A Bandwidth Allocation/Sharing/Extension Protocol for Multimedia Over IEEE 802.11 *Ad Hoc* Wireless LANs

Shiann-Tsong Sheu, *Associate Member, IEEE,* and Tzu-Fang Sheu, *Student Member, IEEE*

***Abstract*—In this paper, we will propose a novel bandwidth allocation/sharing/extension (DBASE) protocol to support both asynchronous traffics and multimedia traffics with the characteristics of variable bit rate (VBR) and constant bit rate (CBR) over IEEE 802.11 ad hoc wireless local area networks. Overall quality of service (QoS) will be guaranteed by DBASE. The designed DBASE protocol will reserve bandwidth for real-time stations based on a fair and efficient allocation. Besides, the proposed DBASE is still compliant with the IEEE 802.11 standard. The performance of DBASE is evaluated by analysis and simulations. Simulations show that the DBASE is able to provide almost 90% channel utilization and low packet loss due to delay expiry for real-time multimedia services.**

***Index Terms*—Ad hoc, CBR, CSMA, VBR, WLAN.**

I. Introduction

As THE speed and capacity of wireless local area networks (WLAN) increase [1], so does the demand for improving quality of service (QoS) for real-time multimedia applications. The IEEE 802.11 wireless local area networks (WLAN) standard [2] includes a basic distributed coordination function (DCF) and an optional point coordination function (PCF). The DCF uses carrier sense multiple access with collision avoidance (CSMA/CA) as the basic channel access protocol to transmit asynchronous data in the contention period. An attractive feature of CSMA/CA protocol is that it is simple to implement; however, this contention-based MAC protocol cannot guarantee transfer delay for real-time traffics. The delay bound can be provided by employing the PCF. The PCF is a polling-based protocol, which is not designed for the distributed environment. With PCF, real-time stations will access channel in a round-robin manner in each contention free period. However, the use of centralized scheme in PCF constrains the operation of WLAN. Furthermore, several researchers have pointed out the centralized protocol results in a poor performance [3], [4].

Papers [5], [6] have proposed a distributed multiple access protocol for voice services over WLAN. In this protocol, voice stations sort their access rights by jamming the channel with pulses of energy, which is named BB contention, before sending their voice packets. Since voice packets must be transmitted repeatedly in a constant interval, sending bursts of energy for each packet will waste bandwidth considerably. Moreover, the BB contention is not a regular scheme defined in IEEE 802.11 standard, thus it is difficult to be overlaid on current CSMA implementations. Although the chaining scheme proposed in [6] is used to enhance the efficiency of BB contention, the splitting of chain, which occurs while a node ends a session or its packet is corrupted, will also reduce the efficiency. Furthermore, the asynchronous data might transmit its packet in the hole of chain and make the access instants of real-time packets stretched. As a result, these real-time packets might be dropped because the delay of packet exceeds the delay-bound. Another protocol supporting real-time multimedia traffic in a WLAN based on group allocation multiple access (GAMA) was proposed in paper [7]–[9]. In GAMA, owing to at most one new session could access the channel and reserve the bandwidth in a cycle, the medium access delay might increase sharply when the traffic load becomes high. Besides, the prior reserved members have the higher priorities to occupy more bandwidth, which might decrease the qualities of the others. Thus GAMA is not a fair protocol and doesn't consider the overall QoS. A different approach to achieve the fairness and QoS guarantee in WLAN is to employ the distributed packet scheduling algorithm [10], [11]. With fair scheduling, different flows sharing a wireless channel can be allocated bandwidth in proportion of their "weights." Paper [11] presented a fully distributed fair scheduling (DFS) algorithm, which is derived from the DCF in the IEEE 802.11, for fair scheduling in WLAN. The essential idea of DFS is to choose a proper backoff interval that is proportional to the "finish tag" of packet. In DFS, a bigger weight will result in a smaller "finish tag" and packets with a smaller "finish tag" will be assigned a smaller backoff interval. Consequently, a flow with a larger weight will obtain a higher throughput. Obviously, performance of DFS depends on the weights assigned to various flows. However, how to assign appropriate weights for real-time and nonreal-time flows is still an open issue. Furthermore, the DFS does not consider the delay bound of real-time packet. Therefore, in this paper, we will propose a new distributed bandwidth allocation/sharing/extension (DBASE) protocol to support both asynchronous and multimedia traffics over IEEE 802.11 *ad hoc* WLAN where no fixed access point coordinates accesses.

For simplicity, a station with real-time (nonreal-time) traffic is denoted as rt-station (nrt-station) throughout this paper.

Manuscript received December 20, 2000; revised May 5, 2001. This work was supported by the National Science Council, Taiwan, R.O.C., under Contract NSC89-2218-E-032-012. A preliminary version of this paper appeared in the *Proc. IEEE INFOCOM,* 2001.

S.-T. Sheu is with the Department of Electrical Engineering, Tamkang University, Tamsui, Taipei Hsien, Taiwan, R.O.C. (e-mail: stsheu@ee.tku.edu.tw).

T.-F. Sheu is with the Institute of Communication Engineering, National Tsing-Hua University, Hsin-Chu, Taiwan, R.O.C. (e-mail: sunnie@ieee.org).

Publisher Item Identifier S 0733-8716(01)08471-2.

In the proposed DBASE protocol, time-sensitive rt-packets always have higher priorities than ordinary nrt-packets to make sure the rt-traffic meets the delay restrictions. To do this, the rt-stations will transmit their rt-packets during the contention free period (CFP) in every superframe. To obtain the periodic access right, a new rt-station needs first contend the channel for reservation before sending its rt-packets. A modified CSMA/CA protocol, which is still compatible with IEEE 802.11 standard, is proposed to support rt-stations have a higher priority to request the periodic reservation and access channel bandwidth than ordinary data accesses. Furthermore, to support both constant bit rate (CBR) services and variable bit rate (VBR) services over WLAN, the channel bandwidth will be dynamically allocated, shared, and released by DBASE protocol. The basic concept is that each time rt-station transmits packet it will also declare and reserve the needed bandwidth at the next CFP. Every rt-station collects this information and then calculates its actual bandwidth at the next cycle. For these rt-stations with temporary light-load, the redundant bandwidth will be shared by the other overloaded rt-stations automatically. Besides, the proposed DBASE allows donaters have the right to take their bandwidth back by scheduling them at the front of the access sequence in CFP. Once any donater desires to extend its bandwidth, it will directly access the needed bandwidth and inform the others rt-stations to recalculate their badnwdith quota. Based on this concept, the DBASE allocates the bandwidth for each session fairly and efficiently. In this paper, terms rt-station and session are interchangeable.

In the Section II, we describe the DBASE MAC protocol for both rt-traffic and nrt-traffic. The boundary condition and throughput of DBASE are analyzed in Sections III and IV, respectively. Simulation models and results are discussed in the Section V and we give some conclusions in the Section VI.

II. DBASE Protocol

In IEEE 802.11 WLAN, the independent basic service set (IBSS) is the most basic type, which IEEE 802.11 stations are able to communicate directly. This mode of operation is possible when 802.11 WLAN stations are close enough to form a direct connection without preplanning. This type of operation is often referred to as an *ad hoc* network. In this section, we will describe the access procedures for transmitting nrt-packets and rt-packets separately in an *ad hoc* network. We divide the frames into three priorities as the standard does. The frames of different priorities have to wait different inter-frame spaces (IFSs) before they are transmitted. The short IFS (SIFS) is used by immediate control frames, which always have the highest priority, such as clear to send (CTS) and acknowledgment (ACK). The priority IFS (PIFS) is used by the rt-frames, such as reservation frame (RF) and the request to send (RTS) of voice/video packets in DBASE protocol. The DCF IFS (DIFS) is the longest IFS and is used by the nrt-frames, which always have the lowest priority.

A. The Access Procedure for Asynchronous Data Stations

The basic access method for nrt-stations is based on the conventional DCF [2]. After the medium is detected as an idle duration of a DIFS period by a waiting nrt-station, who attempts to transmit a nrt-packet, the backoff procedure of this nrt-station will start. The data backoff time (DBT) is derived by

$$\mathrm{DBT} = \mathrm{rand}(a,\, b) \times \mathrm{Slot\_time}$$

where $\mathrm{rand}(a,\, b)$ returns a pseudo-random integer within interval $[a,\, b]$, which b grows exponentially for each retransmission attempt and the range of b is from $b_{\min}$ to $b_{\max}$. In the IEEE 802.11 standard, a is set to 0; $b_{\min}$ and $b_{\max}$ are set as 32 and 1024, respectively. That is

$$b = b_{\min} \times 2^r < b_{\max}$$

where r is the number of retransmission times. The Slot_time, which is defined as the time needed for a station to detect a packet, to accumulate the time needed for the propagation delay, the time needed to switch from the receiving state to the transmitting state, and the time to signal to the MAC layer the state of the channel (busy detect time) [12]. The Slot_time is set as 20 μs for DSSS PHY in [2]. The DBT counter is decreased as long as the channel is idle, and suspended while the medium becomes busy. When its DBT counter becomes zero, that means the backoff procedure is finished, then the nrt-station transmits its data packet (or RTS). When the destination receives the packet correctly, it will transmit an ACK to the source within a SIFS.

B. The Access Procedures for Real-Time Stations

In DBASE, every rt-station needs build and maintain a *ReSerVation Table (RSVT)*. This table records the information of all rt-stations that have finished the reservation procedure successfully. The information includes the access sequence, the MAC address, the service type and the required bandwidth of each rt-station. The way of a new rt-station contending for the medium is by first issuing the RTS packet to join the *ReSerVation Table (RSVT)* and reserve its needed bandwidth. Any rt-station ($\mathrm{STA_{RT}}$), which has joined into the *RSVT,* does not need to contend the medium for each transmission during whole session. In order to maintain the correct access sequence, each rt-station needs to be equipped with a sequence ID (SID) register and an active counter (AC). The SID register is used to record the access order among all active rt-stations and the AC counter is used to record the number of active rt-stations. In the following four subsections, we will describe how the rt-stations reserve, allocate, share and extend the bandwidth.

1) Reservation Procedure: If $\mathrm{STA_{RT}}$ intends to start a session at time t, it will first monitor the channel for detecting the RF in the interval $(t,\, t + D_{\max})$, where $D_{\max}$ means the smallest maximal tolerance delay among all multimedia sessions. The $D_{\max}$ is usually a predefined constant value and is strongly depending on the characteristics of rt-services. The repetition interval of rt-packets is equal to $D_{\max}$ because the real-time information delayed for more than $D_{\max}$ will cause unacceptable quality and must be discarded. Since each multimedia service type has its corresponding maximal delay bound $D_{\max}$, we should set the $D_{\max}$ as the smallest maximal tolerance delay among all multimedia sessions. The RF is a broad-

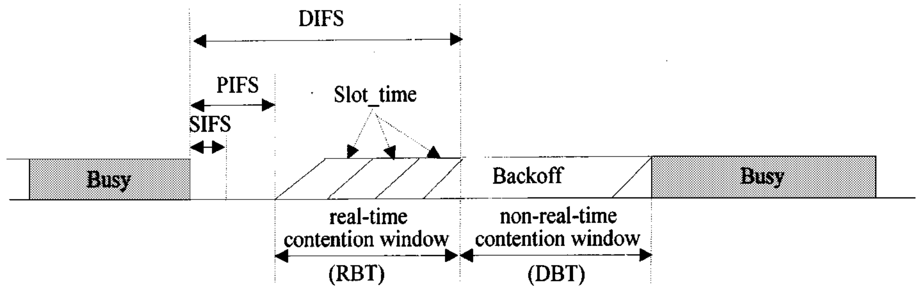


Fig. 1. The relations between three interframe spaces (SIFS, PIFS and DIFS) and two contention windows (DBT and RBT).

