CHAPTER 1

Introduction

The IEEE 802.11 has been a well-known Wireless Local Area Network (WLAN) standard. After the 802.11 standard, a brand new standard which aims at Wireless Metropolitan Area Network (WMAN) was proposed.

In 1997 July, Institute of Electrical and Electronics Engineers (IEEE) made the standard of 802.11 [1]. The IEEE 802.11 contains Medium Access Control (MAC) layer and physical (PHY) layer. It uses 2.4 GHz radio frequency, and supports data rate of from 1 Mbps to 2 Mbps by Direct Sequence Spread Spectrum (DSSS). Later in 1999 September, the standard of 802.11b [2] was made in PHY layer using Frequency Hopping Spread Spectrum (FHSS) for increasing the data rate up to 11 Mbps. In 1999 September, another standard was called 802.11a [3]. 802.11a contains only PHY layer and uses another ratio frequency of 5.8 GHz with the technology of Orthogonal Frequency Division Multiplexing (OFDM). Its physical data rate can be up to 54 Mbps. 802.11g is like 802.11a, but it uses the same ratio frequency of 2.4 GHz as 802.11b. 802.11a can support data rate up to 54 Mbps, which is the same as 802.11g [4].

Since the radio frequencies used by the 802.11 wireless families are free, enterprises, companies, families, or individual persons can build their own wireless LAN, and avoid pulling network wires. The prices of IEEE 802.11 b/a/g dropped quickly, and data rates increased. More and more researches and commercial projects focus on 802.11 WLAN dramatically these years.

Since the higher applications such as video or voice have different QoS requirements for bandwidth, delay, jitter, and packet loss. The legacy IEEE 802.11 is not sufficient because it can not transmit data streams with different QoS requirements. The IEEE 802.11e has been drawn up for supporting QoS further.

The members of IEEE 802.16 had finished the original 802.16 standard in 2001[1], and 802.16a, 802.16c in 2002, 2003, respectively. In 2004, a revision of IEEE 802.16[2] standard was published, which is a revision of original 802.16, 802.16a, and 802.16c and also known as 802.16 REVd. All standards above belong to fixed broadband wireless access. Another version of 802.16 called 802.16e, which is an amendment to 802.16 on enhancements to support mobility, is under development.

The MAC part of 802.16 standard is not Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) that was adopted by 802.11 standard. The 802.16 uses Time Division Duplex (TDD) or Frequency Division Duplex (FDD) to access medium resource, these mechanisms have better efficiency than those of CSMA/CA. The 802.11 standard does not provide QoS support until the 802.11e standard was proposed. Unlike 802.11, the first 802.16 standard has QoS support, it divides all traffic flows into four classes according to their application type and each class has different priority.

1.1. Background

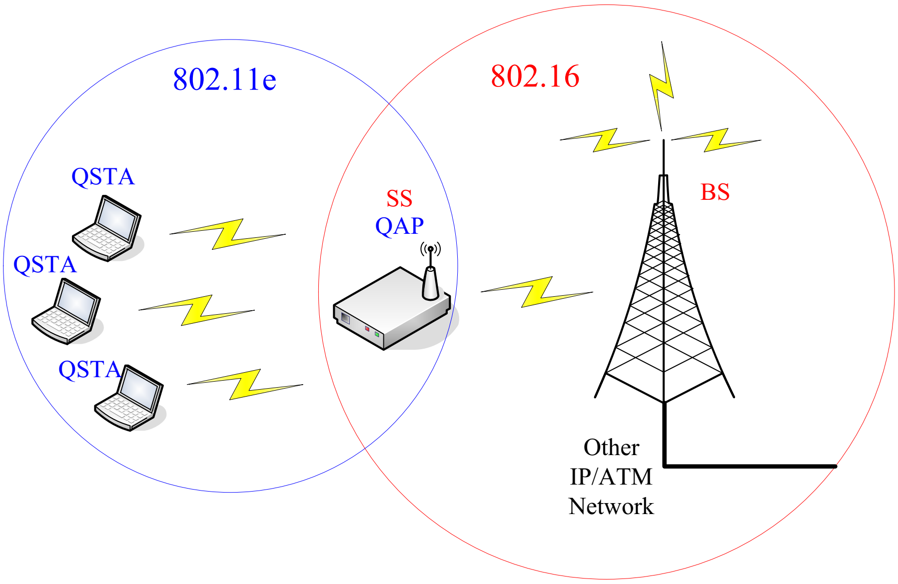

Figure 1-1: The architecture in 802.11e and 802.16 interworking network.

