An Efficient Bandwidth Reclaim Scheme for the Integrated Transmission of Real-Time and Non-Real-Time Messages on the WLAN

Junghoon Lee¹, In-Hye Shin¹, Gyung-Leen Park¹, Wang-Cheol Song²,
Jinhwan Kim³, Pankoo Kim⁴, and Jiman Hong^{5,*}

¹ Dept. of Computer Science and Statistics, Cheju National University

² Dept. of Computer Engineering, Cheju National University

³ Dept. of Multimedia Engineering, Hansung University

⁴ Dept. of Computer Engineering, Chosun University

⁵ School of Computing, Soongsil University

jhlee@cheju.ac.kr, kimjh@hansung.ac.kr, pkkim@chosun.ac.kr,
gman@ssu.ac.kr

Abstract. This paper proposes and analyzes bandwidth reclaim scheme for IEEE 802.11 WLAN, which may suffer from severe bandwidth waste resulting from not only the variation of transmission rate and message length but also the overallocation to the real-time traffic in compensating for the delay due to the intervention of non-real-time messages. Built on top of the weighted round robin scheduling policy, we address that the polling order rearrangement according to the degree of overallocation can enhance reclaimability of unused network time and that the rearranegable slot has its message pending at the rearranging time. The simulation results show that the proposed scheme is able to reclaim up to 52.3 % of bandwidth waste when the number of streams is 2 and that it also provides stable throughput for utilization of 0.5 through 0.8.

1 Introduction

¹ According to the expansion of WLAN (Wireless Local Area Network), real-time and non-real-time messages coexist in the wireless media. Real-time traffic such as video and sensor data requires bounded delay, but is usually tolerant of some packet losses. As contrast, non-real-time traffic requires loss-free transmission without demanding bounded delay[1]. The IEEE 802.11 was developed as a MAC (Medium Access Control) standard for WLAN and this standard consists of both an essential DCF (Distributed Coordination Function) and an optional PCF (Point Coordination Function)[2]. The DCF exploits collision-based CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) protocol for non-real-time messages, aiming at enhancing their average delivery time as well as overall

* Corresponding author.

¹ This research was supported by the MIC, Korea, under the ITRC support program supervised by the IITA (IITA-2006-C1090-0603-0040).

network throughput. On the other hand, the collision-free PCF can provide a real-time guarantee by developing a deterministic access schedule[3].

Network management consists of two parts on real-time communications, namely, static bandwidth allocation and dynamic adaptation parts, respectively [4]. Based on the static informations that do not change for a long time, for example, period and maximum transmission time of each stream, the bandwidth allocation procedure determines the network access schedule for the given set of active streams. However, the dynamic change in network condition needs additional management that can cope with such situations as non-real-time traffic load oscillation, channel status change, and so on. One of challenging problems in WLAN is a bandwidth reclaim scheme that reassigns the network time reserved but unused to another node[3]. In particular, the auto-selection mechanism can create much of such unused bandwidth, as it chooses the rate to be used for each packet that is submitted to the physical layer[5]. If a stream meets better rate than expected for a time interval, it can complete its transmission early.

The reclaim scheme is very crucial to the network throughput, as hard real-time guarantee inevitably brings bandwidth overallocation resulted from a pessimistic assumption that the stream has the worst case available time at each period. Moreover, a phenomenon called as a *deferred beacon problem*, which will be discussed in Section 3.1, further deteriorates the worst case available time for the real-time stream, increasing the amount of overbooking. It is desirable that as much unused bandwidth as possible should be reclaimed and allocated to another stream to minimize bandwidth waste. To this end, this paper is to propose and analyze a bandwidth reclaim scheme for WLAN that strictly obeys the IEEE 802.11 standard, assuming that PCF operates according to the weighted round robin schedule. We can improve the amount of reclaimed bandwidth by adjusting the polling order.

This paper is organized as follows: After issuing the problem in Section 1, Section 2 introduces the related works on both network scheduling and bandwidth reclaim schemes. With the description on network and message models along with a bandwidth allocation scheme in Section 3, Section 4 proposes a reclaim procedure. Section 5 discusses the performance measurement results and then Section 6 finally concludes this paper with a brief summarization and the description of future works.

2 Related Works

Based on the attribute that most real-time traffic is periodic, several MAC protocols have been proposed to support the hard real-time communication over a wireless channel[6]. However, they cannot be easily exploited to the IEEE 802.11 WLAN standard, as they ignored the CSMA/CA part defined as mandatory in the WLAN standard, or just aimed to enhance the ratio of timely delivery for soft multimedia applications[4]. For example, Choi and Shin suggested a unified protocol for real-time and non-real-time communications in wireless networks[1]. Based on frame-structured access mechanism, a BS (Base Station) polls every

station, be it a real-time or a non-real-time one, according to the corresponding policy. Though unpredictability due to message collision is eliminated, this scheme is neither compatible with the standard CSMA/CA protocol, nor takes into account any resource reclaim scheme.

