RTS/CTS control frames are ignored as the amount of channel time occupied by these is neglible.

4.3 Cross Layer Communication

In the proposed scheme, cross layer communication takes place between the application layer and both the MAC and physical layers. The admission decision procedure is performed at the application attempting to make a VoIP call. The MAC layer channel utilisation metric and the physical layer transmission data rate are passed to the application upon receipt of application layer triggers. These parameters are then be used to make the admission control decision. The data rate is needed to determine channel utilisation of the flow to be admitted, this will be discussed further in the next section.

5 Simulation Results

Implementation and simulations of the proposed CAC scheme were carried out using the Qualnet network simulator [9]. Development of the scheme required modification to multiple layers within Qualnet, in particular the 802.11 MAC and CBR application.

This section provides a brief overview of the simulation setup and a discussion of the results obtained.

5.1 Simulation Setup

The wireless environment was simulated using the 802.11e MAC layer and the 802.11g physical layer giving a maximum data rate of 54Mbps. To simulate VoIP traffic, constant bit rate (CBR) sources are used. Each CBR source provides a stream of packets that simulates simplex voice data. Full duplex traffic was modelled using two simplex traffic sources, one on the uplink and one on the downlink. The motivation behind simulating voice calls as full duplex was based on the most commonly used VoIP program Skype, which uses full duplex (i.e. no silence suppression is used)[2].

The VoIP codec used for the simulations was G.711 with a frame size of 20ms and a packet size of 160 bytes. This was chosen as it gives the highest voice quality of all standardised VoIP codecs. Lower bit rate codecs such as G.729 were also considered. However, it was found that using lower bit rate codecs did not offer any increase in capacity or throughput and simply reduced the maximum attainable quality of each call. This is due to the overhead introduced by the four way handshake of RTS/CTS, as each RTS/CTS control frame is transmitted at a low data rate.

---

[2] Skype uses full duplex for two reasons. Transmitting silent packets maintains UDP bindings at NAT (Network Address Translation). Also, if data is being transmitted over TCP the silent period packets prevent a reduction in the congestion window size during the silent period.

5.2 Determining the Channel Utilisation Thresholds

Simulations were performed to calculate the channel utilisation threshold values that are used by the application layer to make CAC decisions. The amount of channel utilisation a station absorbs is dependent on the physical data rate of the station. Stations in low data rate zones of the WLAN require more channel time to transmit their data and hence place more load on the network.

To calculate the threshold values individual experiments for each of the 802.11g supported data rates were performed. In each experiment a single VoIP call at the data rate being investigated was initially added to the system with subsequent 54Mbps calls being added incrementally until the capacity was exceeded. Each experiment was run for 1000 seconds with the channel utilisation being measured every 100ms. The mean and standard deviation of the channel utilisation was then calculated at the end of each simulation. Each data point in the presented results is the average of five such simulation runs.

The threshold value for each case could then be obtained by examining the channel utilisation as measured by each station as the additional calls were added. The threshold value for each data rate was chosen as the mean value plus twice the standard deviation of the channel utilisation two calls before capacity was exceeded. The mean channel utilisation value two calls before the capacity is exceeded, was used so that when a station measures the channel utilisation and it is found to be below the threshold value the additional call admitted will not exceed the system capacity.

Figure 1 and figure 2 shows the channel utilisation and average end-to-end delay for a varying number of 54Mbps calls. As can be seen the system capacity is 24 VoIP calls at the highest possible data rate of 54Mbps. When the 25th call is added to the system the end-to-delay increases dramatically and can not be supported[3].

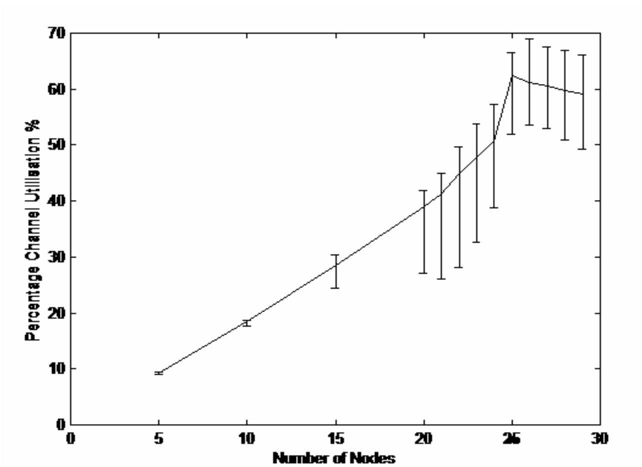

Fig. 1. Channel Utilisation for varying number of 54Mbps calls

[3] Note that 802.11g has be shown to accommodate a significantly higher number of VoIP calls, however this is in the case where RTS/CTS signalling is not used.

