ACM Mobile Networks and Applications (MONET) August 29, 1997

A Cellular Wireless Local Area Network with QoS Guarantees for Heterogeneous Traffic*

SUNGHYUN CHOI AND KANG G. SHIN
Real-Time Computing Laboratory
Department of Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, Michigan 48109–2122, USA
TEL: (313)647-6977 FAX: (313)763-4617
E-mail: {shchoi,kgshin}@eecs.umich.edu

A wireless local area network (WLAN) or a *cell* with quality-of-service (QoS) guarantees for various types of traffic is considered. A centralized (i.e., star) network is adopted as the topology of a cell which consists of a base station and a number of mobile clients. Dynamic Time Division Duplexed (TDD) transmission is used, and hence, the same frequency channel is time-shared for downlink and uplink transmissions under the dynamic control of the base station. We divide traffic into two classes: class I (real-time) and II (non-real-time). Whenever there is no eligible class-I traffic for transmission, class-II traffic which requires no delay-bound guarantees is transmitted, while uplink transmissions are controlled with a reservation scheme. Class-I traffic which requires a bounded delay and guaranteed throughput is handled with the framing strategy [1] which consists of a smoothness traffic model and the stop-and-go queueing scheme. We also establish the admission test for adding new class-I connections. We present a modified framing strategy for class-I voice uplink transmissions which utilizes the wireless link efficiently at the cost of some packet losses. Finally, we present the performance (average delay and throughput) evaluation of the reservation scheme for class-II traffic using both analytical calculations and simulations.

Keywords: Wireless LANs and MAC protocol, QoS guarantees, star topology, dynamic Time Division Duplexing (TDD), stop-and-go queueing, framing strategy, admission tests, reservation scheme.

*A subject of materials in this paper was presented at the *IEEE INFOCOM'97*, Kobe, Japan, April 9–11, 1997. The work reported in this paper was supported in part by the US Department of Transportation under Grant No. DTFH61-93-X-00017. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agency. All correspondence should be directed to Sunghyun Choi or Kang G. Shin at the above address.

1 Introduction

Wireless local area networks (WLANs) are emerging as an attractive alternative or complementary to wired LANs [2, 3], because they allow us to set up and reconfigure LANs easily without incurring the cost of wiring. They are generally characterized as high-speed wireless systems which cover relatively small areas compared to other wireless systems such as cellular, PCS, and mobile data radio systems. It is expected that they will meet the growing demand that mobile clients should have access to the existing high-speed wired networks. As the interest in broadband multimedia communications involving digital audio and video grows, a number of researchers have been looking into ways of providing QoS guarantees in wired point-to-point WANs [4–8] and LANs [9, 10].

In this paper, we consider how to provide QoS guarantees for heterogeneous traffic on a WLAN. The following three types of QoS for each connection are considered: (1) maximum packet delivery delay, measured from its generation (or arrival) at the transmitter to its arrival at the receiver; (2) transmission throughput, defined as the long-term fraction of time the channel carries the connection's traffic; and (3) packet loss probability, which is the percentage of packet losses for the connection. Heterogeneous traffic is categorized into two classes according to the required QoS as shown in Table 1: (1) class-I real-time traffic requires bounded delay and guaranteed throughput, but is usually tolerant of some packet losses with a certain probability. Voice, video, and real-time data belong to this class. (2) Class-II traffic like the conventional data services requires loss-free transmission, but requires no bounded delay nor guaranteed throughput. Class II can be divided further into two subclasses [11]: (a) class II-A which is delay-sensitive like FTP and remote log-in; and (b) class II-B which is delay-tolerant like paging and e-mail. Class II-A is given priority over class II-B.

We adopt a reservation scheme which is similar to the reservation ALOHA [12, 13] or packet reservation multiple access (PRMA) [14] for uplink class-II traffic transmissions. (The reservation scheme proposed in this paper differs from the previous work, but appears similar in the sense of adopting collision-based reservation schemes.) This reservation scheme is a promising multiple access protocol for class-II traffic, as it provides higher throughput and smaller average delay than other collision-based random access protocols like ALOHA. Basically, class-II traffic is transmitted when there is no class-I traffic to be transmitted since it has lower priority than class I. The reservation scheme is also used for mobile clients to request new class-I connections.

The framing strategy [1, 5], which was originally proposed as a framework for congestion management in integrated service packet networks, is, with some modifications, used to provide guaranteed delay bound and guaranteed throughput for class I traffic. The

	Class I	Class II
Name	Real-time	Non-real-time
Examples	Voice & video	Data services
Delay	Bounded	Unbounded
Throughput	Guaranteed	Not guaranteed
Loss	Loss-tolerant	Zero-loss

Table 1: The classification of traffic

framing strategy is composed of a smoothness traffic model and stop-and-go queueing, and provides packet-delay bound and guaranteed transmission throughput. Each class-I connection should follow a smoothness traffic model, and each new connection needs to pass *a priori* the admission test with a traffic model, implying that the framing strategy reserve slots for class-I connections according to their traffic model. To implement the framing strategy, it is necessary to schedule uplink and downlink transmissions. For this purpose, slots and control mini-slots are made to alternate. We also present a modified framing strategy for class-I voice uplink transmissions, where some packet losses do not affect the voice quality much. We define the traffic model of speech, and describe a scheme that utilizes the wireless link efficiently by transmitting class-II packets while the voice transmitter is in "silent" mode.

