Design of an Efficient Guarantee Scheme for Hard Real-Time Messages on the IEEE 802.11 Wireless LAN

Mikyung Kang$^1$ and Junghoon Lee$^2$

$^1$ Department of Computer Science and Statistics, Jeju National University, 690-756, Ara 1-Dong, Jeju City, Republic of Korea
Tel.: +82-64-754-3594, Fax.: +82-64-725-2579

$^2$ Department of Computer Science and Statistics, Jeju National University, 690-756, Ara 1-Dong, Jeju City, Republic of Korea
Tel.: +82-64-754-3594, Fax.: +82-64-725-2579
E-mail: mkkang@cheju.ac.kr, jhlee@cheju.ac.kr

Abstract: This paper proposes and analyzes a round-robin style polling method and corresponding bandwidth allocation scheme for hard real-time messages on the wireless LAN conforming to IEEE 802.11 standard. Directly considering the deferred beacon problem caused by the intervention of non-real-time message, the bandwidth allocation scheme decides the capacity vector that corresponds to the polling schedule on CFP. The simulation results show that the proposed scheme not only enhances the schedulability of wireless network by up to 18%, but also yields more efficient schedule that provides more bandwidth for the non-real-time traffic up to 5.3%.

1. Introduction

Wireless communication technology has gained a widespread acceptance in recent years. As the speed and capacity of WLAN (Wireless Local Area Network) increase, so does the demand for supporting multiple time-sensitive high-bandwidth applications. For example, applications such as VOD (Video On Demand) and interactive multimedia require improved guarantee of QoS (Quality of Service) from WLAN[1]. In addition, the messages exchanged in the wireless sensor network, one of the promising application area of wireless technology, have hard real-time constraints that they should be exchanged within a bounded delay (as long as there is no network error) [2].

Wireless networks are inherently broadcast media, namely, all nodes in the network share one common communication medium. The IEEE 802.11 was developed as a MAC (Medium Access Control) standard for WLANs[3]. The standard includes a basic DCF (Distributed Coordination Function) and an optional PCF (Point Coordination Function) to support an efficient delivery of both real-time and non-real-time messages. The DCF uses CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) protocol as the basic channel access mechanism to transmit asynchronous data in the contention period. It is well known that packet collisions are intrinsic to CSMA. They occur because each node has only a delayed perception of the other nodes’ activity. Especially in wireless environment, they also happen due to hidden nodes[4]. While the CSMA protocol is robust and simple to implement, this contention-based MAC protocol cannot guarantee a bounded network access time for real-time traffics, failing in providing the predictability in message schedule.

A real-time guarantee can be provided only when a deterministic contention free schedule is developed[5]. As a polling-based protocol capable of completely eliminating the possibility of packet collision, the PCF ensures that the real-time station can deterministically access the common channel. However, the use of centralized scheme in PCF constrains the operation of WLAN. Previous researches have already pointed out its poor performance[6]. Even worse, the coexistence of a non-real-time traffic in the network accompanies the deferred beacon problem, seriously deteriorating the schedulability of the network system. As a result, communication protocol should provide an efficient use of the available bandwidth while satisfying the QoS requirement of both real-time and non-real-time applications.

Hard real-time guarantee is difficult because it strongly depends on the underlying polling policy, for example, how polling interval is started, how the coordinator selects the node to be polled, and how long a node transmits when it is polled. With weighted round robin scheme, the coordinator polls each node one by one, and the polled node transmits its message for a predefined time. The set of these predefined times for respective nodes is called as capacity vector. The main focus of this paper is to decide the capacity vector for the given parameters of both network and message stream. It is true that round robin polling policy increases the number of total polls, while each poll wastes the network time. However, it makes the guarantee mechanism much simpler, copes with the deferred beacon problem, and expects the schedulability improvement via dual network architecture.

This paper is organized as follows: After issuing the problem in Section 1, Section 2 introduces the related works focusing on real-time communications on the wireless medium. Then Section 3 proposes the bandwidth allocation scheme after describing the IEEE 802.11 WLAN architecture. Section 4 discusses the performance measurement result and then Section 5 finally concludes this paper with a brief summarization and the description of future works.