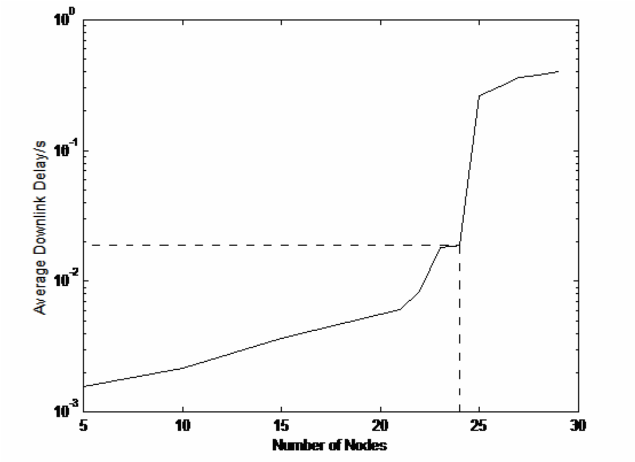

Fig. 2. End-to-End delay for varying number of 54Mbps calls

The mean channel utilisation for 54Mbps calls was found to be 48 percent, with a standard deviation of 3.5 percent. Hence, the threshold value of the mean plus twice the standard deviation was found to be 55 percent.

The same experiments were carried out for the possible 802.11g data rates, 54, 48, 36, 24, 12 and 6 Mbps. Although different data rates were used, it was found that not all data rates had different threshold values; only two differing threshold values were found to be needed. A threshold value of 55 percent channel utilisation was calculated for calls taking place in data rates greater than 12Mbps. The following figures 3 and 4 shows the channel utilisation and the delay for the 12Mbps case. Each data point in the figure 3 represents the average channel utilisation when there are X amount of calls present on the network with all calls at a data rate of 54Mbps with the exception of the final call in each case which is at a rate of 12Mbps.

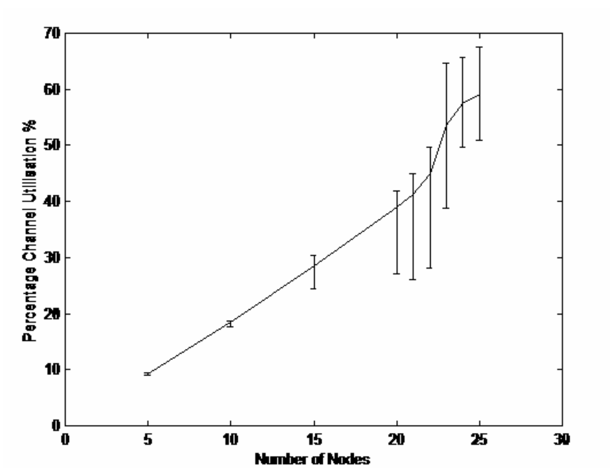

Fig. 3. Channel Utilisation (In the 12Mbps case)

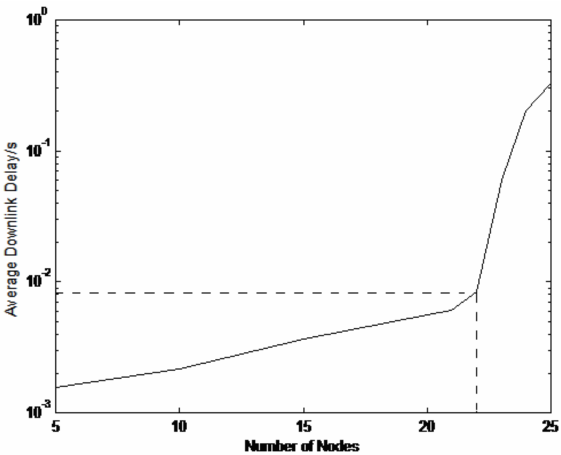

Fig. 4. End-to-End delay (In the 12Mbps case)

The same experiments were performed for the 6Mbps case and from the results obtained, the channel utilisation threshold value for calls with data rates at and below 12Mbps was found to be 49 percent, calculated using the mean value plus twice the standard deviation, as done previously.