cast frame and is used to announce the beginning of contention free period (CFP). The RF is sent by the first active rt-station (with SID $= 1$) in the RSVT, which has the responsibility to initiate the CFP periodically. This particular rt-station is named as contention free period generator (CFPG) throughout this paper. The RF frame mainly carries the information of the number of active rt-stations (AN) in the IBSS and the information of all rt-stations recorded in the *RSVT* of CFPG. Normally, a new active rt-station $\mathrm{STA_{RT}}$ will receive the RF frame in the interval $(t, t + D_{\max})$ if there is any active rt-station already reserved the bandwidth. Otherwise, it means that there is no active rt-station and station $\mathrm{STA_{RT}}$ will be the first one. That is, when the channel still perceives idle in the interval $(t + D_{\max} + \varepsilon, t + D_{\max} + \varepsilon + \mathrm{PIFS}]$ and none RF frame is detected, $\mathrm{STA_{RT}}$ will execute the *backoff* procedure. The symbol ε denotes the remaining transmitting time of the current PDU (protocol data unit) at the time instance $t + D_{\max}$. The contention process is still based on the CSMA/CA protocol. The real-time backoff time (RBT) of a rt-station is defined as

$$\mathrm{RBT} = \mathrm{rand}(c, d) \times \mathrm{Slot\_time}.$$

The RBT counter is decreased as long as the channel is idle, and suspended while the medium is sensed busy. If RBT reaches zero, $\mathrm{STA_{RT}}$ will transmit its RTS to the destination to setup a connection. If no collision occurs, the rt-station, which sent RTS frame, will receive the CTS frame within SIFS, and then it will become the CFPG. Meanwhile, it sets both its SID and AC to one and transmits the RF frame and its rt-packet right away. The length of the first rt-packet following the RTS of each session is limited as the average length decided by its traffic type. In the DBASE, the RTS/CTS control frames for rt-packets will be generated only at the first time access. Contrarily, if collision occurs when transmitting the RTS, the P-*persistent* scheme is used to decide whether the collided stations insist on accessing channel in the next Slot_time. Such collision is detected by a rt-station who does not receive its CTS within the following SIFS period after transmitting the RTS frame. (Note that the SIFS is always shorter than Slot_time.) The collided rt-station will retransmit the RTS in the following Slot_time with a probability p. With a probability $q = 1 - p$, it will defer at least one Slot_time and contend at the next real-time contention window. That is, the deferred one will recalculate the RBT, named as RBTP, by

$$\mathrm{RBTP} = \mathrm{rand}(c + 1, d) \times \mathrm{Slot\_time}.$$

This scheme will efficiently reduce the contention resolving period. As soon as the CFPG is being generated, other rt-stations will detect the RF and the rt-packet of the CFPG and then content for the second access position in the RSVT.

If a new active rt-station ($\mathrm{STA_{RT}}$) detects the RF frame in the interval $(t, t + D_{\max})$, it knows that at least one active rt-station already reserved the bandwidth. To avoid disturbing the rt-stations access channel in the CFP, a new rt-station that wants to join into the *RSVT* by contention must wait until the CFP finishes. The length of CFP is recorded in the RF. During the waiting period, $\mathrm{STA_{RT}}$ monitors the activity of channel. If the channel idles a Slot_time during CFP, it implies a rt-station disconnects session and the CFP should be decreased accordingly. After the CFP finishes, $\mathrm{STA_{RT}}$ follows the *backoff* scheme to contend for its reservation by sending a RTS as mentioned above. While in the backoff period RBT, $\mathrm{STA_{RT}}$ keeps monitoring the channel to check whether any rt-station joins into the *RSVT* successfully. Until the RBT becomes zero, $\mathrm{STA_{RT}}$ will send its RTS. If no collision occurs, the content of AC will be increased by one and the SID of $\mathrm{STA_{RT}}$ is set as the content of AC. (We note that the AC value in each station will be periodically updated by broadcast RF frame.) At this moment, it transmits its first rt-packet right away. Based on this access procedure, every rt-station will increase their AC by one as soon as it listens a CTS. On the contrary, if collision occurs, the contention resolution also follows the P-*persistent* scheme as mentioned above.

To make sure the rt-packets can be transmitted periodically and the repetition cycle will not be longer than $D_{\max}$, we define a parameter real-time transmission period (RTP) to limit the real-time contention period. The RTP is the sum of the CFP and the real-time contention period. The former is the maximal time interval for reserved stations transmitting their packets and the latter is the time interval designed for new rt-stations contending for joining into the *RSVT*. Therefore, every $\mathrm{STA_{RT}}$ has to make sure the instant, which has finished the access, will not over the boundary of RTP. The way of determining RTP will be discussed in Section III.

To prevent nrt-station's transmissions from disturbing rt-station's transmissions, the relationships among three spaces and two contention windows are shown in Fig. 1 and must satisfy the following two constraints:

$$\mathrm{SIFS} + \mathrm{Slot\_time} \le \mathrm{PIFS},$$
$$\mathrm{PIFS} + \mathrm{Slot\_time} + \max\{\mathrm{RBT}\} \le \mathrm{DIFS} + \min\{\mathrm{DBT}\}.$$

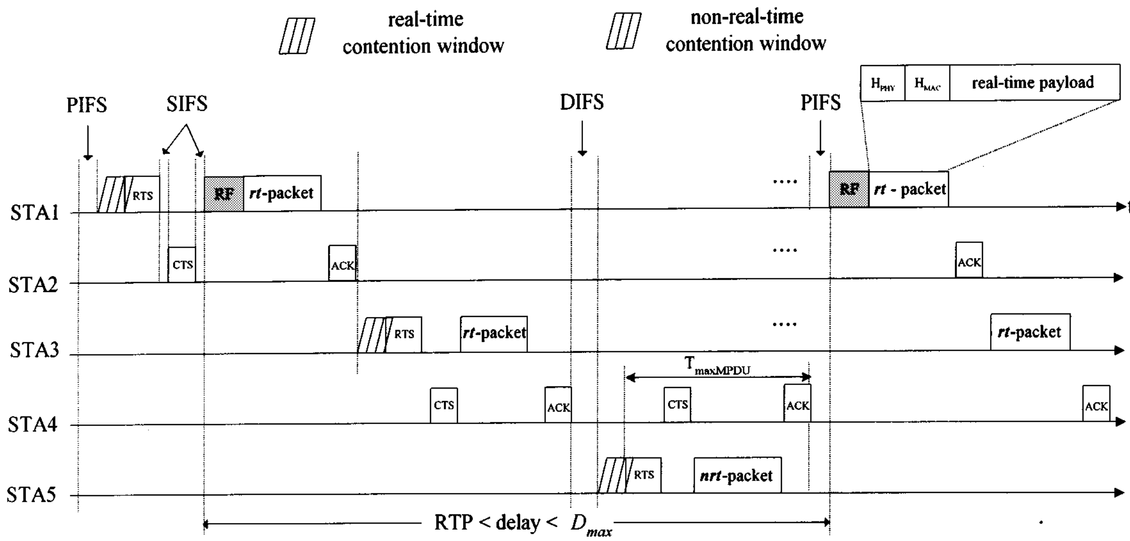


Fig. 2. An example of joining a new rt-station into the *RSVT*.

Thus, c and d of RBT are set as 0 and 3, respectively. We notice that the value d may influence the network throughput. It is clear that a larger value d is, a shorter average contention resolving period for active rt-stations will be obtained. The side effect is that a larger value d will also extend the DIFS period that will decrease the network throughput. Therefore, if we let rt-stations only need to contend channel at the first time access, a smaller value d is acceptable. We also notice that if the traffic load of rt-stations dominates the network throughput, the length of DIFS will not influence the network performance.

The propagation delay can be ignored because the diameter of the basic service area (BSA) of an IBSS in IEEE 802.11 *ad hoc* WLAN is only on the order of 100 feet [12]. That is because an *ad hoc* WLAN is composed solely of stations within mutual communication range of each other and they are able to communicate to each other directly. Moreover, the behavior of RF packet could be considered like that of Beacon in an IBSS. Therefore, the potential hidden node problem in WLAN is not considered in this paper.

Fig. 2 shows an example of how to add a session into a *RSVT* by contending the medium for its reservation. (For simplicity, we omit the timings of issuing Beacons in this example.) We assume there are five stations where STA1 and STA3 want to transmit rt-packets to STA2 and STA4, respectively. Moreover, STA5 has a nrt-packet that is waiting for transmission. In this case, if no RF is received by STA1 and STA3 after listening to the channel for $D_{\max}$, they believe that no rt-station exists. Then, if STA1 and STA3 detect an idle period of PIFS after a detecting period $D_{\max}$, each of them generates a backoff time (RBT) and starts to count down. We assume that $\mathrm{RBT}_{\mathrm{STA1}}$ is smaller than $\mathrm{RBT}_{\mathrm{STA3}}$. When $\mathrm{RBT}_{\mathrm{STA1}}$ counts down to zero, STA1 sends out a RTS and waits for its CTS. If STA1 receives a CTS within SIFS, there is no collision occurred and STA1 adds into the *RSVT* successfully. Because STA1 is the first active station in the *RSVT*, it has the responsibility to transmit a RF before its rt-packet. After STA1 finishes transmitting its first rt-packet, STA3 continues to count down. When $\mathrm{RBT}_{\mathrm{STA3}}$ counter becomes zero, STA3 sends out its RTS and waits for receiving its CTS. When STA3 receives its CTS, it knows that it already reserves the bandwidth and joins into the *RSVT*. When the nrt-station STA5 detects channel being idle for DIFS (it will happen after the rt-traffic contention period), it can try to send out its RTS as soon as its backoff time (DBT) becomes zero.

After a rt-station $\mathrm{STA}_{\mathrm{RT}}$ has joined into the *RSVT* and the passing time from the last access is over the RTP, it will start to monitor the radio channel for its next contention-free access. If the $\mathrm{STA}_{\mathrm{RT}}$ is the CFPG (SID $=1$), it will issue the RF frame and transmit its rt-packet as soon as the channel is detected idle for duration of PIFS. On the other hand, as soon as the other $\mathrm{STA}_{\mathrm{RT}}$ receives the RF, $\mathrm{STA}_{\mathrm{RT}}$ will update its *RSVT* by the broadcast information in the RF and the access instant of each rt-station will be decided. We note that when the channel is idle for a Slot_time, we assume that a rt-station with AID (i.e., SID $=$ AID), which should deliver its packet at this moment, has stopped transmitting. (We assume that the "idle" condition can be distinguished from "interference" case and will be handled in different way.) Due to the characteristic of the multimedia traffic, it is reasonable to release the reserved bandwidth when a station tears down a session. Thus, every following rt-station with larger SID will shift forward its access sequence in the *RSVT*. If the channel is still idle for the next Slot_time, the release phase will be repeated. After each station finishes sending its packet at the current cycle, it still keeps monitoring the channel to check whether any session behind it is being teared down or any new session succeeds to add into the *RSVT*. In the following subsections, we will describe how to dynamically allocate the amount of the reserved bandwidth for each session.