The paper is organized as follows. Section 2 shows the specifications and assumptions of the WLAN under consideration. Section 3 describes the proposed protocol, including the reservation scheme for class-II traffic. Section 4 considers the framing strategy with QoS guarantees for class I and defines the admission test for establishing a new connection. As part of the framing strategy, a smoothness traffic model and stop-and-go queueing are described there. In Section 5, we present a modified framing strategy for voice transmissions, which achieves efficient link utilization at the cost of some packet losses. Section 6 presents the analysis and simulation results of the average delay and throughput of the reservation scheme for class II. Finally, the paper concludes with Section 7.

2 System Specifications and Assumptions

As shown in Figure 1, the WLAN under consideration consists of a base station (denoted by B) and several mobile clients (denoted by numbers) forming a star network, called a *cell*. The base station is connected to a wired high-speed network (e.g., ATM LAN) via a wired link. In this topology, the uplink (mobile-to-base) is not a broadcast channel while the downlink (base-to-mobile) is. Hence, mobile clients are not able to listen directly to other mobiles using the same frequency channel. This assumed situation can occur in real world due to the existence of *hidden* terminals [15]. As shown in Figure 2, the transmission ranges of mobile 1 and 2 do not allow them to hear each other, but can both be heard by the base station in between. Mobile 1 and 2 are hidden terminals to each other.

The entire wireless network may consist of several cells, and mobile clients may move from one cell to another. However, we will in this paper focus on the communication within a single cell, hence the uplink and downlink transmissions only. Since wireless links usually have much less bandwidth than the wired counterpart, the former might become a bottleneck. Since the base stations are usually connected to a wired network, the communicating party of each mobile in a cell can be a node in the wired network, or a mobile in another cell, or another mobile in the same cell. In any case, the wireless link in the cell is considered as the end-most link (for downlink) or the front-most link (for uplink) of the entire multi-hop communication as shown in Figure 3. Note that the downlink traffic comes from the wired network or mobiles in the same cell, and the uplink traffic is generated by local mobiles.

Dynamic time division duplexed (TDD) transmission is used in the network, and hence, the base station multiplexes the uplink and downlink packet transmissions dynamically according to the traffic load in a frequency channel. We could instead use frequency division duplexed (FDD) transmission, in which two different frequency channels are allocated for uplink and downlink transmissions, or static TDD in which a portion (usually a half) of each time frame is allocated for the uplink and the remaining portion is allocated for downlink transmissions. FDD, as in AMPS (FDMA), IS-54 (TDMA), and IS-95 (CDMA), is the common duplexing mode in cellular systems, and static TDD was adopted in the DECT system [2]. But, dynamic TDD allows for more efficient link utilization in the case of unbalanced uplink and downlink traffic, e.g., non-interactive data transmissions as shown in [16]. We assume all packets, like ATM cells, to be of the same fixed size. Throughout the remainder of this paper, we will ignore the packet-propagation delay, because it is usually small relative to the other delay components like queuing and transmission delays in a cell. (A cell in this paper refers to a *micro-cell*, which has coverage of a few hundred meters, or a *pico-cell*, which covers small indoor areas [2]).

Since the wireless channel is inherently unreliable (due to noises, interferences, and multipath fadings), we need a special means to ensure the error-free delivery of packets through each wireless link. Usually, a combined channel coding and diversity scheme [17] is used to meet this need. To handle various types of traffic in our system, we can apply error-handling schemes adaptively. We adopt an ARQ (Automatic Retransmission reQuest) scheme for class II to ensure virtually error-free transmission of data. But, it is ruled out for class I because of its difficulty in making delivery-delay guarantees. We use an FEC (Forward Error Correction) scheme for class I, instead.[1] To this end, the receiver is equipped with a dual-mode channel decoder (i.e., both error correction and detection modes). The transceiver pair works as follows, assuming use of a specific channel code: (1) at the transmitter, a packet is channel-encoded, and then transmitted; and (2) at the receiver, the received packet is decoded by an error-correction decoder (if class I) or an error-detection decoder (if class II) as shown in Figure 4. The dual-mode receiver is expected to work well since using a channel code, the decoder can detect more errors than those correctable [18]. We will not consider error-combating techniques any more since they are not within the primary scope of this paper, but we assume that a packet is received correctly unless that packet collides with one or more concurrent packets.

3 Protocol Description

When a mobile client wants to send a packet, regardless whether it is destined for another client in the same cell or in a remote cell, the client must send the packet to its base station first, which will then forward the packet to the final destination, sometimes via other base stations. Dynamic TDD transmission is used in the network under consideration, i.e., a wireless channel is time-shared for both downlink and uplink transmissions under the dynamic control of the base station. Based on the star topology, only the downlink

[1]Although combined FEC and diversity seems to be the only way for error-protection of class I, it is extremely difficult to guarantee the virtually error-free transmission of packets of these classes over the wireless link due to the limited error-correcting capability of the underlying FEC scheme — the more capability, the more redundancy needed. So the proposed scheme might not be applicable for reliability-critical real-time data traffic of class I.

channel is assumed to be the broadcast type. Thus, when the base station transmits packets, all but their destination mobiles in the cell ignore them. By contrast, a mobile cannot hear other mobiles' uplink transmissions, and only the base station can determine if a collision has occurred in the uplink channel.