5.3 Admission Control with Heterogeneous Data Rates

Figure 5 shows the proposed call admission control mechanism in operation. A simulation network is setup such that there are five stations in each data rate zone of the WLAN. Beginning at 500 seconds and in increments of 10 seconds, each station attempts to make a G.711 VoIP call. The simulation was configured so that all stations in the 54Mbps zone start their calls first, then all stations in the 48Mbps zone and so on through all the data rates. As can be seen the channel utilisation increases as the number of VoIP calls present in the network increases.

At 710 seconds the 22nd call is added to the system. This is a 12Mbps call and at this point there are five 54Mbps calls, five 48Mbps calls, five 36Mbps calls, five 24Mbps calls and two 12Mbps calls. All previous calls have been admitted by the CAC scheme as the channel utilisation was not above the predetermined threshold values.

At 720 seconds another 12Mbps station attempts to make a VoIP call, however the channel utilisation exceeds the predetermined threshold for 12Mbps calls of 49 percent, hence the call is rejected. Another eight call attempts are made by the eight remaining stations over the next 80 seconds, these too are rejected. The maximum end-to-end delay experienced by a node in the network did not exceed 0.0104 seconds (10ms). This shows the ability of the proposed scheme to provide QoS guarantees for VoIP calls in a heterogeneous data rate network.

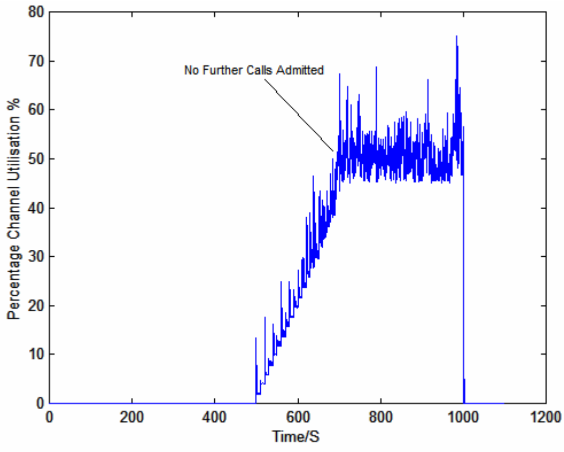

Fig. 5. Heterogeneous Data Rate Simulation

6 Conclusion

In this paper an endpoint based call admission control mechanism for VoIP over heterogeneous transmission rate IEEE 802.11g/e networks has been proposed. The scheme can operate in the presence of varying traffic types and with legacy 802.11 stations.

The CAC decision was based on an estimate of the current channel utilisation. This was determined using information transmitted in the RTS/CTS signalling messages. This was then compared to some predetermined thresholds. If the channel utilisation exceeded the threshold, then the call was rejected.

An empirical approach to determining the thresholds was used. These threshold values were then validated in a scenario with nodes connecting using different date rates.

Future work will be to further evaluate the operation of the scheme in the presence of different traffic types. Also the impact on the CAC decision of node mobility and varying calls arrival patterns needs to be assessed.

Acknowledgements

The support of the Irish Research Council for Science, Engineering and Technology (IRCSET) is gratefully acknowledged.

References

1. IEEE Std. 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 1999.
2. ITU-T, "Series G: Transmission Systems and Media Digital Systems and Networks" G.113, February 2001.

3. IEEE Std. 802.11e-2004, “Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements,” Nov. 2005.
4. D. Gao, J. Cai, and K.N. Ngan, “Admission Control in IEEE 802.11e Wireless LANs,” IEEE Network Magazine, July/August 2005.
5. M Barry, AT Campbell, A Veres “Distributed Control Algorithms for Service Differentiation in Wireless Packet Network,” INFOCOM, 2001.
6. A Veres, A Campbell, M Barry, S Li-Hsiang, “Supporting service differentiation in wireless packet networks using distributed control,” IEEE Journal on Selected Areas in Communications, October 2001.
7. H. Zhai, J. Wang, and Y. Fang, “Providing Statistical QoS Guarantee for Voice over IP in the IEEE 802.11 Wireless LANs” IEEE Wireless Communications Magazine, February 2006.
8. W. Wang, S. C. Liew, and V. O.K. Li, “Solutions to Performance Problems in VoIP over a 802.11 Wireless LAN”, VTC, January 2005.
9. “Scalable Network Technologies, Inc., QualNet simulator,” available at http://www. qualnet.com