2) Allocation Procedure: The *RSVT* in each rt-station records the sending sequence, the packet length, the traffic type and the MAC address of each active rt-station. The information of *RSVT* can be got and updated by the RF frame and by checking the duration field of MAC header in each MAC PDU (MPDU). The DBASE MAC header format is shown in Fig. 3. In which, the duration field consists of five sub-fields: control field, type field, next degree (ND) field, extension flag (EF) and

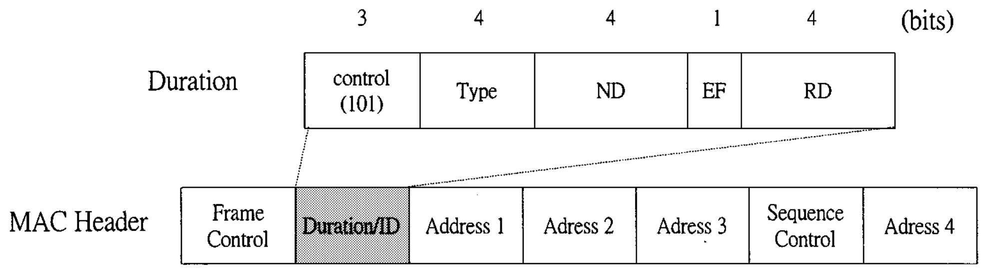


Fig. 3. The duration field format of an IEEE 802.11 MPDU.

raise degree (RD) field. The first three bits in the duration field of a packet is control field. We assume that the control field with "101" indicates that this packet is a rt-packet. The value "101" has been reserved by IEEE 802.11 [12]. The 4-bit type field indicates what type of the rt-packet is (e.g., voice, video, MPEG bit stream, etc.). These traffic types are predefined in system. Each station utilizes the ND field to inform other stations its demanded bandwidth at the next cycle. The request is decided by the amount of buffered packets. Since ND field only occupies 4 bits, the required bandwidth is represented as 16 degrees for each service type. To reduce the overhead of describing the length of required bandwidth, we let $u(i)$ denote the unit bandwidth of the multimedia type i of a session. Then the required bandwidth can be described by identifying how many bandwidth units $u(i)$ a station needs. The $u(i)$ can be obtained by considering two characteristic parameters: $\mathrm{PBR}(i)$ (peak bit rate) and $\mathrm{MBR}(i)$ (minimal bit rate) of multimedia type i. Thus we have

$$u(i) = \left\lceil \frac{\mathrm{PBR(i)} - \mathrm{MBR(i)}}{16} \times \frac{D_{\max}}{\mathrm{UST} \times \mathrm{CDR}} \right\rceil$$

where UST is the unit slot time in system and CDR is the channel data rate. In this equation, we know that a minimal amount of bandwidth is prereserved for each session in each cycle.

Let $MS(i)$ and $ND(i)$, respectively, denote the minimum number of UST needed for an active session in every cycle and the required bandwidth degree [excluding the minimum guaranteed bandwidth $MS(i)$] of a session with service type i at the next cycle. We define

$$MS(i) = \mathrm{MBR}(i) \times \left\lceil \frac{D_{\max}}{\mathrm{UST} \times \mathrm{CDR}} \right\rceil$$

and

$$ND(i) = \left\lceil \frac{QL}{u(i)} \right\rceil, \qquad u(i) > 0$$

where QL denotes the queue length of each rt-station which is measured in UST. We note that in the case of $u(i) = 0$ (i.e., CBR service), we let $ND(i) = 0$. Moreover, the derived ND will be no mare than 16. As long as a rt-station has the right to access the channel, the ND is calculated (according to its current queue length) and broadcast immediately. As a result, all stations will be aware of the reserved bandwidth for each session in every CFP. Now we will describe how to dynamically allocate/share/extend bandwidth for stations with different bandwidth requirements in the *ad hoc* WLAN.

Let $\mathrm{AVD}(i)$ denote the average degree of a multimedia type i. The $\mathrm{AVD}(i)$ can be obtained from the $\mathrm{ABR}(i)$ (average bit rate) of multimedia type i. That is,

$$\mathrm{AVD}(i) = \frac{(\mathrm{ABR}(i) - \mathrm{MBR}(i)) \times D_{\max}}{u(i) \times \mathrm{UST} \times \mathrm{CDR}}.$$

In DBASE, if the demand of each rt-station for bandwidth in the next cycle does not excess its average bandwidth requirement (ABR), the demanded bandwidth will be allocated. Otherwise, only the average bandwidth quota of its multimedia type will be first allocated. Therefore, the maximal bandwidth reserved for all active sessions in every CFP is the sum of AVDs of all active sessions, and simply described as the $\mathrm{CFP}_{\max}$. If the $\mathrm{CFP}_{\max}$ is larger than the actually required bandwidth, there is residual bandwidth that can be shared by the other overloaded rt-stations. In the next subsection, we will describe how residual bandwidth is shared.

Because the interference exists in the wireless environment and the delay bound for the rt-traffics is limited, an efficient *retransmission* scheme must be designed to improve the quality of rt-traffics and the stability of a distributed system. In DBASE protocol, after passing CFP, the CFPG will broadcast the retransmission mapping (RTM) frame. The RTM is a bit mapping to inform all rt-stations which stations can retransmit their packets. For example, RTM "001," means that the rt-station whose $\mathrm{SID} = 3$ can retransmit its rt-packet after RTM frame. The length of the retransmitted packet is limited as the negotiated average packet length (can be derived from AVD). If the retransmission succeeds, each rt-station monitoring the channel will detect the ND field of the retransmitted packet and record this information into its *RSVT*. Otherwise, the ND field for this retransmitted station in the *RSVT* will be set as $\mathrm{AVD}(i)$ if the multimedia type of the retransmitted station is i. During the *retransmission* process, every retransmitted rt-stations need to check whether the time instant will over the boundary of RTP or not. If the finish time will exceed the boundary, the *retransmission* process is terminated immediately.

3) Sharing Procedure: Before reallocating residual bandwidth for the overloaded rt-stations, every overloaded rt-station will first accumulate the spare bandwidth from those sessions whose ND requests are less than AVD. Let CD be the required bandwidth degree of a session in the current cycle, which is copied from the ND of the MPDU transmitted in the previous cycle. For simplicity to demonstrate the following equations, we omit the service type parameter in notations AVD, CD, MS, and u. Let SS denote the number of USTs that can be shared by those rt-stations whose demanded degree excesses its AVD. (The SS can be treated as the residual bandwidth in a contention free cycle.) Therefore, we have

$$SS = \sum_{k=1}^{AC} (\mathrm{AVD}_k - \min(\mathrm{AVD}_k,\, CD_k)) \times u_k.$$

To fairly share the residual bandwidth among over-loaded stations, the proportional approach is used. The actual number of reserved unit slot (RS) for session j in current cycle will be

$$RS_j = \begin{cases} \left\lceil \dfrac{(CD_j - \mathrm{AVD}_j) \times u_j}{\sum\limits_{CD_k > \mathrm{AVD}_k} (CD_k - \mathrm{AVD}_k) \times u_k} \right\rceil \\ \quad \times SS + \mathrm{AVD}_j \times u_j + MS \\ \quad + \dfrac{2 \times \text{SIFS} + T_{ACK}}{\text{UST}}, \\ \qquad\qquad\qquad \text{where } CD_j > \mathrm{AVD}_j \\ CD_j \times u_j + MS + \dfrac{2 \times \text{SIFS} + T_{ACK}}{\text{UST}}, \\ \qquad\qquad\qquad \text{otherwise } CD_j \le \mathrm{AVD}_j \end{cases}$$

where T_{ACK} is the time needed for transmitting the ACK frame. According to the above equation, the packet length of active rt-station in this cycle can be calculated by each station individually. Consequently, the access instant and the length of reserved period for every station can be got easily.

4) Extension Procedure: In the case that the burst traffic arrives just after a rt-station has issued the ND for the next cycle, the delay bound of the excess data may be violated. To solve this critical problem, the 1-bit extension flag (EF) in duration field of each MAC frame is used to indicate whether a rt-station needs more bandwidth than the amount that had been announced at previous cycle. If the EF is set in the duration field of MPDU, the following 4-bit raise degree (RD) will record the renewal demand at the current cycle. But the RD will not be larger than AVD. As long as the EF is set, the shared slot (SS), the CD and the RS must be recalculated by the other rt-stations.

For the sake of providing the right for sessions, whose ND is lower than the negotiation (ND $<$ AVD), to extend their transmission bandwidth, the transmission instances of these sessions are arranged at the beginning of the CFP by scheduler. Based on the access sequence, once they intend to take their bandwidth back, the latter rt-stations (with larger SID) will detect the extention and then each of them will recalculate the currently remaining SS and the new RS allocated in the current cycle. In other words, when any session extends its degree, the overloaded stations will reduce their RSs since SS is decreased. Based on this concept, a rt-station may change its SID in every contention free cycle in DBASE. This *scheduling* scheme is an efficient and flexible method for dealing with the statistical multiplexing for multimedia services over WLAN.

Fig. 4(a) shows an example of reservation and scheduling process. We simply assume that there are three active rt-stations A, B and C. The traffic type of station A is CBR and the traffic types of stations B and C are both VBR. The characteristics of the traffic type CBR and VBR are listed in Table I. In this example, we consider CDR $=$ 11 Mb/s and the $D_{\max}$ for both CBR and VBR services is 25 *ms*. The value of CD in the Fig. 4(a) indicates the allocated bandwidth at this CFP according to the reserved value (ND) in the previous CFP by piggyback. As mentioned above, we schedule the transmitting sequence at the second cycle as "A, C, B" according to values of NDs in the previous cycle. Session C with a ND (level 5) that is lower than its AVD (level 7) will be put in front of any session with a ND higher than the AVD (i.e., session B). Therefore, the sessions with light load could have enough time to take back their bandwidth by setting their EFs. Fig. 4(b) shows that session C extends its transmitting bandwidth by setting EF, but the maximum extended degree is still limited by AVD. We emphasize that session B will reduce its degree from 8 units to 7 units (i.e., AVD) because no more SS can be shared at this cycle. Fig. 4(c) is the case that the CFPG closes its session. Other sessions recognize it by a Slot_time idle after an PIFS. Then the sessions C and B reduce their SIDs and ACs by one and remove session A from their own *RSVT*. Therefore, session C becomes the new CFPG (whose SID $=$ 1) and sends a RF as usual.

In Fig. 5, we show two examples of *retransmission* scheme. The retransmission process can reduce the effects caused by the interference. If the CFPG does not detect the correct frame or the acknowledge frame of any session at the instant the bandwidth reserved for it, the retransmission procedure will reserve an extra bandwidth for this session as long as the retransmission will not cross the boundary of RTP. In Fig. 5(a), the session B fails at its first transmission and then succeeds to send its rt-packet in the retransmission period and the ND of session B will be recorded in the *RSVT*. If the retransmission still fails, the system will only reserve the amount of AVD for session B in the next cycle due to the ND of session B is not being detected by others. The second case is shown in Fig. 5(b).

III. Boundary Conditions and Saturated Connections

In this section, we will discuss the system parameters and analyze the saturated connections of multimedia type i in the proposed DBASE. Due to the characteristic of time-sensitive for multimedia services, we should consider the saturation of sessions of service type i (N_i^{sat}) under the maximum delay bound ($D_{\max}$). The event in which the CFP starts later than the nominal start of repetition period is called stretching. We show the stretching event in Fig. 6. To prevent the transmission of the asynchronous MPDU from stretching the real-time repetition period ($D_{\max}$) even in the worst case (i.e., the stretching packet is the maximum MPDU), the following equation must be satisfied:

$$D_{\max} \ge \text{RTP} + \text{DTP} + \text{T}_{\max \text{MPDU}}$$

where RTP is sum of the rt-packets reserved period (CFP) and the rt-stations access/contention period, and DTP is the preserved data transmission period for nrt-packets if any. Let T_{RF} denote the transmission time of a RF frame. Based on the proposed DBASE, the RTP can be derived by

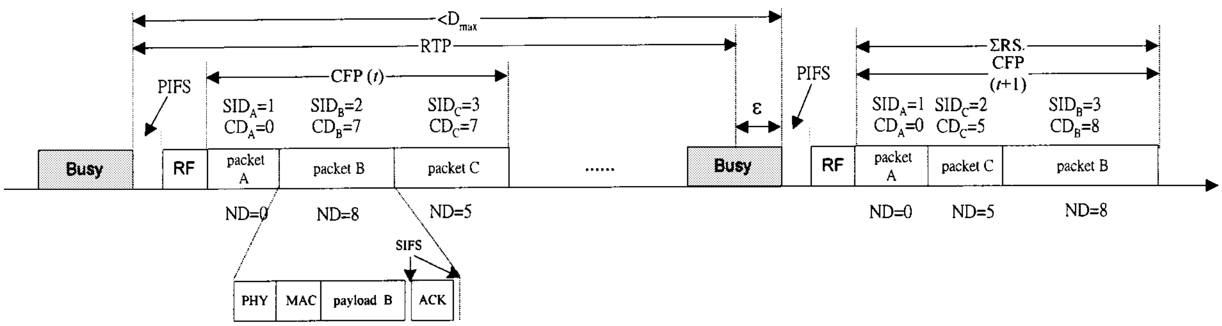
$$\text{RTP} = \text{PIFS} + \text{T}_{\text{RF}} + \sum (N_i \times AB_i)$$