We adopt two different strategies for class I and II. First, class-I traffic is transmitted via connections, i.e., for each class-I (downlink or uplink) connection, a finite number of slots are reserved to meet its required QoS. Each connection (between the base station and a mobile client) is identified by (1) which client the connection is for and (2) whether the connection is for downlink or uplink transmissions. For QoS provision, class-I traffic has priority over class-II traffic, where the transmission is controlled by the framing strategy (to be discussed in the next section). Class-II traffic doesn't need the concept of connection, but if there is a pending message (which consists of a number of fixed-size packets), it will be transmitted when there are available slots, i.e., when there is no class-I traffic to be transmitted over the link. For uplink class-II traffic, a request of slot reservation is made for the transmission of each message.

A slot and a control mini-slot alternate continually as shown in Figure 5. In a slot of duration T_s, a (fixed-size) packet is transmitted. By dynamic TDD transmission, each slot can be used for either downlink or uplink transmission under the control of the base station. A control mini-slot of duration T_{ms} is used to transmit a control packet. Control packets are used by the base station to announce to the mobiles information on the next slot like (1) for the downlink or uplink, (2) for class I or II, and (3) for which client. These regularly alternating slots and mini-slots are expected to help each mobile client synchronize to the global transmission system. We will henceforth use T_{ms} as a basic time unit. Assume that a slot duration is an even number multiple of a mini-slot duration, i.e., $K = T_s/T_{ms}$ is an even number.

There exists a slot, called the *reservation slot*, which is used for requesting an uplink class-I connection establishment or an uplink class-II message transmission by mobiles. A reservation slot consists of K mini-slots of duration T_{ms}. It is divided into two parts: (1) the first half is a set of $K/2$ reservation mini-slots used by mobile clients to request uplink transmissions; and (2) the second half is a set of $K/2$ result-announcing mini-slots for each of the corresponding previous reservation mini-slots. The reservation mini-slots are accessed by a slotted ALOHA-like random access protocol: when a reservation slot is issued by the base station, each mobile client with a pending request chooses one of $K/2$ mini-slots randomly, and then sends the request in that chosen mini-slot with such traffic information as the class of traffic it wants to transmit. If class-I traffic is requested, the mobile client needs to send the traffic characteristics (defined in the next section) as well. The result of each of $K/2$ mini-slots is success (of which mobile[2]), or collision, or empty/unused. Using each of the next $K/2$ downlink mini-slots, the result of the corresponding reservation mini-slot is announced.

If a reservation slot contains only "collided" reservation mini-slots, the base station will issue reservation slots consecutively until a successful reservation mini-slot appears in a reservation slot. Using this policy, the base station will obtain at least one successful reservation for mobile clients who want to make a slot reservation. A mobile whose reservation request collided with others will retransmit the request again in the subsequent

[2] Due to the *capture effects* [3], a reservation request can be transmitted successfully even in the presence of concurrent reservation requests from other mobiles.

reservation slots with the probability q_{ret} until it is successful. (The retransmission probability q_{ret} can be determined adaptively according to the results of all of $K/2$ reservation mini-slots.) If a reservation request is successful, the base station will be informed that the mobile client who made the request wants (1) to send a pending class-II message if the request was for class II, or (2) to establish a class-I connection if it was for class I.

For dynamic TDD transmission, the base station needs to multiplex between downlink and uplink transmissions. The base station does not know if a mobile has a pending message without receiving a reservation request. Basically, a reservation slot is issued after completing the transmission of a (downlink or uplink) class-II message as shown in Figure 6, where only class-II traffic (without indicating subclasses) is shown for a better understanding of how it works. Two first-in-first-out (FIFO) class-II service queues are implemented at the base station for two subclasses of class II, in which both the uplink transmission requests (from mobile clients) and the downlink messages are queued together. The contents of the queues are updated at the end of every reservation slot (marked with arrows in Figure 6): at the end of the k-th reservation slot, E_k, all the downlink messages which arrived at the base station between E_{k-1} and E_k, and all the uplink transmission requests which were successfully received during the k-th reservation slot are queued in a random order. By this random queueing policy, the uplink transmission achieves fairness since the uplink reservation requests might suffer excessive delays compared to the downlink transmission due to the collision-based reservation request access. The second queue for class II-B can be served whenever the first queue for class II-A is empty. When both queues are empty, the base station issues a reservation slot consecutively for each available slot.

With the scheme explained above, the maximum achievable class-II throughput might be very low depending on the average message length, i.e., the smaller the average message length, the smaller maximum achievable throughput. To solve this problem, we assign the Minimum Next Reservation Slot Length (*MNRSL*) defined as the minimum number of slots between two consecutive reservation slots. Assume that a (downlink or uplink) message transmission was completed after a reservation slot. If less than *MNRSL* slots were issued since the reservation slot, the base station will serve more messages until the total number of slots issued exceeds *MNRSL*. Thus, the maximum achievable throughput is guaranteed to be greater than or equal to $(MNRSL \cdot T_s)/((MNRSL + 1)(T_s + T_{ms}))$.

4 Framing Strategy for QoS Provision

In this section, we describe the framing strategy to guarantee QoS for class-I traffic. The time axis is divided into frames, each of which is composed of a finite number of slots (and so mini-slots) as shown in Figure 5. If there are N slots and N mini-slots in a frame of time duration T, then $T = N \cdot (T_s + T_{ms})$.