The Figure 1-1 is the architecture in our environment. There are few QSTA download or upload with QAP in 802.11e. At the same time, the QAP plays another role SS which connects with BS in 802.16. Thus, we call it SSQAP. We describe the basic knowledge of our research in this section. First, we introduce the token bucket mechanism we use in our research. Then we give detail description about the IEEE 802.11, IEEE 802.11e and IEEE 802.16 standard, which includes system architecture and QoS specifications.

1.1.1. Token Bucket Mechanism

Token bucket is a mechanism of controlling burst size and average transmission rate of a traffic flow. Figure 1-1 shows the token bucket mechanism. Every packet in the packet queue must get a token from the bucket before leaving the packet queue. There are r tokens generated per time unit and the generated tokens are put in the bucket. The bucket has size of b, which means it can hold b tokens at most. When the bucket is full, tokens newly generated will be discarded. When the bucket is empty, packets that want to leave the packet queue will wait until new tokens are generated. Of course, if the packet queue is full, the new coming packets will be dropped. We can control the burst size of a traffic flow by set different bucket size and control the average transmission rate of a traffic flow by set different token rate.

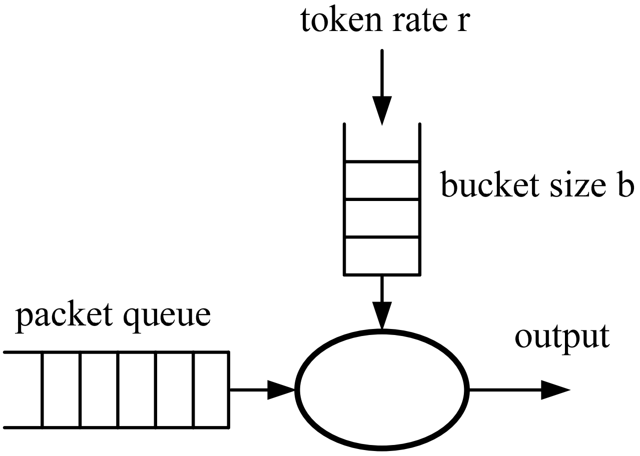

Figure 1-2: The token bucket mechanism.

1.1.2. The IEEE 802.11 Standard

The DCF is also known as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). For a 802.11 station to transmit, it shall sense the medium to determine if another station is transmitting. If the medium is not determined to be busy for greater than or equal to a DIFS (DCF Inter-Frame Space) period, the transmission may proceed. If the medium is determined to be busy, the station shall defer until the end of the current transmission. After deferral, or prior to attempting to transmit again immediately after a successful transmission, the station shall select a random backoff interval and shall decrement the backoff interval counter (or called backoff number) while the medium is idle. The basic access mechanism is illustrated in Figure 1-3.