SID	MAC	ND	CD	RD	$Flag_{retrans}$	Type	AVD	MS
1	A	0	0	-	Fault	CBR	0	15
2	B	8	7	-	Fault	VBR	7	28
3	C	5	7	-	Fault	VBR	7	28

CFP cycle (t)

SID	MAC	ND	CD	RD	$Flag_{retrans}$	Type	AVD	MS
1	A	0	0	-	Fault	CBR	0	15
3	B	8	8	-	Fault	VBR	7	28
2	C	5	5	-	Fault	VBR	7	28

CFP cycle (t+1)



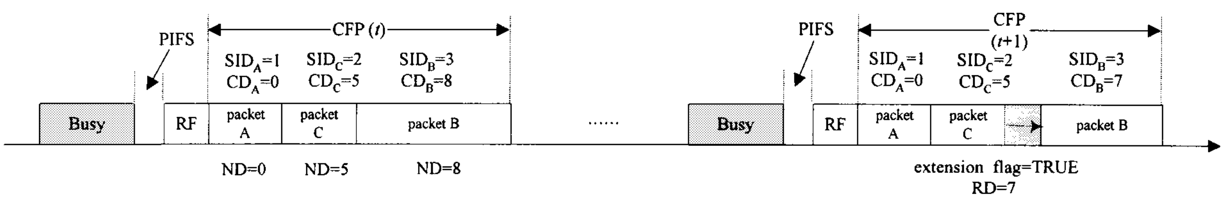
(a)

SID	MAC	ND	CD	RD	$Flag_{retrans}$	Type	AVD	MS
1	A	0	0	-	Fault	CBR	0	15
2	B	8	8	-	Fault	VBR	7	28
3	C	5	5	-	Fault	VBR	7	28

CFP cycle (t)

SID	MAC	ND	CD	RD	$Flag_{retrans}$	Type	AVD	MS
1	A	0	0	-	Fault	CBR	0	15
3	B	9	7	-	Fault	VBR	7	28
2	C	7	5	7	Fault	VBR	7	28

CFP cycle (t+1)



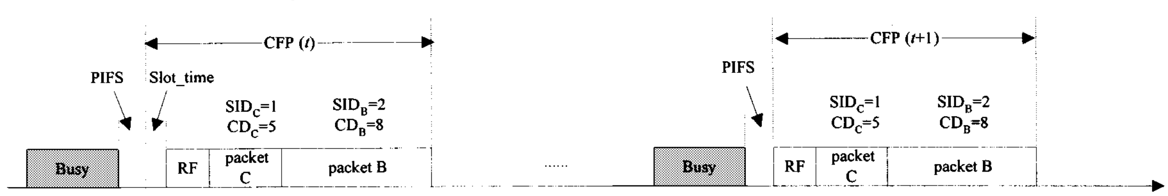
(b)

SID	MAC	ND	CD	RD	$Flag_{retrans}$	Type	AVD	MS
1	A	0	0	-	False	CBR	0	15
3	B	8	8	-	False	VBR	7	28
2	C	5	5	-	False	VBR	7	28

CFP cycle (t)

SID	MAC	ND	CD	RD	$Flag_{retrans}$	Type	AVD	MS
2	B	8	8	-	False	VBR	7	28
1	C	5	5	-	False	VBR	7	28

CFP cycle (t+1)



(c)

Fig. 4. Examples of DBASE protocol.

TABLE I
THE CHARACTERISTICS OF TWO TRAFFIC TYPES FOR SIMULATIONS AND EXAMPLES

Type	PBR	ABR	MBR	MS (in UST)	AVD	u (in UST)
CBR	64 Kbps	64 Kbps	64 Kbps	15	0	0
VBR	420 Kbps	240 Kbps	120 Kbps	28	6	5

where AB_i means the average bandwidth of multimedia type i and N_i is the number of simultaneously active sessions of multimedia type i. Thus, we have

$$AB_i = u(i) \times \mathrm{AVD}_i \times \mathrm{UST}.$$

Now we can easily calculate N_i of DBASE in WLAN by a given maximum delay bound $D_{\max}$ and the numbers of active

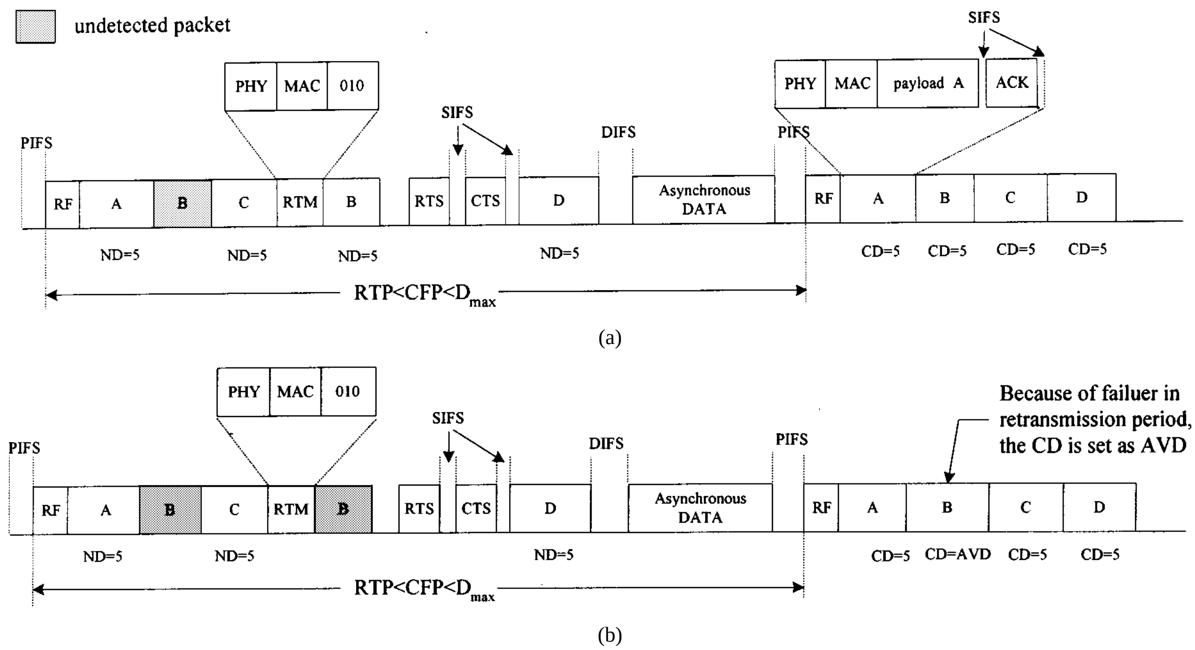


Fig. 5. Examples of retransmission processes.

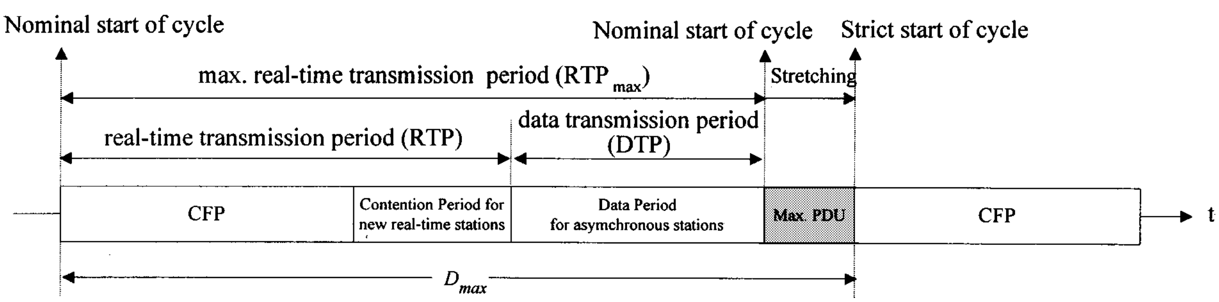


Fig. 6. The stretching situation in the superframe structure.

sessions of multimedia type k (N_k, $i \neq k$), as shown in (1) at the bottom of the page. Thus, we can derive $N_i^{\rm sat}$ of service type i by (2), also shown at the bottom of the page.

IV. Throughput Analysis

In this section we will develop a P-*persistent* model for rt-sessions and a nonpersistent model for nrt-stations to analyze the saturated throughput of DBASE. The repetition period consists of the contention free period of reserved rt-sessions, the contention period of new rt-sessions, and the contention period of nrt-stations. We circumstantiate these three periods in the following three subsections.

A. P-Persistent Model

The P-*persistent* model is developed for the contention period of rt-sessions, as shown in Fig. 7. We consider a fixed number n of contending rt-stations. Let $b^{rt}(t)$ be the stochastic process representing the size of the backoff window for a given rt-session at time t and $s^{rt}(t)$ be the stochastic process representing the backoff stage of the rt-session at time t. The max-

$$N_i \leq \frac{D_{\max} - [\mathrm{PIFS} + \mathrm{T_{RF}} + \sum_{k\neq i}(N_k \times AB_k) + \mathrm{DTP} + \mathrm{T_{max\,MPDU}}]}{AB_i} \tag{1}$$

$$N_i^{\rm sat} = \left\lfloor \frac{D_{\max} - [\mathrm{PIFS} + \mathrm{T_{RF}} + \sum_{k\neq i}(N_k \times AB_k) + \mathrm{DTP} + \mathrm{T_{max\,MPDU}}]}{AB_i} \right\rfloor \tag{2}$$

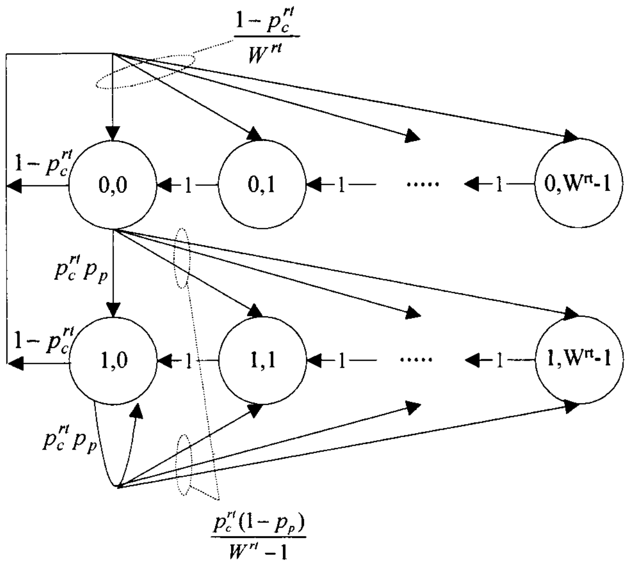


Fig. 7. Markov chain model of P-*persistent* sessions.

imum backoff window size for each backoff stage is a constant value W^{rt}. We assume that the collision probability of rt-sessions p_c^{rt} is independent on the backoff state $s^{rt}(t)$. In this condition, the bi-dimensional process $\{s^{rt}(t), b^{rt}(t)\}$ is a discrete-time Markov chain with the transition probabilities shown as follows:

For simplicity, we let $P\{i_1, k_1 \,|\, i0, k_0\} = P\{s^{rt}(t+1) = i_1, b^{rt}(t+1) = k_1 \,|\, s^{rt}(t) = i_0, b^{rt}(t) = k_0\}$. To describe the decrement of the backoff time counter, we have

$$P\{i, k \,|\, i, k+1\} = 1, \qquad 0 \le k \le W^{rt} - 2, 0 \le i \le 1.$$

In saturation condition, to describe a new session starts with a backoff stage 0, we have

$$P\{0, k \,|\, i, 0\} = \frac{(1 - p_c^{rt})}{W^{rt}}, \qquad 0 \le k \le W^{rt} - 1, 0 \le i \le 1.$$

After an unsuccessful transmission in backoff stage i, the transition probabilities that the size of backoff window is zero are described as

$$P\{1, 0 \,|\, i, 0\} = p_c^{rt} p_p, \qquad 0 \le i \le 1;$$

where p_p is the persistent probability.