4.1 Traffic Model

For each connection i of class-I traffic, we adopt the (M_i, T_i)-smooth model. During each frame of length T_i, no more than M_i packets arrive for connection i. If connection i is for uplink transmissions, the mobile regulates its uplink traffic to follow the (M_i, T_i)-smooth model using a packet admission policy; under this policy, any packet that violates the smoothness is assumed not to be generated until the beginning of the next frame. In the wired part of the network, we assume the existence of a traffic regulator like the (M_i, T_i)-smooth admission policy and leaky-bucket [6] in the source end nodes and a flow/congestion control or packet scheduling scheme such as the framing strategy and Weighted Fair Queueing (or PGPS) [6] in intermediate nodes. Thus, the traffic arriving at the base station from the wired network will have the smoothness property, which can then be converted to the (M_j, T_j)-smooth model. Moreover, the downlink traffic originated from a mobile within the cell for intra-cell communications will also have the the smoothness property (to be explained later). So, it is possible to adopt the (M_j, T_j)-smooth model for a downlink connection j as well.

Suppose there are G frame sizes, $T_1^f, T_2^f, \cdots, T_G^f$, and each frame size is a multiple of next smaller frame size, i.e.,

$$T_g^f = I_g \cdot T_{g+1}^f, \quad g = 1, 2, \cdots, G-1, \tag{1}$$

for some integer I_g. For all g,

$$T_g^f = K_g \cdot (T_s + T_{ms}), \tag{2}$$

for some integer K_g, i.e., there are a finite number of slots and mini-slots in each frame. Each frame of duration T_g^f is called a *type-g frame*. For each connection i, $T_i = T_g^f$ for some g, and the connection is called the *type-g connection*. Figure 7 shows the case of $G = 2$ and $T_1 = 3T_2$. Note that all frames are incorporated into a single frequency channel. As shown in the next subsection, the packets in a type-g connection will be guaranteed to have a delivery-delay bound $2T_g^f$, implying the existence of G delivery-delay bounds.

4.2 Stop-and-Go Queueing

Downlink Transmissions: The communication from the base station to mobile clients can be viewed as taking place over a single wireline link since it is broadcast-type communication. The base station is equipped with G FIFO class-I service queues, one for each type of connections. Stop-and-go queueing is used for downlink transmissions as follows.

- **Rule (a)** A downlink packet of a type-g connection that has arrived at the base station during a frame is not *eligible* for transmission until the beginning of the next frame (Figure 8). Eligible packets of certain type connections are transmitted during the corresponding frame.
- **Rule (b)** Any eligible packet of a type-g connection, $g = 2, 3, \cdots, G$, has priority over eligible packets of type $g' < g$.
- **Rule (c)** The wireless link should not be left idle whenever there are eligible packets in the class-I service queues.

Uplink Transmissions: During each type-g frame, the base station issues (via mini-slots) up to $\sum_{\{T_i = T_g^f\}} M_i$ slots for type-g uplink connections with the same priority given in **Rule (b)**. Within the same type, uplink connections are given priority over downlink connections, so in each frame, uplink slots are issued first, and then eligible downlink packets are transmitted. For uplink connection i, the base station will issue up to M_i

uplink slots or until the corresponding mobile completes the transmission of all its eligible packets. Then, the next connection in the priority list is served. When a slot is issued for connection i in a frame, the connection-i mobile transmits a packet which was generated during the previous frame, (i.e., eligible), and marks the last packet arrived in the previous frame. Now, the traffic arrived at the base station from each mobile client also has the (M_i, T_i)-smooth property.

Using the above transmission rules, all of the connection-i packets, which conform to the smoothness, will be transmitted before the end of the next type-g frame (when $T_i = T_g^f$), thus guaranteeing their transmission within a delay of $2T_i$. For each connection i, up to M_i packets can be transmitted during a frame of duration T_i, and hence, it is guaranteed to have the throughput of $M_i \cdot T_s/T_i$.

Example 1
We ignore the effects of mini-slots assuming that $T_{ms} \ll T_s$. Suppose there are three class-I connections, where connection 1 and 3 are for uplink communications and connection 2 is for downlink communications with the smoothness parameters $\{(M_i, T_i)\} = \{(2, 4T_s), (1, 4T_s), (2, 8T_s)\}$. So, there are two frame sizes, (i.e., $G = 2$), where $T_1^f = 8T_s$ and $T_2^f = 4T_s$. Figure 9 shows an example of stop-and-go queueing for this set of connections. The slots for connection 1, 2, and 3 are marked with $U1$, $D2$, and $U3$, respectively. The packet arrivals of connection 2 are also marked by arrows. One can see that connection 1 has priority over connection 2 even though they are of the same frame size since connection 1 is uplink. Connection 3 has the lowest priority. We observe that the slot at time T_1^f appears to be empty because even though the base station issued that slot for connection 1, the mobile of connection 1 didn't have any packet to transmit at that time. The base station doesn't issue the next slot for connection 1, and transmit a downlink packet from connection 2 instead. Using the shaded slots, which are left unused by class-I connections, in the figure, class-II traffic, if any, will be transmitted according to the rules presented in Section 3. □