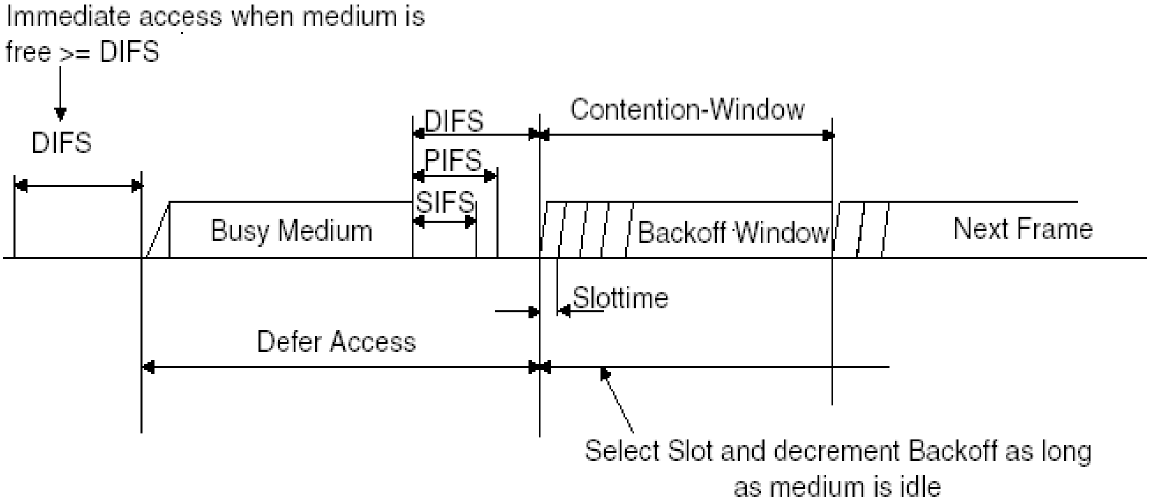

Figure 1-3: The Basic Access Method.

1.1.3. The IEEE 802.11e Standard

The 802.11e station is named QoS Station (QSTA). The QoS support is realized with the introduction of Access Categories (ACs). MAC Service Data Units (MSDUs) are delivered through multiple backoff instances within one QSTA. The 802.11e indicates four queues in MAC for Voice (VO), Video (VI), Best Effort (BE), and Back Ground (BK) traffic. The QoS services are such as Voice over Internet Protocol (VoIP), Videoconference, and Video on demand (VoD).

Each AC within the QSTA contends for a transmission opportunity (TXOP) and independently starts a backoff after detecting the channel being idle for an Arbitration Inter-Frame Space (AIFS); the AIFS is at least DIFS, and can be enlarged individually for each AC. A TXOP is defined in 802.11e as an interval of time when a QSTA has the right to initiate transmissions, defined by a starting time and a maximum duration. After waiting for AIFS, each backoff sets a counter to a random number drawn from the interval [0,CW[AC]]. CW is the short name of Contention Window. The minimum size (CWmin[AC]) and the maximum size (CWmin[AC]) of the CW are another parameters dependent on the AC. Priority over legacy stations is provided by setting CWmin[AC]<31, CWmax[AC]=1023 (in case of 802.11b PHY) and AIFS=DIFS.

As in legacy DCF, when a MSDU arrivals, it will set that backoff stage equals zero as the initial value, and every unsuccessful transmission attempt will increase the backoff stage. When the backoff stage researches the retry limit defined in 802.11, the backoff instance will reset backoff stage to zero again. The same as legacy DCF when the medium is sensed busy before the backoff number reaches zero, the backoff has to wait for the medium being idle for AIFS again, before continuing to count down the counter. The big difference from the legacy DCF is that when the medium is determined as being idle for the period of AIFS, the backoff number is reduced by one beginning the last slot interval of the AIFS period. Note that with the legacy DCF, the backoff number is reduced by one beginning the first slot interval after the DIFS period. The AIFS is determined by the parameter AIFS Number (AIFSN) defined in 802.11e:

$$\text{AIFS[AC]} = \text{SIFS} + (\ \text{AIFSN[AC]} * \text{slot time}\ )$$

The slot time is the basic unit of backoff number defined in 802.11. Since the minimum AIFS equals DIFS, the minimum AIFSN equals 2. The parameters about CW and AIFS are listed in Table 1.1..

Table 1-1: Default EDCA Parameter set.

AC	***CWmin***	***CWmax***	***AIFSN***
VO	(CWmin+1) / 4 - 1	(CWmin+1) / 2 -1	2
VI	(CWmin+1) / 2 - 1	aCWmin	2
BE	aCWmin	aCWmax	3
BK	aCWmin	aCWmax	7

After any unsuccessful transmission attempt a new CW is calculated with the help of the Persistence Factor (PF[AC]) and another uniformly distributed backoff number out of this new, enlarged CW is drawn, to reduce the probability of a new collision. Whereas in legacy 802.11 CW is always doubled after any unsuccessful transmission (equivalent to PF = 2), 802.11e uses the PF to increase the CW different for each AC:

$$\mathrm{CW[AC]} = (\,(\mathrm{CWmin[AC]}+1) * \mathrm{PF}^{\text{backoff stage}}\,) - 1$$