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 that supports multimedia traffics over IEEE 802.11 *ad hoc* WLAN[4]. The basic concept is that each time real-time station transmits its packet it will also declare and reserve the bandwidth needed at the next CFP. Every station respectively collects this information and then calculates its actual bandwidth at the next cycle. Though the per-packet reservation makes a resource reclaim scheme unnecessary, it does not only increase the runtime burden on member stations, but also demands all stations receive all the transmitted packets.

M. Caccamo and his colleagues have proposed a MAC capable of supporting deterministic real-time scheduling by implementing TDMA[3]. Referred as *implicit contention*, their scheme makes every station respectively run the common real-time scheduling algorithm to determine which message can access the medium. Accompanied with the implicit contention, FRASH (FRAame SHaring) can reclaim the unused bandwidth. Whenever the transmission of the current dispatched message is over and it does not use all the reserved frames, its identifier is put in a field in the header of the last data packet of the current message. However, the identifier should be correctly received by all stations in the network to reach a global agreement. FRASH can perform properly only for TDMA protocols that operate on fixed size slots. Moreover, this scheme cannot be implemented in the current 802.11 WLAN standard without the addition of new management frames and thus causes additional overhead.

3 Backgrounds

3.1 Network and Message Models

In BSS (Basic Service Set), the time axis of WLAN is divided into a series of superframes, each of which alternately operates CP (Collision Period) and CFP (Collision Free Period) phases, which are mapped to PCF and DCF, respectively. PC (Point Coordinator) node, typically AP (Access Point), sequentially polls each station during CFP according to the polling schedule determined by a specific policy such as EDF (Earliest Deadline First). All stations including the ones even in the polling list contend in the CP to send a management frame or a control frame. Even in the *ad hoc* mode, it is possible to designate a specific node to play a role of PC in a target group.

The phase of network operation is managed by the exchange of control packets which have higher priority than other ordinary packets. The PC attempts to initiate CFP by broadcasting a *Beacon* at regular intervals derived from a network parameter of *CFPRate*. Round robin is one of the commonly used polling

policies for CFP, in which every node is polled once a polling round. A polling round may be completed within one superframe, or spread over more than one superframe. In case the CFP terminates before all stations have been completely polled, the polling list is resumed at the next node in the ensuing CFP cycle. The polled node transmits its message for up to a predefined time interval, and always responds to a poll immediately whether it has a pending message or not.

To prevent starvation of stations that are not allowed to send during the CFP, a superframe is forced to include a CP of minimum length that allows at least one data packet delivery under DCF[2]. Hence, a non-real-time packet may occupy the network when the coordinator is to send the beacon frame. However, only after the medium is idle the coordinator will get the higher priority due to the shorter IFS (InterFrame Space). Thus, the delivery of a beacon frame can get delayed, resulting in the *deferred beacon problem*, possibly invalidating the network schedule determined for real-time messages. The maximum amount of deferment coincides with the maximum length of a non-real-time packet.

The real-time traffic is typically *isochronous* (or synchronous), consisting of message streams that are generated by their sources on a continuing basis and delivered to their respective destinations also on a continuing basis[6]. For example, a sensor node periodically reports the collected sensor data to a remote server. Accordingly, the general real-time message model consists of n streams, namely, $S_1, S_2, \dots, S_n$, and for each S_i , a message sized to at most C_i arrives at the beginning of its period, P_i , and it must be transmitted by P_i . M_i is the transmission time of C_i , estimated with a reference transmission rate, R_i , which can be set empirically or analytically. Small value of R_i increases the probability that the actual transmission rate is above R_i , bringing more unused bandwidth. Finally, the destination of message can be either within a cell or outside a cell, and the outbound messages are first sent to the AP and then forwarded to the final destination, or vice versa. In case of a change in the stream set, bandwidth should be reallocated[1].

3.2 Bandwidth Allocation

This subsection briefly describes the allocation scheme of Lee's work on which this paper is built. For detailed description of bandwidth allocation, refer to [7]. To begin with, by allocation, we mean the procedure of determining capacity vector, $\{H_i\}$, for the given superframe time, F , and message stream set described as $\{S_i(P_i, M_i)\}$. As shown in Fig. 1, H_i denotes the time amount during which

Fig. 1. Polling procedure and capacity vector

S_i can send its message when it is polled. A stream can timely send C_i only if its average transmission rate is over R_i during the P_i .