4.3 Admission Tests

If a new connection is to be added, it has to pass the following simple admission test depending on the frame-size constraints. Downlink connection requests come from the wired network (or from a mobile requesting a connection within the same cell) with the traffic characteristics (M_i, T_i), while uplink connection requests come from mobile clients via a reservation slot. The admission test is given as

$$\sum_{g=1}^{G} M_g^f \cdot (T_s + T_{ms})/T_g^f \leq 1, \tag{3}$$

where $M_g^f = \sum_{\{T_i = T_g^f\}} M_i$ is the number of the reserved slots within a type-g frame, which includes both the existing connections and the newly-requested connection. If the results of the above test is positive, it is possible to provide the required QoS to the new connection without compromising the existing connections' guarantees, and hence, the base station starts to serve the new connection at the next frame. The basic idea of the admission test is that the total reserved throughput for class-I connections (for both slots and mini-slots) should be less than, or equal to, 1. The readers are referred to

Appendix A for a formal proof. Note that for Example 1, the summation in Eq. (3) is exactly 1, implying that all of the slots be reserved for the three class-I connections. It might sometimes be desirable to set aside a certain portion of throughput for class-II traffic (say S). In such a case, Eq. (3) should be modified by replacing 1 with $1-S(T_s+T_{ms})/T_s$.

5 Modified Framing Strategy for Uplink Voice Transmissions

In this section, we modify the framing strategy for uplink class-I voice transmissions. Speech is encoded by the 8-bit PCM[3] at 8 kHz, and so, the bit rate is 64 kb/s [19]. Assume that the target delay bound is $2T_g$ for some g for a voice connection i. Since the frame size is T_g, up to N_f $(= \lfloor T_g/T_{sample} \rfloor)$ samples can be transmitted during a frame, where the sampling period $T_{sample} = 1/(8 \times 10^3) = 1.25$ msec. If a packet can accommodate N_s samples, $\lceil N_f/N_s \rceil$ packets are necessary during a frame to transmit the voice traffic within the delay bound, so $(M_i, T_i) = (\lceil N_f/N_s \rceil, T_g)$ traffic model is adopted for connection i.

It is well-known that voice traffic is modelled by an "on-off" model since a speech signal is in either *talking* or *silent* mode. That is, a voice client generates bursts of packets, corresponding to *talk-spurts*, while in talking mode, and no packets while in silent mode. From the (M_i, T_i) smoothness model, in each frame of duration T_i, there may or may not be packets to be transmitted. With this traffic model, we can use the above framing strategy for downlink, but it will not be efficient for uplink since the base station will issue at least one slot every T_i, and while in silent mode, there will be no packet transmission in that slot. We modify the previous framing strategy for uplink voice transmissions of connection i so that class-II traffic may be transmitted while in silent mode.

The admission test given in Eq. (3) is also used for a new voice connection. Assuming that connection i has passed the admission test, the base station issues at least one slot for connection i in each frame, and the corresponding mobile transmits voice packets via those slots. If an issued slot is empty within a frame, the base station assumes that the mobile is in silent mode, and will stay there for a while. To utilize the link more efficiently, the base station uses M_i slots in each frame for class-II traffic beginning at the next frame after an empty slot, i.e., by the scheme in Section 3, up to M_i downlink packets of class-II are transmitted or uplink slots are issued for class-II traffic in each frame, while announcing that these slots were originally for connection i. While the mobile stays in silent mode, class-II traffic transmissions will be successful. However, after this process, the connection-i mobile will transmit a packet in the issued slot when the mobile returns to talking mode, and then there will be a collision between the voice packet and a class-II packet. Detecting this collision will lead the base station to assume that the mobile has now returned to talking mode. The collided class-II packet should be retransmitted later. The base station also tries to issue another slot in the same frame for the collided connection-i packet. If all slots are reserved for class-I connections (i.e., the equality holds

[3]This voice coding is used in the current wired telephone communications. In the existing wireless communications, other voice coding schemes like ADPCM, QCELP, and VSELP are used to reduce the bit rate to cope with the wireless bandwidth limitation [2]. However, in this paper, we adopt a simple PCM scheme to show how the framing strategy can be modified to accommodate voice communication more efficiently.

in Eq. (3)), it will be impossible to issue another slot in the same frame. But, this is not a serious problem since voice traffic is tolerable of occasional packet losses. After the frame in which the collision occurred, the base station starts to issue M_i slots in each frame exclusively for connection i.

6 Performance Analysis of Class-II Communications

This section analyzes the performance of the reservation scheme for class-II communications, where smaller average delay and larger throughput are desirable. Here, all traffic belongs to class II, where class II-A and II-B are not differentiated for simplicity, and hence, only one base station service queue is implemented. As described in Section 3, when there exists class-I traffic, the reservation scheme considered here is activated whenever there are no eligible class-I packets to be transmitted.

For uplink accesses using the reservation scheme, we use the model of K_u clients with the following assumptions.

- **A1.** Message length has a geometric distribution with parameter p_l measured in the number of packets.
- **A2.** Downlink messages arrive from the wired network according to a Poisson process with the overall arrival rate λ_d (messages/mini-slot).
- **A3.** Messages are generated at each of the K_u clients according to independent Poisson processes with the generation rate λ_u/K_u (messages/mini-slot).
- **A4.** Each collided request must be retransmitted in a later reservation mini-slot until the request is successfully received.
- **A5.** Closed-loop behavior of clients, i.e., backlogged clients will discard newly-generated messages until the successful transmission of the request.