The CW never exceeds the parameter CWmax[AC], which is the maximum possible value for CW. The backoff instance transmits packets like Figure 1-4. There are four ACs in one QSTA, and each AC start to count down the backoff number after AIFS[AC]. Finally, one of the backoff number counts to zero and this backoff instance will try to access wireless medium. In Figure 1-4, the backoff number of VO backoff instance counts down to zero first, and transmits data with RTS/CTS machine.

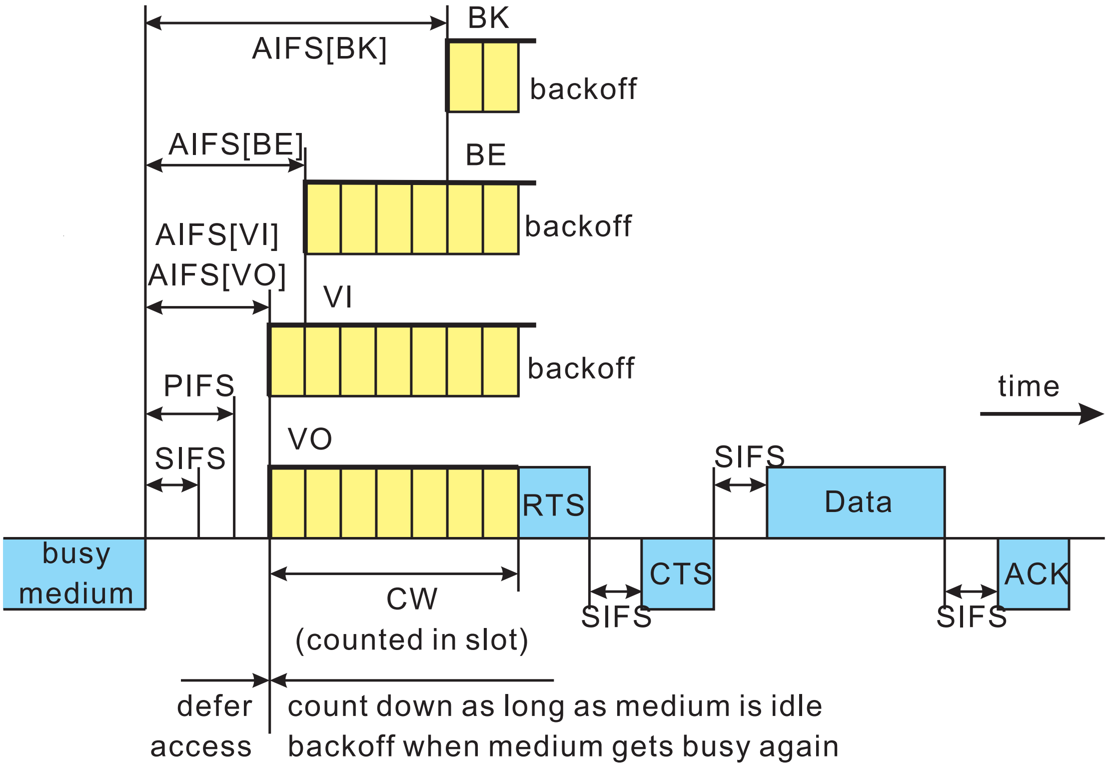

Figure 1-4: Multiple EDCA backoff entities contention with RTS/CTS.

A single QSTA may implement up to four transmission queues realized as virtual stations inside a QSTA, with QoS parameters that determine their priorities. If the counters of two or more parallel ACs in a single QSTA reach zero at the same time, a scheduler inside the QSTA avoids the virtual collision. The scheduler grants the TXOP to the AC with highest priority, out of the ACs that virtually collided within the QSTA, called virtual collision, as illustrated in Figure 1.5. There is then still a possibility that the transmitted frame collides at the real wireless medium with a frame transmitted by other QSTAs.