After an unsuccessful transmission in backoff stage i, the transition probabilities that the size of backoff window is uniformly chosen in the range (1, $W^{rt} - 1$) are described as

$$P\{1, k \,|\, i, 0\} = \frac{p_c^{rt}(1 - p_p)}{W^{rt} - 1}, \quad 1 \le k \le W^{rt} - 1, 0 \le i \le 1.$$

Let $b_{i,k}^{rt}$ mean the stationary distribution of the Markov chain and we have

$$b_{i,k}^{rt} = \lim_{t \to \infty} P\left\{s^{rt}(t) = i, b^{rt}(t) = k\right\}, \qquad 0 \le k \le W^{rt} - 1, 0 \le i \le 1.$$

By investigating the chain regularities, the following relations are derived:

$$b_{0,k}^{rt} = \frac{W^{rt} - k}{W^{rt}} b_{0,0}^{rt} \qquad 0 \le k \le W^{rt} - 1,$$

$$b_{1,k}^{rt} = \frac{(W^{rt} - k)(1 - p_p)}{W^{rt} - 1} b_{1,0}^{rt} \qquad 0 \le k \le W^{rt} - 1,$$

and

$$b_{1,0}^{rt} = \frac{p_c^{rt}}{1 - p_c^{rt}} b_{0,0}^{rt}.$$

We can obtain $b_{0,0}^{rt}$ by imposing the normalization condition,

$$\begin{aligned} 1 &= \sum_{i=0}^{1} \sum_{k=0}^{W^{rt}-1} b_{i,k}^{rt} \\ &= b_{0,0}^{rt} \sum_{k=0}^{W^{rt}-1} \frac{W^{rt} - k}{W^{rt}} \\ &\quad + b_{1,0}^{rt} \sum_{k=1}^{W^{rt}-1} \frac{(W^{rt} - k)(1 - p_p)}{W^{rt} - 1} + b_{1,0}^{rt} \\ &= \left[\frac{W^{rt} + 1}{2} + \frac{W^{rt}(1 - p_p) + 2}{2} \frac{p_c^{rt}}{1 - p_c^{rt}}\right] b_{0,0}^{rt}, \end{aligned}$$

and thus we obtain

$$b_{0,0}^{rt} = \frac{2(1 - p_c^{rt})}{W^{rt}(1 - p_p p_c^{rt}) + 1 + p_c^{rt}}.$$

The rt-station will transmit its packet only when the size of backoff window is decreased to zero. Let the probability that a station transmits in a slot time be represented by τ^{rt}, and we have

$$\tau^{rt} = \sum_{i=0}^{1} b_{i,0}^{rt} = \frac{b_{0,0}^{rt}}{1 - p_c^{rt}} = \frac{2}{W^{rt}(1 - p_p p_c^{rt}) + 1 + p_c^{rt}}.$$

If more than one station of N^{rt} rt-stations transmit their packets simultaneously in a slot time, collision will occur. Thus the collision probability can be denoted by the following equation:

$$p_c^{rt} = 1 - (1 - \tau^{rt})^{N^{rt}-1}.$$

By solving the previous two equations, the values τ^{rt} and p_c^{rt} can be found. Once τ^{rt} is known, the probability P_T^{rt} that at least one rt-station transmits its packet and the probability P_S^{rt} that a packet is transmitted successfully can be obtained by the following two equations:

$$P_T^{rt} = 1 - (1 - \tau^{rt})^{N^{rt}},$$

and

$$\begin{aligned} P_S^{rt} &= \frac{N^{rt} \tau^{rt} (1 - \tau^{rt})^{N^{rt}-1}}{P_T^{rt}} \\ &= \frac{N^{rt} \tau^{rt} (1 - \tau^{rt})^{N^{rt}-1}}{1 - (1 - \tau^{rt})^{N^{rt}}}. \end{aligned}$$

Additionally, the mean of consecutive idle slots between two consecutive transmissions of rt-packet $E[\Psi^{rt}]$ can be derived by

$$E[\Psi^{rt}] = \frac{1}{P_T^{rt}} - 1.$$

According to our designed protocol, if the number of consecutive idle slots is larger than that of DIFS plus minimum DBT, the period of nrt-sessions will start. Thus the maximum idle slots in the period of rt-sessions I^r will be

$$I^r = \min\{(\text{DIFS} + \min\{\text{DBT}\}),\, E[\Psi^v]\}$$

where $\min\{A\}$ returns the minimum value in set A.

B. Reservation Model

The reservation model is developed to analyze the average number of active rt-sessions in the contention free period. The state of the Markov chain in Fig. 8 is defined to be the number of rt-sessions in the reservation list, where the value m is the maximum number of active rt-sessions [19]. The value m can be decided by N_i^{sat} as analyzed in Section III. The probability that a rt-session completes its whole transmission during a Slot_time is given by $\mu = 1/(L \times \text{NS}_{\text{slot}})$ where L is the average number of frames (each frame is transmitted in each CFP) of rt-session, and NS_{slot} is the number of Slot_times of $D_{\max}$. Let $P_{S(k)}^v$ be the arrival probability of a member in the reservation list during a unit slot, when there are k rt-sessions already in the reservation list. Thus $P_{S(k)}^{rt}$ can be expressed as

$$P_{S(k)}^{rt} = \begin{cases} \dfrac{N^{rt}\tau^{rt}(1-\tau^{rt})^{N^{rt}-1}}{1-(1-\tau^{rt})^{N^{rt}}} & k < m; \\ 0 & k = m. \end{cases}$$

We note that $P_{S(i)}^v = P_{S(j)}^v, \forall 1 \le i, j \le m$. Let P_S^v be the probability that a RTS is transmitted successfully in the real-time contention period. Thus, we have $P_{S(i)}^v = P_S^v, \forall 1 \le i \le m$. We also assume that the members can both enter and leave states m or state 0 in the same slot. Let X_r be the state at slot r and then the transition probability $a_{i,j}$ from state i to state j is defined as

$$a_{i,j} = \lim_{r\to\infty} P(X_{r+1} = j | X_r = i), \quad \text{where } i, j \in [0, m].$$

Thus the one-step transition probabilities from state n to $n-l$ are

$$\begin{aligned} a_{n,n-l} = {} & P_S^v \times \mu^{l+1} \times (1-\mu)^{n-l} + (1-P_S^v) \\ & \times \mu^l \times (1-\mu)^{n-l}, \qquad 0 \le l \le n \le m. \end{aligned}$$

Moreover, the transition probability from state n to state $n+1$ is

$$a_{n,n+1} = P_S^v \times (1-\mu)^n, \qquad 0 \le n < m.$$

We assume that in equilibrium the flow of probability flux across a vertical boundary between state k and $k+1$ balances

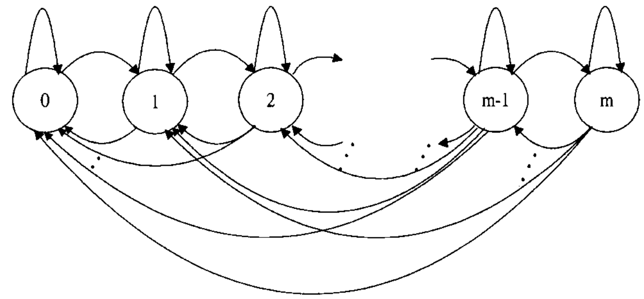


Fig. 8. Markov chain defining the average number of rt-sessions in the reservation list.

in both directions. For the state transition diagram of Fig. 8, it can be expressed as

$$a_{k,k+1} \times Q_k = \sum_{i=k+1}^{m+k} \sum_{j=0}^{k} a_{i,j} \times Q_i$$

where Q_k means the equilibrium state probability of there being k rt-sessions in the reservation list. Reducing the equation leads to

$$Q_k = \frac{1}{a_{k,k+1}} \sum_{i=k+1}^{k+m} \sum_{j=0}^{k} a_{i,j} \times Q_i.$$

According to the normalization regularities, the sum of all of the value Q_k (where $0 \le k \le m$) is equal to 1. That is $\sum_{k=0}^{m} Q_k = 1$. By the previous two equations, the value Q_k can be derived. Let ξ denote the average number of active rt-sessions in the contention free period. We have

$$\xi = \sum_{i=1}^{m} (i \times Q_i).$$

In order to analyze the time length of a contention period spent on arrival rt-sessions contending, we define two parameters T_S^{rt} and T_C^{rt}. The T_S^{rt} is the average time that the channel is sensed busy because of a successful transmission, and the T_C^{rt} denotes the average time that the channel is sensed busy by a station during contention period. Since the contentions of rt-sessions follow RTS/CTS/packet/ACK handshakings, if the propagation delay is ignored, we have

$$\begin{aligned} T_S^{rt} &= \text{T}_{\text{RTS}} + \text{SIFS} + \text{T}_{\text{CTS}} + \text{SIFS} + \text{T}_{\text{H}} \\ &\quad + E[T_{\text{payload}}^{rt}] + \text{SIFS} + \text{T}_{\text{ACK}} \\ T_C^{rt} &= \text{T}_{\text{RTS}}. \end{aligned}$$

where $E[T_{\text{payload}}^{rt}]$ is the average time length used to transmit a rt-packet, $\text{T}_{\text{RTS}}(\text{T}_{\text{CTS}})$ is the time spent on transmitting RTS(CTS) packet, and T_{H} is the time needed to transmit both MAC and PHY headers. The propagation delay is also ignored in above equations. Therefore, we can obtain the length of the period used by rt-stations (RP)

$$\begin{aligned} \text{RP} = {} & \text{PIFS} + \text{T}_{\text{RF}} + \xi \\ & \times (\text{T}_{\text{H}} + E[T_{\text{payload}}^{rt}] + \text{SIFS} + \text{T}_{\text{ACK}} + \text{SIFS}) \\ & + I^r + P_S^{rt} \times T_S^{rt} + (1 - P_S^{rt}) \times T_C^{rt}. \end{aligned}$$

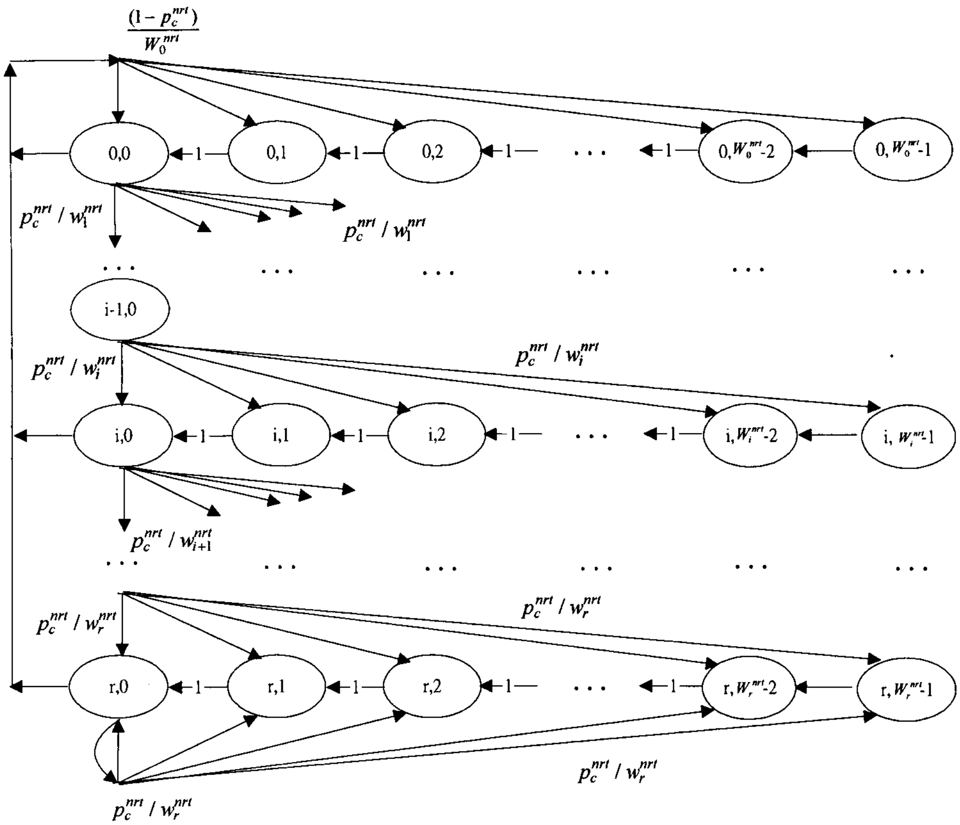


Fig. 9. Markov chain model for nonpersistent procedure of nrt-sessions.