A client is said to be *backlogged* when it was notified by the base station to have a collided request and hence must retransmit it. Note that from **A**1, we consider the inter-cell communications only since downlink messages are assumed to arrive exclusively from the wired network. We also make the following simplifications of the scheme to facilitate the derivation.

- **S1.** Even if a reservation slot contains only collided reservation mini-slots, the base station will not issue another reservation slot.
- **S2.** *MNRSL* will be set to 1.
- **S3.** The retransmission probability q_r will be assigned a fixed constant.

By **S**1 and **S**2, a message transmission and a reservation slot will alternate continuously when the base station service queue is not empty.

6.1 Markov Chain Modelling

The pair $(M(k), N(k))$ is modelled by a 2-dimensional Markov chain, where $M(k)$ is the number of the backlogged clients requesting uplink message transmission and $N(k)$ is the number of the downlink messages or uplink requests in the base station service queue at the end of the k-th reservation slot. Figure 6 shows a timing diagram of class-II communications with the state transition moments and the change of state $N(k)$. Each of the $M(k)$ backlogged clients will transmit a request in the $(k+1)$-th reservation slot, independently of each other, with probability q_r. Each of the $K_u - M(k)$ other clients will transmit a request in the $(k+1)$-th reservation slot if one (or more) such messages are generated since the last reservation slot. $T_{nr}(k)$ $(= L_{nr}(k)(T_{ms} + T_s))$ is the time period from the k-th reservation slot to the $(k+1)$-th reservation slot, and so $L_{nr}(k)$ is the number of corresponding slots including the reservation slot during $T_{nr}(k)$ with the following conditional distribution given $N(k) = n$:

$$q_l(l|n) = \begin{cases} 1, & \text{if } l = 1, n = 0, \\ 0, & \text{if } l \neq 1, n = 0, \\ 0, & \text{if } l < 2, n \neq 0, \\ p_l(1-p_l)^{l-2}, & \text{if } l \geq 2, n \neq 0. \end{cases} \tag{4}$$

The distribution of the number of the downlink message arrivals, $N_a(k)$, from the end of the $(k-1)$-th reservation slot to the end of the k-th reservation slot given $L_{nr}(k-1) = l$ is

$$q_a^l(i) = e^{-\lambda_d l(T_{ms}+T_s)} \frac{(\lambda_d l(T_{ms} + T_s))^i}{i!}. \tag{5}$$

The probability that a non-backlogged client requests in the k-th reservation slot given $L_{nr}(k-1) = l$ is

$$q_g^l = 1 - e^{-\lambda_u l(T_{ms}+T_s)/K_u}. \tag{6}$$

Let $Q_g^l(i,m)$ be the probability that i out of $K_u - m$ non-backlogged clients transmit requests in the k-th reservation slot, and let $Q_r(i,m)$ be the probability that i of m backlogged clients transmit requests given $M(k-1) = m$ and $L_{nr}(k-1) = l$, then

$$\begin{aligned} Q_g^l(i,m) &= \binom{K_u - m}{i} (1 - q_g^l)^{K_u-m-i} (q_g^l)^i, \\ Q_r(i,m) &= \binom{m}{i} (1 - q_r)^{m-i} (q_r)^i. \end{aligned} \tag{7}$$

Now, $N_r(k)+N_g(k)$ clients will transmit requests in the k-th reservation slot. Accordingly, we obtain the following state transition relationship:

$$\begin{aligned} N(k) &= N(k-1) + N_s(k) + N_a(k) - T(k-1), \\ M(k) &= M(k-1) + N_g(k) - N_s(k), \end{aligned} \tag{8}$$

where $T(k) = 0$ if $N(k) = 0$ and 1 if $N(k) \geq 1$, and $N_s(k)$ is the number of the successful requests during the k-th reservation slot.

The probability $P_u(\hat{j}, \hat{k}, \hat{l})$ that $\hat{j}$ out of $\hat{k}$ clients succeed in the k-th reservation slot (with $\hat{l}$ reservation request mini-slots) is given by

$$P_u(\hat{j}, \hat{k}, \hat{l}) = \begin{cases} 0, & \text{if } \hat{j} > \hat{l} \text{ or } (\hat{j} = \hat{l} \text{ and } \hat{k} > \hat{l}), \\ \binom{\hat{k}}{\hat{j}} \frac{\hat{l}!}{(\hat{l}-\hat{j})!} A(\hat{k} - \hat{j}, \hat{l} - \hat{j})/\hat{l}^{\hat{k}}, & \text{otherwise}, \end{cases} \tag{9}$$

where $A(k',l')$ is the number of cases such that k' clients requested during one of l' mini-slots, and all of them failed:

$$A(k',l') = \begin{cases} 1, & \text{if } k'=0, \\ 0, & \text{if } k'=1, \\ \sum_{g=1}^{\lfloor k'/2 \rfloor} \binom{l'}{g} \sum_{\mathbf{C}(\mathbf{n})} \binom{k'}{n_1 n_2 \cdots n_g} & \\ \cdot \binom{g}{m_1 m_2 \cdots m_{g'}}, & \text{if } k' \geq 2, \end{cases} \tag{10}$$