2. Related Works and Background

2.1 Real-time Communication on WLAN

Several MAC protocols have been proposed to provide bounded delays for real-time messages along with non-real-time data over a wireless channel. However, these protocols
are typically based on a frame-structured access which consists of a contention part and reservation part, so they cannot be easily applied to the IEEE 802.11 WLAN. As an example, Choi and Shin suggested a unified protocol for real-time and non-real-time communications in wireless networks based on an infrastructure that is composed of a wired backbone network and a number of base stations which communicate with the wireless mobile nodes[7]. This scheme devises a minislot for the network reservation.

M. Caccamo and et. al have proposed a MAC that uses deterministic real-time scheduling to implement TDMA Time Division Multiple Access)[5]. Referred as implicit contention, this scheme makes each node concurrently run a real-time scheduling algorithm to determine which message has access to the medium. Each message implicitly contends for the medium through the scheduling algorithm, for example, with priorities, instead of on the physical medium. However, with this implicit contention, each node must schedule all messages in the network, making difficult to scale to large networks.

As another example, in BB (Black Burst) contention scheme, stations sort their access rights by jamming the channel with pulses of energy before sending their packet[8]. Since packets must be transmitted repeatedly in a constant interval, sending burst of energy for each packet will waste bandwidth considerably. Moreover, the BB contention is not a regular scheme defined in IEEE 802.11 standard.

Most works that conform to the IEEE standard are aiming at enhancing the ratio of timely delivery for soft multimedia applications, rather than providing a hard real-time guarantee. DBASE (Distributed Bandwidth Allocation/Sharing/Extension) is a protocol to support both synchronous and multimedia traffics over IEEE 802.11 ad hoc WLAN. The basic concept is that each time real-time station transmits packet it will also declare and reserve the needed bandwidth at the next CFP. Every station collects this information and then calculates its actual bandwidth at the next cycle. This scheme can be ported to WLAN standard, but it is not essentially designed for hard real-time message streams.

In the non-academic effort, IETF has produced new drafts, EDCF (Enhanced DCF) and HCF (Hybrid Coordination Function) to replace the CSMA/CA based and centralized polling based access mechanism, respectively[6]. No guarantees of service are provided, but EDCF establishes a probabilistic priority mechanism to allocate bandwidth based on traffic categories. According to HCF, a hybrid controller polls stations during a contention-free period. The polling grants a station a specific start time and a maximum transmit duration.

2.2 IEEE 802.11 LAN

Wireless LAN operates on both CP (Collision Period) and CFP (Collision Free Period) intervals alternatively as shown in Figure 1. A superframe, or frame in short, consists of a CFP and a CP. In CFP interval, under the control of PCF, only the node polled by coordinator node, typically AP (Access Point), can transmit its messages. Note that it is mandatory that a superframe includes a CP, and its length should be large enough to send at least one data packet under DCF.

![Figure 1 Time axis of wireless LAN](image)

The frames of different priorities have to wait different IFSs (InterFrame Space) before they are transmitted. The SIFS (Short IFS) is used by immediate control frames, which always have the highest priority such as acknowledgement. The PIFS (Priority IFS) is used by real-time frames, while DIFS (DCF IFS), the longest IFS, is used by non-real-time frames[3].

According to the network parameters configured in priori ori, the PC (Point Coordinator) initiates CP and CFP by announcing a management frames such as StartBeacon and EndBeacon frames with PIFS, respectively. PC is responsible for generating beacons which generally contain a time information. As an example, this information includes the traffic indication map which identifies the station and length of unicast/multicast messages that will occupy in the following CFP. In IEEE 802.11, there are several problems with the PCF that led to the development of enhancements to the protocol. The problems of the PCF scheme are as follows:

- The transmission of the beacon by the point coordinator depends on whether the medium is idle at the time of TBTT (Target Beacon Transmission Time). But after the medium is idle the PC will get the priority due to the shorter PIFS. But the time at which the medium becomes idle is unpredictable. Thus the beacon frame can get delayed affecting the time allocated to time-bounded traffic. This is referred to as the deferred beacon problem of PCF.
- The duration of the transmission that happens after the polling is not under the control of the point coordinator.