Let δ denote the total overhead of a superframe including polling latency, IFS, exchange of beacon frame, and the like, while D_{max} the maximum length of a non-real-time data packet. In addition, P_{min} denotes the smallest element of set $\{P_i\}$. Then the requirement for the superframe time, F , can be summarized as in Ineq. (1). Within this range, the scheme can select F and modify some of P_i 's such that they are harmonic[8].

$$\sum H_i + \delta + 2 \cdot D_{max} \leq F \leq P_{min} \quad (1)$$

In addition, the least bound of H_i that can meet the time constraint of S_i is calculated as in Eq. (2).

$$H_i = \frac{M_i}{(\lfloor \frac{P_i}{F} \rfloor - 1)} \quad \text{if } (P_i - \lfloor \frac{P_i}{F} \rfloor \cdot F) \leq D_{max}$$

$$H_i = \frac{M_i}{\lfloor \frac{P_i}{F} \rfloor} \quad \text{Otherwise} \quad (2)$$

The allocation vector calculated by Eq. (2) is a feasible schedule if the vector meets Ineq. (1). Finally, we can determine the length of CFP (T_{CFP}) and that of CP (T_{CP}) as follows:

$$T_{CFP} = \sum H_i + \delta, \quad T_{CP} = F - T_{CFP} \geq D_{max} \quad (3)$$

4 Bandwidth Reclaim Scheme

4.1 Reclaim Test

Hard real-time guarantee is given by the worst case available transmission time which is calculated with a pessimistic assumption that the transmission rate of S_i is just R_i while the size of message is always C_i . So a stream can meet extra slots in some periods if any of above conditions are not met. As a result, a node

Fig. 2. Bandwidth reclaiming

may have no pending message when it receives a poll, in which case it responds with a null frame containing no payload. How to cope with this unused slot is critical to the network throughput. The first step to reclaim the bandwidth is to determine whether to advance the rest of the polling or leave the slot unused.

To begin with, let's assume that if a slot is unused, AP simply moves ahead every subsequent poll. Fig. 2 shows the example in which the predecessors of S_s generate unused slots. Fig. 2(b) illustrates that the unused slots are reclaimed, CFP terminates earlier than scheduled to extend the CP for non-real-time message transmission. However, this method may deprive S_s of one access and the real-time guarantee can be broken. If we let $P_s = k \cdot F + \Delta$, where Δ is a value from 0 to F , then the least bound of network access within P_s is k or $k - 1$, as noted in Eq. (2). (The figure shows the case of $k = 2$). If AP simply advances H_s , S_s loses one scheduled access as shown in Fig. 2(b) provided that the new arrival of message falls in between new and original polling instants. As contrast, that access can be preserved if the AP does not reclaim the unused bandwidth as shown in Fig. 2(a).

The main idea of proposed scheme is that the rest of polling schedule can be advanced if all the subsequent streams have their messages to send, that is, if none of them are waiting for a new message arrival. As the PC can finish the polling schedule of that round earlier than the original CFP duration, CP can be extended to transmit more non-real-time messages. In addition, as the AP receives all the informations on period and transmission time before bandwidth allocation, it can estimate the status of each stream, namely, whether its transmission buffer is empty or not[3]. Finally, in case the slot cannot be reclaimed, it can be used for error control of that stream.

4.2 Runtime Operation

Polling order is important not only in deciding whether a stream will be affected by a deferred beacon but also in improving the probability of reclaim. The more a stream generates unused slots, it would be better to put the stream in the latter place, as small number of successor increases the probability of being reclaimed. How much a stream generates unused slot depends on the amount of overallocated bandwidth. This amount consists of static and dynamic factors. The static factor does not change during the whole life time of a stream, calculated by subtracting the actual bandwidth requirement, $\frac{C_i}{P_i}$, from allocated bandwidth, $\frac{H_i}{F}$. On the other hand, the dynamic factor keeps changing period by period, as it is decided by the current transmission rate. As a result, the overallocation, O_i , is calculated as in Eq. (4).

$$O_i = (\frac{H_i}{F} - \frac{\bar{C}_i}{P_i}) + \max\{0, (1 - \frac{A_i}{R_i}) \cdot M_i\} \quad (4)$$

where A_i is the actual transmission rate S_i is now experiencing for this period. With this information, the polling order should be decided such that the larger O_i , the latter the stream is polled. The order is rearranged for each beginning of a super-frame, taking into account the current transmission rates of respective streams.