C. Nonpersistent Model

The nonpersistent model shown in Fig. 9 is developed to analyze the contention period of nrt-stations. A Markov chain model has been proposed in [12] to analyze the saturation throughput of IEEE 802.11 that only employs the DCF. Because the nrt-stations only operates with the DCF, we can obtain the τ^{nrt}, which is the collision probability of nrt-stations in a Slot_time. Let W^{nrt} be the basic size of the backoff window and p_c^{nrt} is the probability that a transmitted nrt-packet collides with others in a time slot. That is, τ^{nrt} is

$$\tau^{nrt} = \sum_{i=0}^{1} b_{i,0}^{nrt} = \frac{b_{0,0}^{nrt}}{1-p_c^{nrt}} = \frac{2(1-2p_c^{nrt})}{(1-2p_c^{nrt})(W^{nrt}+1)+p_c^{nrt}W^{nrt}(1-(2p_c^{nrt})^r)}$$

where r is the maximum number of backoff stage. Recall that $b_{i,k}^{nrt}$ means the stationary distribution of the Markov chain.

We assume that $W_i^{nrt} = 2^i \times W^{nrt}$ where W_i^{nrt} is the backoff window size of backoff stage i and N^{nrt} is the number of active nrt-stations. Obviously, we have

$$p_c^{nrt} = 1-(1-\tau^{nrt})^{N^{nrt}-1}.$$

Similar to the analysis of rt-sessions, we can derive the probability P_T^{nrt} that at least one nrt-station transmits packet and the probability P_S^{nrt} that a nrt-station transmits its packet successfully. Thus we have

$$P_T^{nrt} = 1-(1-\tau^{nrt})^{N^{nrt}}$$

and

$$P_S^{nrt} = \frac{N^{nrt}\tau^{nrt}(1-\tau^{nrt})^{N^{nrt}-1}}{1-(1-\tau^{nrt})^{N^{nrt}}}.$$

The mean idle slots between two consecutive transmissions of nrt-packet $E[\Psi^{nrt}]$ can be derived by P_T^{nrt} as deriving $E[\Psi^{rt}]$. That is, $E[\Psi^{nrt}] = (1/P_T^{nrt}) - 1$. Furthermore, let T_S^{nrt} and T_C^{nrt}, respectively, denote the average time of a successful transmission and the average time of an unsuccessful transmission sensed by a nrt-station. Since we consider that the contentions of nrt-stations also follow RTS/CTS/packet/ACK handshakings, if the propagation delay is ignored, we have

$$\begin{aligned} T_S^{nrt} &= \text{DIFS} + \text{T}_{\text{RTS}} + \text{SIFS} + \text{T}_{\text{CTS}} + \text{SIFS} + \text{T}_{\text{H}} \\ &\quad + E[T_{\text{payload}}^{nrt}] + \text{SIFS} + \text{T}_{\text{ACK}} \\ T_C^{nrt} &= \text{DIFS} + \text{T}_{\text{RTS}} \end{aligned}$$

where $E[T_{\text{payload}}^{nrt}]$ is the average time used to transmit the asynchronous packet of the nrt-station and T_{ACK} is the time used to transmit the ACK packet. Because the cycle is limited by the maximum delay bound ($D_{\max}$), the length of asynchronous data transmitting period (DTP) will also be limited. That is

$$\begin{aligned} \text{DTP} = \min\{&(D_{\max} - \text{RTP}),\ (E[\Psi^{nrt}] + P_S^{nrt} \times T_S^{nrt} \\ &+ (1-P_S^{nrt}) \times T_C^{nrt})\}. \end{aligned}$$

TABLE II
NUMERICAL VALUES FOR THE CBR MODEL

Parameter	Value
Conversation Length	180 sec
Principle Talkspurt	1.00 sec
Principle Silent Gap	1.35 sec
Data Bit Rate (CBR)	64 Kbps
Maximum Packet Delay	25 ms

Finally, we define the normalized saturation throughput S as the fraction of the time that the channel is sensed busy by the successful transmission of the payloads, and it is

$$S = \frac{E\left[\begin{array}{c}\text{time used for successful trans.}\\ \text{in repetition period}\end{array}\right]}{E[\text{length of a new repetition period}]}.$$

According to the previous analyses, the saturation throughput S can be expressed as

$$S = \frac{\begin{array}{c}\xi \cdot E[T^{rt}_{\text{payload}}] + P^{rt}_S \cdot E[T^{rt}_{\text{payload}}]\\ + P^{nrt}_S \cdot E[T^{nrt}_{\text{payload}}]\end{array}}{\text{RTP} + \text{DTP}}.$$

We note that in this analysis, the considered overheads include all control packets, collisions, idle time, and the headers of packet, therefore, S is a precise measurement.

V. PERFORMANCE EVALUATION

The performance of the proposed protocol is evaluated by simulations. In this section, the traffic models and the performance measurements are defined. In order to investigate the multimedia services over WLAN, we consider the WLAN with 11 Mb/s, which WLAN adapter has been announced recently. Each simulation run sustains 3×10^6 Slot_times.

A. Traffic Models

In order to evaluate the DBASE performance, three different traffic models are considered:

Voice Traffic Model (CBR): The voice traffic is usually considered as a service with CBR traffic. In the simulations, the voice traffic is modeled as a two-state Markov process with talkspurt and silent-gap states. Each voice source is assumed to equip a slow speech activity detector (SAD) [13], [14]. In the talkspurt state, we consider that the voice source generates a continuous bit-stream; in the silent state, there is no packet to be generated. The duration of talkspurt and silent-gap both follow exponential distribution with the mean duration equal to 1 and 1.35 s, respectively. A voice packet is assumed to be dropped if it suffers a delay longer than $D_{\max}$(=25 ms). The parameters of voice model are summarized in Table II.

Video Traffic Model (VBR): This model is a multiple-state model (shown in Fig. 10) where a state generates a continuous bit stream for a certain holding duration [15], [16]. The bit rate values of different states are obtained from a truncated exponential distribution with a minimum and a maximum bit rate values. The holding times of the states are assumed to be statistically independent and exponential distributed. We assume that each state has the same mean holding time. In the simulations, the generated VBRs are quantized as 16 levels. Table III summarizes the numerical values used for the VBR model.

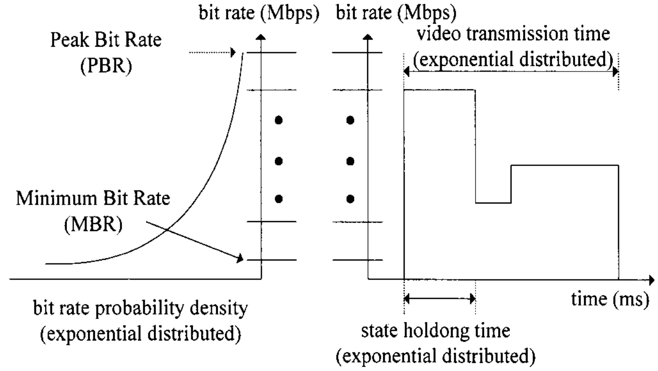


Fig. 10. The source model of VBR.

TABLE III
NUMERICAL VALUES FOR THE VBR MODEL

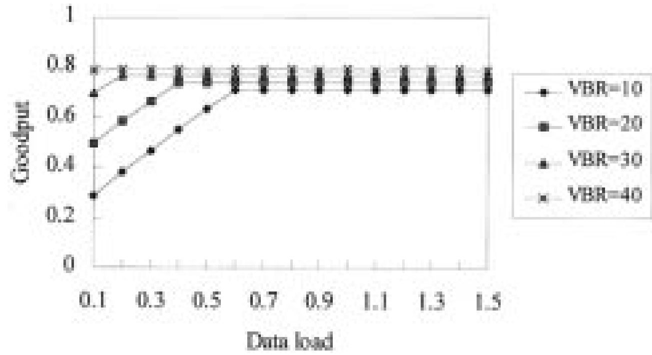
Parameter	Value
Peak Bit Rate (PBR)	420 Kbps
Minimum Bit Rate (MBR)	120 Kbps
Average Bit Rate (ABR)	240 Kbps
Mean State Holding Time	160 msec
Mean Video Call Length	180 sec
Maximum Packet Delay	75 ms

TABLE IV
SYSTEM PARAMETERS OF SIMULATIONS

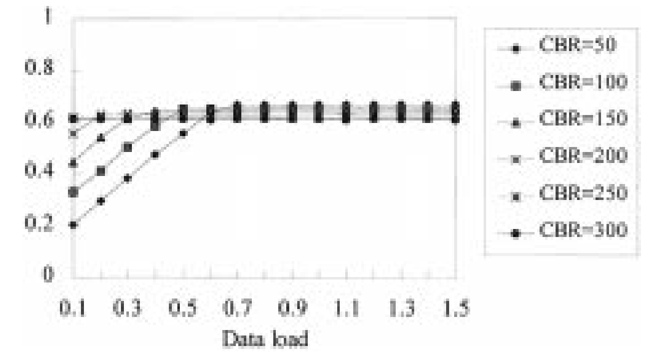
Parameter	Symbol	Value
Channel bit rate	CDR	11 Mbps
Maximum delay	D_{max}	25 ms
Slot time	Slot_time	20 μs
Unit slot time	UST	10 μs
Short interframe space	SIFS	10 μs
Priority interframe space	PIFS	30 μs
DCF interframe space	DIFS	110 μs
MAC header	H_{MAC}	272 bits
PHY header	H_{PHY}	128 bits
The packet length of RTS	$T_{RTS}\times$CDR	288 bits
The packet length of CTS	$T_{CTS}\times$CDR	240 bits
The packet length of ACK	$T_{ACK}\times$CDR	240 bits
Persistent probability	P_p	0.8

Data Traffic Model: We assume that data packets arrive at each station following the Poisson process with the mean value $\lambda = (\lambda^{nrt}/N^{nrt})$, where λ^{nrt} is the total packet arrival rate of asynchronous traffic. The total data load ρ can be estimated as $\rho = \lambda^{nrt} \times E[T^{nrt}_{\text{payload}}]$. In simulations, the $E[T^{nrt}_{\text{payload}}]$ is set as 744 μs (=8184 bits/CDR = 8184 bits/11 Mb/s), λ is set as 0.1, buffer size is fixed at 100 packets and the length of asynchronous packet is fixed at 8184 bits (=1023 bytes).

According to the traffic models defined above, the traffic parameters of DBASE for DSSS PHY are summarized in Table IV.

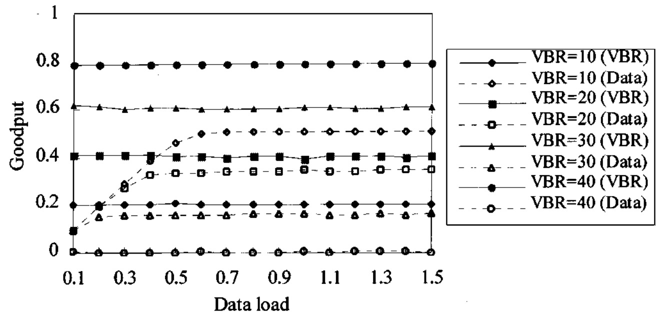


(a)

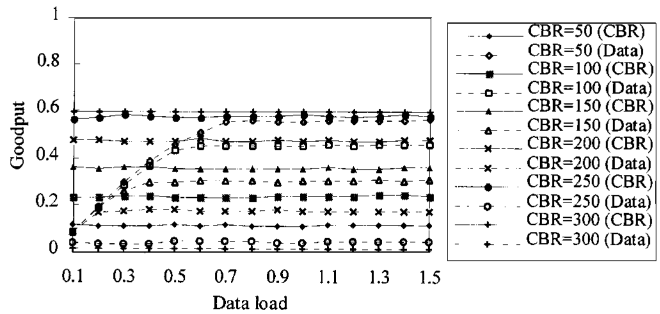


(b)

Fig. 11. The derived Goodputs by DBASE under different numbers of VBR and CBR rt-sessions.