Figure 2 illustrates the effect of deferred beacon frame. The figure shows that the start of CFP and hence the network access of a node may be delayed by up to $D_{max}$, which coincides with the maximum non-real-time message size. As the hard real-time guarantee is based on the worst case available time, deferred beacon is another serious schedulability degradation factor in the wireless LAN.

![Figure 2 Deferred beacon problem](image)
3. Bandwidth Allocation

3.1 Network Specification and Message Model

In IEEE 802.11 WLAN, the IBSS (Independent Basic Service Set) is the most basic type, where the stations are able to communicate directly. This mode is possible when the stations are close enough to form a direct connection without preplanning. We also assume that a PC polls each node during CFP in turn, and only the polled node can transmit its message up to a predefined time amount according to the capacity vector calculated offline. The PC tries to initiate the CFP at fixed time, though a non-real-time message may defer the initiation. The ideal value of superframe is denoted as $F$. In real-time communication literature, the term real-time traffic typically means isochronous (or synchronous) traffic, consisting of message streams that are generated by their sources on a continuing basis and delivered to their respective destinations on a continuing basis[9]. The traffic characteristics such as period and message size are usually known in priori of system operation. We assume that there are $n$ streams of real-time messages, $S_1, S_2, ..., S_n$. Each stream can be modeled as follows: A message arrives at the beginning of its period and must be transmitted by the end of the period. The period of stream, $S_i$, is denoted as $P_i$, as well as the maximum length of a message is $C_i$. The first message of each stream arrives at time 0. As is the case of other works, we assume that each stream has only one stream, and this assumption can be generalized with virtual station concept[11].

3.2 Allocation Procedure

By allocation, we mean the procedure of determining capacity vector, $\{H_i\}$, for the given superframe time and message stream set. Figure 3 illustrates that more network bandwidth is assigned to the stream with more utilization. That is, $H_i \geq H_j$ if $\frac{C_i}{P_i} \geq \frac{C_j}{P_j}$ for different $S_i$ and $S_j$. For a minimal requirement, $F$ should be sufficiently large enough to make the polling overhead ($\delta_p$) insignificant. Additionally, if $S_{min}$ is the stream which has the smallest period among $\{P_i\}$, $F$ should be less than $P_{min}$ so that $S_{min}$ can meet at least one frame within $P_{min}$. That is,

$$\delta_p + D_{max} \leq F \leq P_{min} \quad (1)$$

, where $\delta_p$ is total overhead in a superframe, for example, polling latency, IFS, and so on. In (1), the term, $D_{max}$, in the left hand side reflects the mandatory requirement of IEEE 802.11 WLAN that $F$ should reserve the time amount for at least a data packet to be delivered.

![Figure 3](image)

It is desirable that the superframe time is a hyperperiod of the set and it is known that a message set can be made harmonic by reducing message periods by at most half[10]. So we assume that the superframe time is also given in priori. For the simplicity, one polling round is assumed to complete within a frame. We will remove the assumption after a while. In WLAN, the guarantee mechanism strongly depends on the polling policy, so they cannot be considered separately. How many polls a station receives is different period by period. Guarantee requires that even in the period which meets the smallest number of poll, the message can be transmitted within its period. Figure 4 illustrates the worst case available time for $S_i$. $P_i$ contains a series of superframes, and let $R_i$ be the residual of dividing $P_i$ by $F$. Then the following equation holds

$$P_i = F \cdot s + R_i, \quad R_i = P_i - \left\lfloor \frac{P_i}{F} \right\rfloor \cdot F$$

, where $s$ is an integer part of $\frac{P_i}{F}$.

Without the deferred Beacon, the CFP precisely starts at the fixed time instants. In that case, for $S_i$, the least bound of network access within $P_i$ is $\left\lfloor \frac{P_i}{F} \right\rfloor$, as natural. On the contrary, if we consider the deferred beacon, the deferred start of the last superframe may invalidate one access of stream when $R_i$ is less than $D_{max}$, as show in Figure 4(b). If $R_i$ is greater than $D_{max}$, the number of available time slots is not affected in spite of deferred beacon.