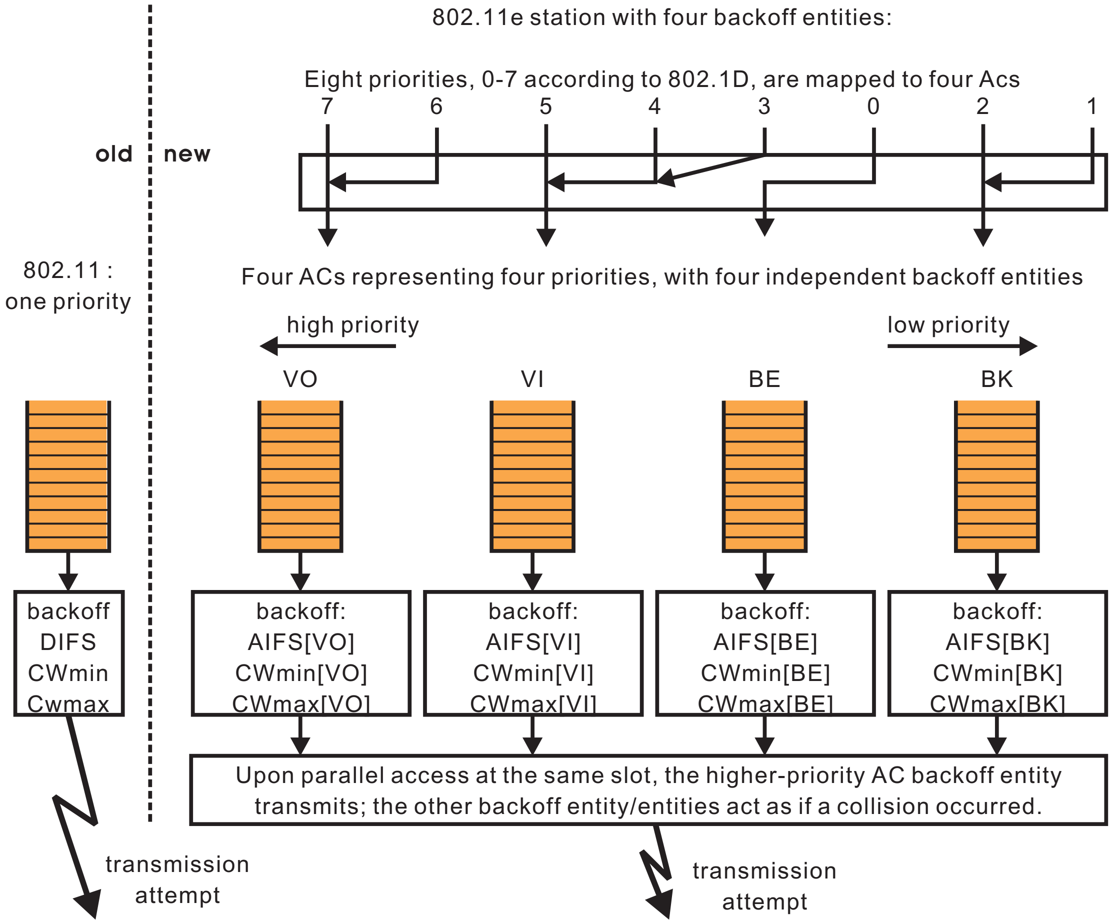

Figure 1-5: Access Categories in the IEEE 802.11e.

1.1.4. The IEEE 802.16 Standard

In 802.16, the client side is called Subscriber Station (SS), and the server side is called Base Station (BS). The IEEE 802.16 standard has two operation modes. One is infrastructure mode, which is like traditional client-server mode, and all SSs must contact its BS before transmitting. The other is mesh mode, which allows a SS transmit data directly to other SSs. We focused on the former mode in this paper.

There are four QoS classes defined in 802.16, which are Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), non-real-time Polling Service (nrtPS), Best Effort

(BE). Table 1.1 shows the QoS classes of 802.16, and the upper class has higher priority than lower one. The higher priority means the more right to use the resource.

Table 1-2: Four QoS classes of 802.16.