(a)



(b)

Fig. 12. The derived pure Goodputs of VRB/CBR rt-sessions and nrt-sessions by DBASE under different numbers of VBR and CBR rt-sessions.

B. Performance Measurements

The performance measurements considered in our simulations are defined as follows:

- **Goodput:** the goodput is defined as the percentage of the time used by both rt- and nrt-stations to successfully transmit their pure payload data. The control and management signals, the idle time, the collided packets, the error packet caused by the interference and the header bits are excluded from the goodput.
- **Packet delay dropped probability (PDDP):** the packet delay dropped probability is defined as the fraction of discarded rt-packets caused by violating the delay bound.

C. Simulation Results

To observe the performance of DBASE, we consider the goodputs and PDDPs under different numbers of rt-stations and different traffic service types individually. Figs. 11 and 12 plot the network goodputs and PDDPs of CBR and VBR traffic, respectively. The number of nrt-stations N^{nrt} is fixed at 10 and the asynchronous data load (ρ) increases from 0.1 to 1.5 in a step of 0.1. The number of VBR sessions ranges from 10 to 40, and the number of CBR sessions ranges from 50 to 300.

Fig. 11 shows the relations of the goodput and the data load under different number of rt-stations. In Fig. 11, we can find that the goodput can be up to about 80% for VBR traffics and 67% for CBR traffics. We know that the overhead caused by header of packet will be quite obvious when the average payload size of packet is small. Therefore, the saturated goodput of VBR is higher than that of CBR since the packet length of VBR (average payload length is 28 USTs as shown in Table I) is longer than that of CBR (average payload length is 15 USTs). Because the number of data stations is fixed at 10, the increasing data load will only increase the queue length of the data buffer of nrt-station but not the collision probability when the traffic load is saturated. Accordingly, the goodput will not decrease by heavy nrt-traffic load after saturation. In Fig. 11(b), because the payload length of nrt-stations is much larger than that of CBR sessions, curves with different number of CBR sessions will cross each other. While the number of CBR stations decreases, the number of nrt-packets transmitted successfully will become more. This is why the saturated goodput with 50 CBR sessions is slightly higher than others when the asynchronous data load is heavy.

Fig. 12 shows the pure goodputs of rt-stations and data stations. From Fig. 12, we can find that the goodput for VBR stations and CBR stations are not affected by nrt-traffic at all. This implies that rt-stations always have a higher priority to access channel bandwidth than asynchronous data. Moreover, when the total traffic load of rt-stations becomes heavy, the network bandwidth is almost occupied and shared by them.

Fig. 13 plots the PDDPs of CBR and VBR traffics, respectively. According to Section III, we can estimate the saturated capacity of the system by those simulation parameters. If the traffic type is CBR, the saturated capacity is about 112 CBR rt-sessions. Similarly, the saturated capacity for the VBR traffic model is about 39 VBR rt-sessions. In Fig. 13(a), we can see that no packet loss will occur in the cases of less than 40 VBR sessions. Since a number of 40 VBR sessions is just a little over the saturated number of active stations, the PDDP in this case is obvious higher than the other cases. Since DBASE is a reservation-based protocol, packets will be dropped only when the rt-sessions are still in the contention procedure. Therefore, even

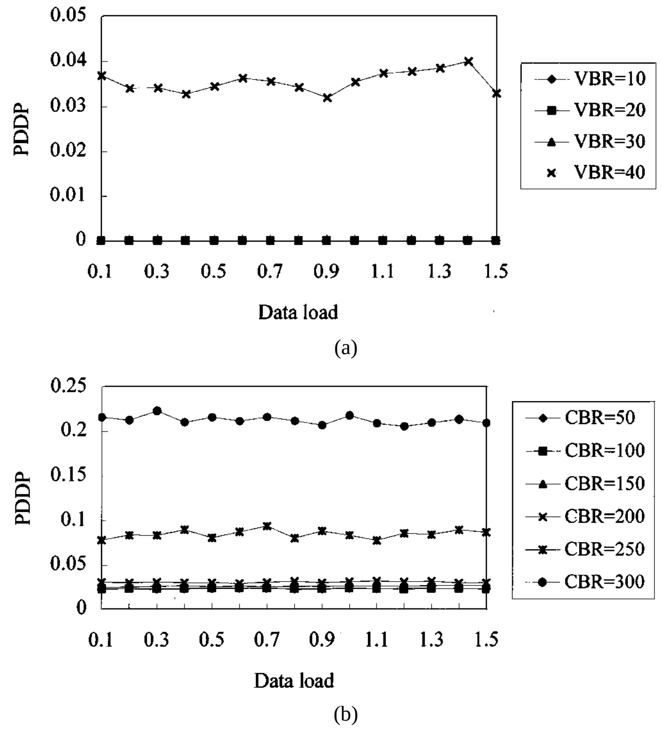


Fig. 13. The derived PDDPs by DBASE under different numbers of VBR and CBR rt-sessions.

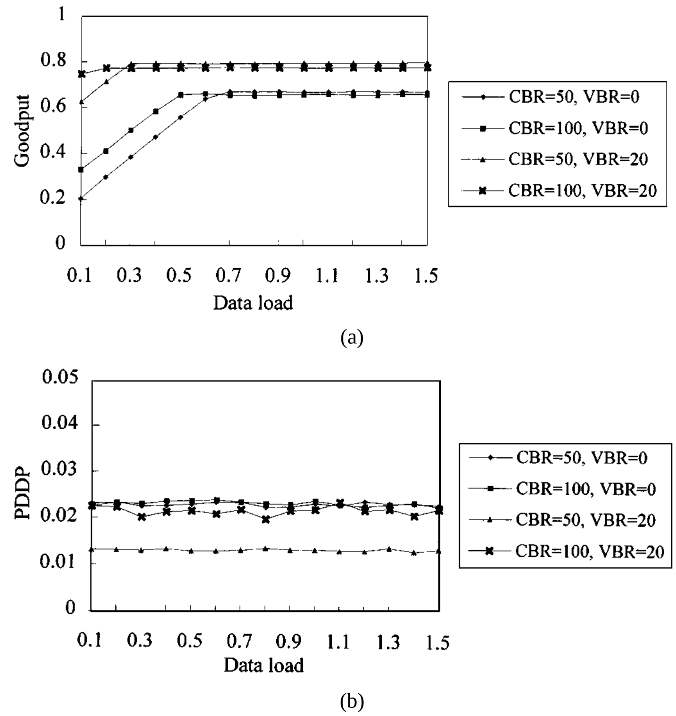


Fig. 14. The derived Goodputs and PDDPs by DBASE under different numbers of VBR and CBR rt-sessions.

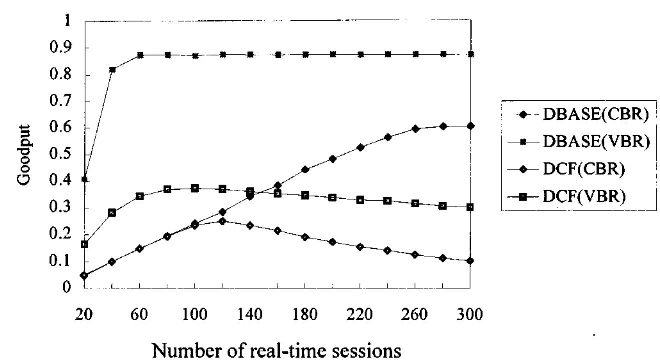


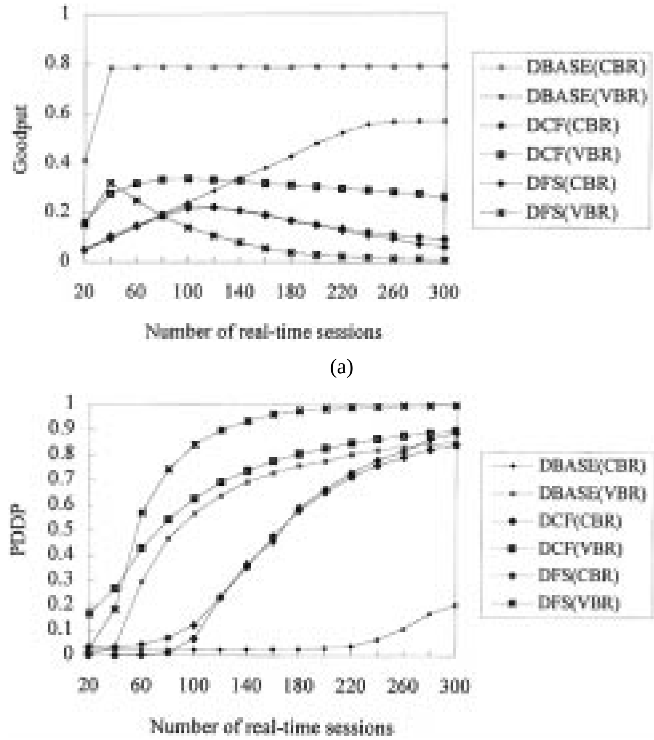
Fig. 15. Comparisons of the Goodputs derived by DBASE and DCF under different traffic types in the clear radio environment, when $N^{nrt} = 10$.

though the number of VBR rt-sessions is over the saturated capacity (39), the PDDP will not increase rapidly as increasing of data load. In Fig. 13(b), we find that the PDDP is not over 3% even when the number of CBR sessions is up to 200, which is far beyond the saturated active sessions of CBR traffics. This is because the ON–OFF model is used in the CBR traffic model. In papers [17], [18], authors concluded that the packetized voice communications can tolerate only a small amount (1%∼3%) of dropped packets before suffering a large quality degradation. Based on these criterias, in Fig. 13(b), the maximum number of simultaneously active CBR stations can be up to 200. We also find that the asynchronous data load does not affect the PDDP because the rt-traffic has a higher priority than the nrt-traffic. The low PDDP of DBASE implies that DBASE can perform a high quality of service even when the traffic load is heavy.

Fig. 14 illustrates the network goodputs and PDDPs derived by DBASE protocol under mixed rt-traffic load in WLAN. We emphasize that when both CBR and VBR rt-sessions are active at the same time, the $D_{\max}$ will be set as the smallest one (=25 ms). From Fig. 14(a), we can see that the derived goodputs are similar as that of Fig. 11 in all cases. Fig. 14(b) shows that the PDDP of rt-sessions is still bounded within 3% even when the number of simultaneously active CBR stations and VBR stations are 100 and 20, respectively. This implies that the DBASE protocol can also support heterogeneous rt-traffic situations in WLAN.

To compare our proposed protocol DBASE with the IEEE 802.11 MAC protocol based on pure DCF, the following simulations are made. For simplicity, the conventional protocol based on IEEE 802.11 DCF is named as DCF in short. In Fig. 15, we assume that the number of nrt-stations is still fixed at 10 ($N^{nrt} = 10$) and the λ of each station is 0.1 (heave network load). The curves DCF(CBR) and DBASE(CBR) indicate the performance of DCF and DBASE with the CBR rt-traffic model, respectively, and so as the DCF(VBR) and DBASE(VBR) with VBR rt-traffic model. For the CBR model, when the active number of rt-sessions is small, the DBASE performs similar to DCF. However, as the number of CBR sessions is larger than 100, the DBASE performs much better than DCF. This is because DBASE will reserve the bandwidth and the increasing contentions caused by the increasing number of rt-sessions will not waste the channel resource. For the VBR model, the DBASE also performs much better than DCF because of the same reasons mentioned above. We emphasize that the derived goodput of DBASE(VBR) can up to almost 90%. This implies that almost network bandwidth is fully utilized by our DBASE protocol.