![Figure 4](image) Worst case analysis

As a result, the minimum of available transmission time, $X_i$ is calculated as Eq. (1). Namely,

$$X_i = \left( \left\lfloor \frac{P_i}{F} \right\rfloor - 1 \right) \cdot H_i \quad if \left( P_i - \left\lfloor \frac{P_i}{F} \right\rfloor \cdot F \right) \leq D_{max} \quad (2)$$

$$X_i = \left\lfloor \frac{P_i}{F} \right\rfloor \cdot H_i \quad Otherwise$$

For each message stream, $X_i$ should be greater than or equal to $C_i$. With this property, we can decide the minimum bound of $H_i$ that can guarantee the timely delivery for the hard real-time message streams. Namely,

$$X_i \geq C_i \quad (3)$$

Note that $D_{max}$ is the least bound of CP. The allocation vector obtained as Eq. (1) is a feasible schedule if the vector meets the following constraints;

$$\sum H_i + \delta \leq F - D_{max} \quad (4)$$

With these constraints, we can determine the length of CFP ($T_{CFP}$) and that of CP ($T_{CP}$), according to the following equation.

$$T_{CFP} = \sum H_i + \delta_p, \quad T_{CP} = F - T_{CFP} - \delta_m \quad (5)$$
where $\delta_m$ is the overhead invoked to start and end the beacon. As this allocation scheme generates the larger CP for the given $F$, we can expect the more efficient use of network bandwidth. Additionally, this calculation is easily fulfilled with simple arithmetic operations.

4. Performance Measurement and Discussion

A comparison with TDMA is impossible because it essentially requires so many assumptions to be decided on slot size, slot allocation scheme, and the way to cope with deferred beacon. Moreover, TMDA-based schemes are built on top of an assumption that there is no CP. Most previous works guarantee meeting the time constraint for the hard real-time messages based on a pessimistic assumption that every stream suffers from the deferred beacon[1]. This is resulted from the fact the their schemes cannot accurately estimate the intervention of non-real-time message. We compare the schedulability of our scheme with this pessimistic guarantee mechanism.

We generate 2000 streams sets whose utilization is greater than or equal to 0.68 and less than 0.70. The number of streams randomly distributes from 2 to 10. In the experiment, every time variables are aligned to the frame time. Accordingly, the period of each stream ranges from $5.0F$ to $10.0F$, while its message length from $0.3F$ to $3.0F$. Figure 5 plots schedulabilities of proposed and pessimistic schemes according to $D_{max}$. $D_{max}$ is also presented as the ratio to the frame time. At low $D_{max}$, both schemes show same schedulability, but our scheme generates the $T_{CP}$ up to 5.3% larger than that generated by the pessimistic scheme, inviting more non-real-time message to occupy the network. The figure also shows that our scheme improves the guarantee ratio by 18% at maximum.

![Figure 5 Measured guarantee ratio](image)

Up to now, a polling round is contained in a superframe. However, it doesn’t matter even if a polling round spreads over more than one frame, say $m$. In that case, a frame performs only a partial polling round. The worst case access time for $S_t$ arises when the frame belonging to the last frame of its period is deferred. Only in the allocation part of Eq. (2), $F$ should be replaced with $mF$ while $F$ in the condition part intact.

5. Conclusion and Future Work

The deferred beacon problem is a serious hindrance to the real-time guarantee in the IEEE 802.11 WLAN standard. This paper has proposed and analyzed an hard real-time guarantee scheme based on the round robin style polling mechanism on the wireless network that conforms to the standard. The key idea is to decide the CFP polling schedule in the form of capacity vector for the given network parameters and time constraints, after analyzing the effect of deferred beacon to the hard real-time message stream. The experiment demonstrates that the proposed scheme improves the schedulability for the experimental real-time stream sets by up to 18%. As a work in progress, we are developing another hard real-time guarantee scheme that can minimize the deferred beacon problem via dual or multiple network architecture. As natural, this works is based on the round-robin style polling mechanism.

References