Finally, the stream which has higher error rate brings more unused slots, so it seems better to place such a stream to the latter part. However, the error dynamics, conforming to Guilbert model, are so unpredictable that the average behavior cannot provide meaningful criteria[1]. If we are to consider the error characteristics, the channel probing mechanism should be reinforced to the reclaim scheme.

5 Performance Measurements

This section measures the performance of the proposed reclaim scheme via simulation using NS-2 event scheduler[9]. The experiments focus on measuring the achievable throughput to demonstrate the effectiveness of reclaiming scheme. We define *achievable throughput* as the virtual throughput for a given stream set without any collision even in CP. This can be estimated as the sum of both utilization of real-time message streams and ratio of average length of CP to F .

Fig. 3. Bandwidth reclaiming

Fig. 4. Bandwidth reclaiming

Fig. 3 plots achievable bandwidth according to the average number of streams on the superframe to evaluate the performance of reclaiming scheme. Without overallocation caused by the hard real-time guarantee, the only waste is polling overhead, but overallocation makes the throughput much less than ideal. However, the resource reclaiming scheme can narrow the gap between those two curves, that is, considerably relieves the problem of poor utilization of PCF operation, as shown in Fig. 4. The amount of overallocation does not depend on the number of streams but how much F is harmonic with each P_i . For the experiment, 200 stream sets are generated for each number of streams ranging from 2 to 20 with utilization between 0.64 and 0.65. At last, it is certain that the improvement increases when the number of streams is small, and the 52.3 % of waste was recovered. As shown in the figure, the improvement gets smaller as the number of streams increases. This is due to the fact that the reclaimed portion gets smaller and the probability of reclaim decreases.

Fig. 4 plots the reclaimed throughput measured by changing the utilization of stream set from 0.5 to 0.8, while the number of streams randomly distributes

from 2 to 10. If only a stream set has a feasible schedule, the throughput goes high as utilization increases. On the contrary, reclaimed scheme provides stable throughput throughout the given utilization range. When the utilization is from 0.5 to 0.65, about 31.3 % of bandwidth waste was reclaimed.

6 Conclusion and Future Work

In this paper, we have proposed and analyzed a bandwidth reclaim scheme that can overcome poor utilization problem of PCF for real-time communication in WLAN. When an unused slot occurs, AP tests whether it can be reclaimed by checking all of its successors have messages to send. This test confirms that the other time-sensitive traffics are not affected by the early termination of polling round. The reclaimed bandwidth is reassigned to CP to improve the response time of connection management, error control, and other non-real-time messages. The simulation results show that the proposed scheme is able to reclaim up to 52.3 % of bandwidth waste when the number of streams is 2 and that it also provides stable throughput throughout the utilization from 0.5 to 0.8.

Finally, we are to apply the bandwidth reclaim scheme proposed in this paper to the EDF style polling framework. In addition, we are also planning to develop a bandwidth reclaim scheme combined with an error control mechanism.

References

1. Choi, S., Shin, K.: A unified wireless LAN architecture for real-time and non-real-time communication services. *IEEE/ACM Trans. on Networking* (2000) 44-59
2. IEEE 802.11-1999: Part 11 - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (1999) also available at <http://standards.ieee.org/getieee802>
3. Caccamo, M., Zhang, L., Sha, L., Buttazzo, G.: An implicit prioritized access protocol for wireless sensor networks. *Proc. IEEE Real-Time Systems Symposium* (2002)
4. Sheu, S., Sheu, T.: A bandwidth allocation/sharing/extension protocol for multimedia over IEEE 802.11 ad hoc wireless LANS. *IEEE Journal on Selected Areas in Communications*, Vol. 19 (2001) 2065-2080
5. Khattab, A. and Elsayed, K.: Channel-quality dependent earliest deadline due fair scheduling schemes for wireless multimedia networks. *Proc. of MSWIM*. (2004) 31-38
6. Adamou, M., Khanna, S., Lee, I., Shin, I., Zhou, S.: Fair real-time traffic scheduling over a wireless LAN. *Proc. IEEE Real-Time Systems Symposium* (2001) 279-288
7. Lee, J., Kang, M., Jin, Y., Kim, H., Kim, J.: An efficient bandwidth management scheme for a hard real-time fuzzy control system based on the wireless LAN. *Lecture Notes in Artificial Intelligence*, Vol. 3642. Springer-Verlag, Berlin Heidelberg New York (2005) 644-659
8. Carley, T., Ba, M., Barua, R., Stewart, D.: Contention-free periodic message scheduler medium access control in wireless sensor/actuator networks. *Proc. IEEE Real-Time Systems Symposium* (2003) 298-307
9. Fall, K., Varadhan, K.: Ns notes and documentation. Technical Report. VINT project. UC-Berkeley and LBNL (1997)