Class name	Traffic type	Application
UGS	Real-time Constant Bit Rate (CBR)	Voice over IP (VoIP)
rtPS	Real-time Various Bit Rate (VBR)	Real-time video
nrtPS	Non-real-time	Bandwidth-sensitive FTP
BE	Non-real-time	HTTP, Telnet

Traffic flows in 802.16 are treated as connections. A traffic flow must establish connection with its BS before transmitting data. The operation process of 802.16 is shown in Figure 1.6 [5]. The blocks drawn with dotted line in Figure 1 are the parts undefined in 802.16.

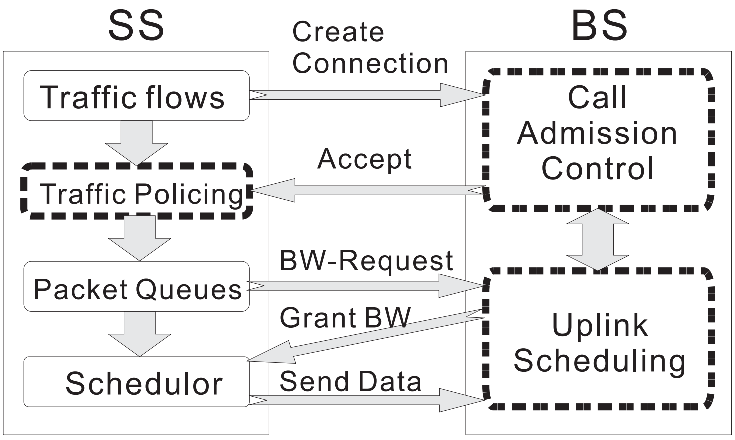

Figure 1-6: The operation process of 802.16.

There are many traffic flows in a SS and each of them is treated as connections. In order to get permission to transmit data, traffic flows first establish connection with its BS.

A traffic flow will transmit some parameters, such as token rate, bucket size, and delay requirement, to BS and the Call Admission Control (CAC) model that resides in BS will decide whether the BS accept this traffic flow. If the CAC model decides to accept it, it will notify the Traffic Policing model that resides in SS and send the two parameters, which are token rate and bucket size of the traffic flow, to the Traffic Policing model in order to enforce the traffic flow on obeying its traffic contract. Here the traffic policing model can be simply achieved by applying token bucket mechanism [5]. After establishing connection with its BS, the traffic flow should tell the BS how much resource it need by sending bandwidth request (BW-request) to the BS. We use the architecture in [5]: a connection sends its queue size as BW request. The uplink scheduling model that resides in BS will gather the BW-requests from all connections and schedule how to distribute resource to all connections. Then the BS will transmit the result of scheduling to SSs, and the scheduler reside in SSs will transmit data according to the result.

The 802.16 standard divides transmission time into super frames and each super frame can be divided into a downlink sub-frame and an uplink sub-frame, both of them contain many time slots. A time slot is the minimum unit of transmission in 802.16 and the duration of a time slot depends on the physical layer (PHY). Downlink means the direction of transmission is from BS to SS, and the uplink means the direction of transmission is from SS to BS. The downlink scheduling is simple because the only sender is BS. Every SS only needs to listen to if there is data sent to it. Hence we focused on uplink scheduling in our research. Figure 1.3 shows the frame structure of 802.16. The downlink MAP (DL-MAP) field in downlink sub-frame contains the information that the where the data in downlink sub-frame is sent to. The uplink MAP (UL-MAP) field in downlink sub-frame contains the information of distributing resource (e.g. the result of the uplink packet scheduling that resides in BS). The remainder of the uplink sub-frame is the transmission area of each SS.