To compare our proposed protocol DBASE with distributed fair scheduling (DFS) protocol and DCF, the following simulations are made. In Fig. 16, we assume that the number of



(a)

(b)

Fig. 16. Comparisons of Goodputs and PDDPs of DBASE and DCF in the interfering environment when PER $= 0.1$ and $N^{nrt} = 0$.

nrt-stations is zero to investigate how these protocols deal with rt-packets. In DFS protocol, the Scaling_Factor is set as 0.02 and the Collision_Window is 4 slots. We also let all rt-sessions have identical weight ($1/N^{rt}$). All packets on CBR (VBR) rt-session contains 200 ($\approx$800) bytes. (CBR: 64 kb/s $\times$ $D_{\max} =$ 64 kb/s $\times$ 25 ms $=$ 1600 bits/packet; VBR: 240 kb/s $\times$ $D_{\max} =$ 240 kb/s $\times$ 25 ms $=$ 6000 bits/packet) The first backoff windows for rt-packet is equal to Scaling_Factor $\times$ packet_length/weight (that is, $4 \times N^{rt}$ in CBR model and $15 \times N^{rt}$ in VBR model). For example, if $N^{rt} = 20$, the first backoff window for CBR (VBR) rt-packet are 80 (300). Once collision occurs, a relatively smaller backoff window (i.e., Collision_Window) is used to inhibit the bandwidth wastage. The range for backoff window grows exponentially with the number of consecutive collisions. To show the effect of the interference in the worse wireless environment on the DBASE, DCF and DFS, we set the packet error rate (PER) as 0.1.

In Fig. 16(a), the curves show that the goodput of DBASE is obviously higher than that of DCF and DFS under the same traffic type. We know that a higher PER will result in a lower goodput. However, the goodput of DBASE(VBR) can still hold on 78% even though the PER is up to 0.1 but the goodput of DCF(VBR) decreases to 30% only. We notice that DFS performs slightly better than DCF when the active number of VBR rt-sessions is small. The reason is that the potential collisions are solved by enlarging the first backoff window in DFS. However, a longer backoff interval may make rt-stations drop more rt-packets due to delay expiry as shown in Fig. 16(b). As the active number of rt-sessions becomes large, the goodput and PDDP achieved by DFS become the lowest and worst one, respectively. This is because the DFS scheme tends to choose greater backoff intervals than DCF, resulting in higher overhead (penality) for DFS. Furthermore, since we assume the weight of each flow is constant and is inverse proportional with the number of active rt-sessions, the large backoff window will waste considerable bandwidth especially when the data arrival rate is not constant and the number of active rt-sessions is large. This is the major shortcoming of DFS protocol.

In Fig. 16(b), we also find that the PDDP of DCF is much higher than that of DBASE under the same traffic type. Fig. 16(b) illustrates that the PDDP of DBASE(CBR) will begin to increase when the number of CBR sessions is larger than 220, which is the saturated capacity under the consideration of ON–OFF model and PER $= 0.1$. However, the PDDPs of DCF(CBR) and DFS(CBR) will increase sharply as long as the number of CBR sessions is larger than 100. Similarly, the PDDP of DBASE(VBR) starts increases at about 40 VBR sessions and the PDDP of DCF(VBR) is up to about 20% while the number of VBR sessions is only 20. This is because the PER and contentions make the packet delay in DCF increase extremely. Since the bandwidth is reserved for the rt-packets to retransmit in DBASE, the packet error rate does not influence the PDDP of DBASE seriously when the number of rt-sessions is small. Although DBASE has the *retransmission* scheme to reduce PDDP of the reserved rt- sessions, the new rt-sessions using CSMA/CA to reserve bandwidth still has the chance to make the PDDP increase as the rt-traffic load increases. This is why the PDDP of DBASE(VBR) is only slightly better than that of DCF(VBR) when the traffic load is heavy and overloaded. Nevertheless, the goodput of DBASE(VBR) is still maintaining about 80% when the number of active VBR stations is more than 39.

D. Analysis Results

In this subsection, we will show the saturation throughput of DBASE by simulation and analysis. In order to simplify the analysis and simulation scenarios, the following assumptions are used.

1) The wireless channel is assumed as error-free and a transmitted packet will be corrupted only when collision occurs.
2) The propagation delay and state transition time are ignored.
3) There are no hidden terminals in the wireless LAN. All stations in a WLAN can hear each other.

Some parameters are decided for analysis. For data traffic, the number of asynchronous data sessions N^{nrt} is increasing from 10 to 60. For rt-traffic, we assume $L = 100$ packets, $W^{nrt} = 4$, and N^{rt} is the number of real-time sessions. To show the saturation throughput of DBASE under different sizes of payloads, we consider three scenarios with CDR $=$ 11 Mb/s. In scenario I, we assume $E[T^{nrt}_{\text{payload}}] =$ 8184 bits/CDR $=$ 744 μs, $E[T^{rt}_{\text{payload}}] =$ 1600 bits/CDR $=$ 145.45 μs ($E[T^{rt}_{\text{payload}}] \times$ CDR$/D_{\max} \cong$ 64 kb/s), and $m = 112$, where m is equal to N_i^{sat} that is discussed in Section III. In scenarios II and III, we consider $E[T^{nrt}_{\text{payload}}] = E[T^{rt}_{\text{payload}}] = 145.45$ μs (where $m = 115$) and $E[T^{nrt}_{\text{payload}}] = E[T^{rt}_{\text{payload}}] = 744$ μs (where

TABLE V
COMPARISONS OF SIMULATED AND ANALYTIC SATURATION THROUGHPUT

	Scenario I		Scenario II		Scenario III	
N^{nrt}	Analyze	Simulate	Analyze	Simulate	Analyze	Simulate
10	0.677000	0.615425	0.676000	0.629632	0.902000	0.83672
20	0.677000	0.615432	0.676000	0.629573	0.903000	0.836793
30	0.677000	0.615408	0.676000	0.629567	0.903000	0.836905
40	0.677000	0.615402	0.676000	0.629583	0.903000	0.836733
50	0.676000	0.615390	0.676000	0.629602	0.903000	0.836907
60	0.676000	0.615382	0.676000	0.629612	0.903000	0.836755

$m = 29$). For these three scenarios, we set $N^{rt} = 112$, 115, and 29, respectively in the analyzes. Since we evaluate the saturation throughput, we also assume that every station transmits their packets permanently.

Table V shows the simulation and analysis results under three different scenarios. We can find that the analysis results are very close to the simulation results. This implies that our analysis model is valid for modeling the DBASE protocol. Furthermore, this table also encourages us that the DBASE always reserves channel bandwidth for time bounded services.

VI. CONCLUSION

In this paper, we proposed a distributed bandwidth allocation/sharing/extension (DBASE) protocol to support multimedia traffic over IEEE 802.11 *ad hoc* WLAN. In DBASE, rt-stations can reserve and free channel resources dynamically. The system capacity of proposed DBASE was analyzed. Simulation results shown that the proposed protocol can support multimedia services of either CBR or VBR in *ad hoc* WLAN. Simulation results also demonstrated that the DBASE performs very well and much better than the conventional IEEE 802.11 standard with DCF. The channel efficiency of DBASE is up to 90% for only supporting VBR traffics. Besides, in DBASE, the packet loss probability of rt-packets, which is caused by the delay limitation and noise interference, is very low even though the total traffic load is heavy.

REFERENCES

[1] R. van Nee, M. Morikura, H. Takanashi, M. Webster, and K. W. Halford, "New high-rate wireless LAN standards," *IEEE Commun. Mag.*, vol. 37, pp. 82–88, Dec. 1999.

[2] *Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*, IEEE 802.11 Std, Aug. 1999.

[3] W. K. Kuo, C. Y. Chan, and K. C. Chen, "Time bounded services and mobility management in IEEE 802.11 wireless LANs," in *Proc. 1997 IEEE Int. Conf. Personal Wireless Commun.*, 1997, pp. 157–161.

[4] J. Sanchez, R. Martinez, and M. W. Marcellin, "A survey of MAC protocols proposed for wireless ATM," *IEEE Network*, vol. 11, pp. 52–62, Nov.–Dec. 1997.

[5] J. L. Sobrinho and A. S. Krishnakumar, "Distributed multiple access procedures to provide voice communications over IEEE 802.11 wireless networks," in *Proc. IEEE GLOBECOM'96*, vol. 3, 1996, pp. 1689–1694.

[6] ——, "Quality-of-service in *ad hoc* carrier sense multiple access wireless networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1353–1368, Aug. 1999.

[7] A. Muir and J. J. Garcia-Luna-Aceves, "Supporting real-time multimedia traffic in a wireless LAN," in *Proc. SPIE Multimedia Comput. Networking*, 1997.

[8] ——, "Group allocation multiple access with collision detection," in *Proc. INFOCOM'97*, Kobe, Japan, Apr. 1997, pp. 1182–1190.

[9] ——, "Group allocation multiple access in single-channel wireless LANs," in *Proc. Commun. Networks Distrib. Syst. Modeling and Simulation Conf.*, Phoenix, AZ, 1997.

[10] H. Luo, S. Lu, and V. Bharghavan, "A new model for packet scheduling in multihop wireless networks," in *Proc. ACM MOBICOM'00*, Boston, MA, 2000, pp. 76–86.

[11] N. H. Vaidya, P. Bahl, and S. Gupta, "Distributed fair scheduling in a wireless LAN," in *Proc. ACM MOBICOM'00*, Boston, MA, 2000, pp. 167–178.

[12] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 wireless LAN: Capacity analysis and protocol enhancement," in *Proc. IEEE INFOCOM'98*, San Francisco, CA, Apr. 1998, pp. 142–149.

[13] A. Iera, A. Modafferi, and A. Molinaro, "Access control and handoff management in multi-tier multimedia wireless systems," in *IEEE Wireless Commun. Networking Conf. WCNC'99*, vol. 3, 1999, pp. 1518–1522.

[14] C. R. Lin and M. Gerla, "Asynchronous multimedia multihop wireless networks," in *Proc. IEEE INFOCOM '97*, vol. 1, 1997, pp. 118–125.

[15] I. F. Akyildiz, D. A. Levine, and I. Joe, "A slotted CDMA protocol with BER scheduling for wireless multimedia networks," *IEEE/ACM Trans. Networking*, vol. 7, pp. 146–158, Apr. 1999.

[16] M. Decina and T. Toniatti, "Bandwidth allocation and selective discarding for a variable bit rate video and burst data cells in ATM networks," *Int. J. Digital Analog Commun. Syst.*, vol. 5, no. 2, pp. 85–96, Apr.–June 1992.

[17] H. Heffes and D. M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE J. Select. Areas Commun.*, vol. SAC-4, pp. 856–868, Sept. 1986.

[18] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE J. Select. Areas Commun.*, vol. SAC-4, pp. 833–846, Sept. 1986.

[19] T. G. Robertazzi, *Computer Networks and Systems Queueing Theory and Performance Evaluation*. New York: Springer-Verlag, 1994.

Shiann-Tsong Sheu (S'94–A'95) was born in Taiwan, R.O.C., in 1968. He received the B.S. degree in applied mathematics from National Chung Hsing University, Taiwan, R.O.C., and the Ph.D. degree in computer science from National Tsing Hua University, Taiwan, R.O.C., in 1990 and 1995, respectively.

Since 1995, he has been an Associate Professor with the Department of Electrical Engineering, Tamkang University, Taiwan, R.O.C.. His current research interests include ATM networks, WDM networks, wireless networks, and design and analysis of protocols and algorithms.

Tzu-Fang Sheu (S'00) received the B.S. degree in electrical engineering from Tamkang University, Taiwan, R.O.C., in 1998, and the M.S. degree from the same university in 2000.

Currently she is working toward the Ph.D. degree in Institute of Communication Engineering at National Tsing Hua University, Taiwan, R.O.C.. Her research interests are in the fields of wireless local area networks, wireless communication networks, and QoS in IP and wireless networks.