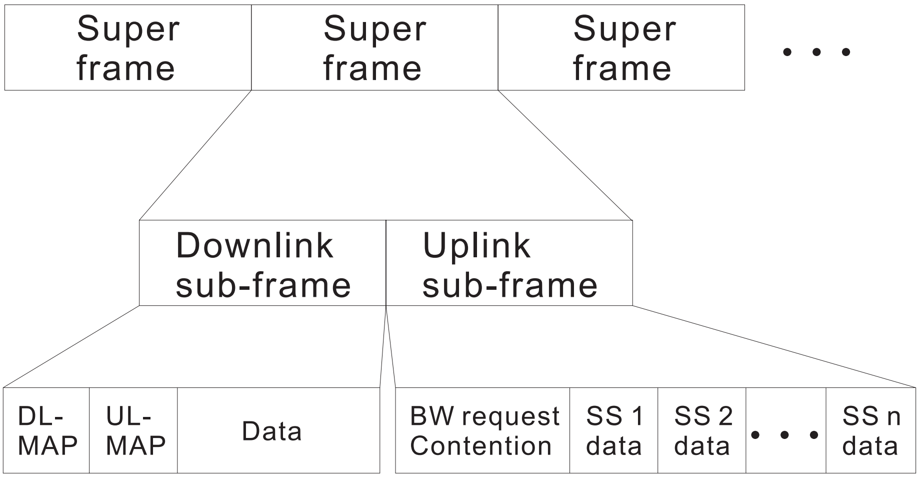

Figure 1-7: The frame structure of 802.16.

Basically, the 802.16 can be seen as a polling-based system. Every connection needs to be polled by BS before transmitting anything including BW-request. If the BS does not poll a connection, this connection cannot transmit anything to BS. The results of the uplink packet scheduling, which reside in BS, indicate that which SS can transmit in certain time slots. So actually the BS indicates the poll rule by UL-MAP. When the load of system is heavy, it is possible that the there is not enough resource for polling all connections. In this situation, the connections of the lower classes may get no transmitting opportunities of BW-requests because the resource is not enough. The connections of the lower classes can’t send BW-requests to the BS; also they can not transmit any data to BS. Hence, there is a BW request contention period in the 802.16 uplink sub-frame. It is for the classes of lower priority, such as nrtPS and BE, to get the opportunities of sending BW requests through contention when system is too busy to poll all connections. The easiest mechanism of contention resolution is binary backoff. Besides polling, a connection can send its BW-request to BS when it sends its data. Some bandwidth for sending data is borrowed for transmitting BW-request. This is called piggyback.

Although the 802.16 defines for QoS classes and mentions that scheduling should be used to support QoS, but the 802.16 standard only defines the scheduling of UGS class: allocate fixed bandwidth during fixed time. Because the scheduling of UGS class is allocating fixed bandwidth during fixed time, an UGS connection does not need to transmit any BW-requests. The scheduling of other classes is undefined in the 802.16 standard.

1.2. Motivation

In recent years, there are more and more research issues focus on QoS. 802.11 without QoS mechanism is not satisfied on wireless network. For example, real-time applications, such as VoIP and real-time video, can not tolerate long delay time, but the non-real-time applications, such as FTP and HTTP, can do. Therefore, real-time applications often have delay requirements. We should pay more attention to the real-time applications because we must guarantee that they have enough bandwidth and do not have longer delay time than its delay requirement. Both 802.11e and 802.16 have four Access Categories to support QoS guarantee. But the MAC of 802.11e is CSMA/CA, which is different from the TDD of 802.16, also is the major problem in interworking network between 802.16 and 802.11e. Thus, we hope to propose a complete QoS architecture in 802.16 integrating with 802.11e. Markov Chain mechanism is used for analyzing 802.11e MAC delay to meet the delay requirement. Token bucket mechanism has some useful characteristics that can help us to predict the traffic behaviors. So we employ these two mechanisms in our research. Details will be shown in the later sections.

1.3. Organization

The rest of this thesis is organized as follows. In Chapter 2, we first introduce related works about the Markov Chain Model about IEEE 802.11 DCF and IEEE 802.11e EDCA QoS. Second, describing the Token Bucket and the Call Admission Control in IEEE 802.16. In Chapter 3, we propose and explain our four-dimension Markov Chain model. Because we just use it for analyzing the MAC delay, we verify the accuracy of the model by comparing the delay with the result in the network simulator Qualnet. We describe our token rate estimation model, CAC, and Packet Drop mechanism in Chapter 4. In Chapter 5, we valid the models proposed by us. Simulation results are shown in this Chapter. Conclusions and future work are described in Chapter 6.