where $\binom{k'}{n_1 n_2 \cdots n_g}$ $(= k'!/n_1!n_2!\cdots n_g!)$ is the g-th order multinomial coefficient, and the condition $\mathbf{C}$ of the g-th order vector $\mathbf{n} = \{n_1, n_2, \cdots, n_g\}$ is: (i) $\sum_{i=1}^{g} n_i = k'$; (ii) for all i, $n_i \geq n_{i+1}$; and (iii) for all i, $n_i \geq 2$. The g'-th order vector $\mathbf{m} = \{m_1, m_2, \cdots, m_{g'}\}$, which directly depends on the vector $\mathbf{n}$, satisfies: (i) $\sum_{i=1}^{g'} m_i = g$; (ii) $g' = \max_{i=1}^{g} n_i$; and (iii) m_i is the number of n_j's such that $n_j = i$.

Finally, we can easily derive the state transition probabilities of $M(k)$ and $N(k)$, respectively, given $(M(k), N(k), L_{nr}(k)) = (m, n, l)$:

$$\begin{aligned} P_{m,m+i}(m,n,l) &= \sum_{g=0}^{K_u-m} \sum_{r=0}^{m} Q_g^l(g,m) Q_r(r,m) \\ &\quad \cdot P_u(g-i, g+r, L_{ms}), \end{aligned} \tag{11}$$

for $K_u - m \geq i \geq -L_{ms}+1$ if $m > L_{ms}$ and $K_u - m \geq i \geq -m$ if $m \leq L_{ms}$, where L_{ms} is the number of the reservation mini-slots in a reservation slot, i.e., $L_{ms} = K/2 = T_s/(2T_{ms})$, and

$$\begin{aligned} &P_{n,n+j}(m,n,l) \\ &= \begin{cases} \sum_{a=0}^{j} q_a^l(a) Q_s(j-a|m,l), & \text{if } n=0, \\ \sum_{a=0}^{j+1} q_a^l(a) Q_s(j+1-a|m,l), & \text{if } n>0, \end{cases} \end{aligned} \tag{12}$$

for $j \geq 0$ if $n=0$ and $j \geq -1$ if $n > 0$, where $Q_s(i|m,l)$ is the probability of $N_s(k) = i$ given $M(k-1) = m$ and $L_{nr}(k-1) = l$:

$$Q_s(i|m,l) = \sum_{g=0}^{K_u-m} \sum_{r=0}^{m} P_u(i, r+g, L_{ms}) Q_g^l(g,m) Q_r(r,m). \tag{13}$$

The conditional state transition probability of the 2-dimensional Markov chain $(M(k), N(k))$ given $L_{nr}(k) = l$ is:

$$P_{(m,n),(m+i,n+j)}^l = P_{m,m+i}(m,n,l) P_{n,n+j}(m,n,l). \tag{14}$$

Averaging the effect of the condition $L_{nr}(k)$, we obtain the state transition probability:

$$P_{(m,n),(m+i,n+j)} = \sum_{l=1}^{\infty} P_{(m,n),(m+i,n+j)}^l P_{L_{nr}(k)|N(k)}(l|n). \tag{15}$$

Finally, we can obtain the steady-state probability:

$$\pi_{m,n} = \lim_{k \to \infty} P(M(k)=m, N(k)=n). \tag{16}$$

6.2 Uplink Request Success Rate

The request success rate from the $(k-1)$-th reservation slot to the k-th reservation slot given $N_s(k) = i$ and $L_{nr}(k-1) = l$ is:

$$R_u^s(i,l) = \frac{i}{l(T_{ms}+T_s)}. \tag{17}$$

By averaging $N_s(k)$ and $L_{nr}(k)$, we get the uplink request success rate given $M(k-1) = m$ and $N(k-1) = n$.

$$R'_s(m,n) = \sum_{i=1}^{L_{ms}} \sum_{l=1}^{\infty} R_u^s(i,l) Q_s(i|m,l) q_l(l|n). \tag{18}$$

We define two new continuous-time processes $\hat{M}(t) = M(k)$ and $\hat{N}(t) = N(k)$, if $t \in [E_k, E_{k+1})$, where E_k is the end time of the k-th reservation slot. Note that $\hat{M}(t)$ denotes the number of backlogged clients at time t. We can obtain the steady-state probability of this process as follows:

$$\hat{\pi}_{m,n} = \frac{\pi_{m,n} E[L_{nr}^n]}{\pi_0^{bsq} E[L_{nr}^0] + (1-\pi_0^{bsq}) E[L_{nr}^1]}, \tag{19}$$

where $E[\cdot]$ is the expectation of a random variable, $\pi_n^{bsq} = \lim_{k\to\infty} P(N(k) = n) = \sum_m \pi_{m,n}$, and L_{nr}^n is the number of the slots between two consecutive reservation slots, $L_{nr}(k)$, given $N(k) = n$. It is easily shown to be $E[L_{nr}^0] = 1$ and $E[L_{nr}^1] = 1 + 1/p_l$. For a given time t, if $\hat{M}(t) = n$ and $\hat{N}(t) = m$, then the conditional request success rate is $R'_s(m,n)$. Thus, by averaging this over time, we get the average request success rate:

$$R_u^s = \sum_m \sum_n R'_s(m,n) \hat{\pi}_{m,n}. \tag{20}$$

6.3 Average Request Success Delay

We derive the delay from the generation of a message to a successful request for its transmission. The first term in the delay is the average time V from the message generation to the beginning of next reservation slot. When $N(k) = 0$, $L_{nr}(k) = 1$. Then the generation time of a message — generated in $[B_k, B_{k+1})$ for an arbitrary k — will be uniformly distributed in $[B_k, B_{k+1}]$ [20], where $B_{k+1} - B_k = T_s + T_{ms}$, since messages are generated according to a Poisson process, and so $E[V|N(k) = 0] = (T_s + T_{ms})/2$. When $N(k) > 0$, $L_{nr}(k)$ has a geometric distribution plus one. Since the geometric distribution is memoryless, when a message was generated, $E[V|N(k) > 0]$ is approximated to be $E[L_{nr}^1 - 1](T_s + T_{ms})$. Consequently, we obtain the mean value of V as:

$$\begin{aligned} E[V] \;\approx\; & \frac{E[L_{nr}^0]}{2}(T_s + T_{ms})\hat{\pi}_0^{bsq} + E[L_{nr}^1 - 1] \\ & \cdot (T_s + T_{ms})(1 - \hat{\pi}_0^{bsq}). \end{aligned} \tag{21}$$

Secondly, according to Little's theorem, the average time spent in the backlog is the ratio of the average of backlogged clients to the average message generation rate G_{new} or

$E[\hat{M}]/G_{new}$, where the average of backlogged clients $E[\hat{M}] = \sum_m \sum_n m\hat{\pi}_{m,n}$. Now, the average delay measured is given as:

$$D_u^s = E[V] + T_s + \frac{E[\hat{M}]}{G_{new}} \quad \text{(mini-slots)}, \tag{22}$$

where the first term corresponds to the time to the next reservation slot, the second term to a reservation slot time, and the third term to the average backlog delay. For the whole system to be stable, the average rate of new message generation must equal the average message transmission request success rate, i.e., $G_{new} = R_u^s$. Finally, we get the desired throughput-delay relation under the stable condition:

$$D_u^s = E[V] + T_s + \frac{E[\hat{M}]}{R_u^s} \quad \text{(mini-slots)}. \tag{23}$$

6.4 Throughput Analysis

Due to the existence of control mini-slots and reservation slots, the maximum achievable throughput W_{total}^{max} is less than one, and is dependent on the message length distribution. Assuming that for all k, $N(k) > 0$, a reservation slot and a message transmission will alternate continuously, thus achieving the maximum possible throughput which is given by

$$W_{total}^{max} = \frac{T_s}{T_s + T_{ms}} \frac{E[L_{nr}^1 - 1]}{E[L_{nr}^1]}. \tag{24}$$

Note that the actual total incoming rate (including both uplink and downlink) to the base station service queue is $\lambda_{total} = \lambda_d + R_u^s$. Now, if $\lambda_{total} E[L_{nr}^1 - 1]T_s \leq W_{total}^{max}$, i.e., if the base station service queue is in the stable condition, the downlink throughput W_d and uplink throughput W_u would be

$$\begin{aligned} W_d &= \lambda_d E[L_{nr}^1 - 1]T_s, \\ W_u &= R_u^s E[L_{nr}^1 - 1]T_s. \end{aligned} \tag{25}$$

6.5 Average Delay

First of all, we need the queueing delay in the base station service queue, i.e., the average delay from the entrance of a downlink message or an uplink request into the service queue to the start of its transmission. We first obtain the average number of the queued downlink messages or uplink requests in the base station service queue which is given by

$$E[\hat{N} - 1 | \hat{N} > 1] = \sum_n (n-1)\hat{\pi}_n^{bsq}, \tag{26}$$

because $\hat{N}(t) - 1$ corresponds to the number of downlink messages or queued requests in the base station service queue for $[E_k, E_{k+1})$. Now, the queueing delay is given using Little's theorem:

$$D_q = E[\hat{N} - 1 | \hat{N} > 1]/G'_{new}, \tag{27}$$

where $G'_{new} = \lambda_{total} = \lambda_d + R_u^s$ for the system to be stable. Now, the downlink delay is given by

$$D_d = E[L_{nr}^1 - 1](T_s + T_{ms}) + E[V] + D_q, \tag{28}$$

where $E[V]$ is the average time from a downlink message arrival to the end of the next reservation slot, which is approximated to the value from Eq. (21), and the uplink delay is given by

$$D_u = E[L_{nr}^1 - 1](T_s + T_{ms}) + D_u^s + D_q. \tag{29}$$

In both equations, the first terms stand for the message transmission delays, the second terms for the delays from the arrival/generation of a message to the entrance into the base station service queue, and the third for the queueing delays in the service queue.

6.6 Numerical and Simulation Results

We show some analytical calculation results using the equations above, and compare them with the simulation results. For the simulations, we generated Poisson traffic, and followed the assumptions given at the beginning of this section. The results are based on $p_l = 0.1$, $q_r = 1.$, $K_u = 5$, $K = T_s/T_{ms} = 10$ (and so $L_{ms} = 5$).

Figure 10 (a) plots the uplink delays D_u as λ_u increases for three different λ_d values, while Figure 10 (b) plots the downlink delays D_d as λ_d increases. We observe that the numeric calculations (with mark *calc*) and the simulations (with mark *sim1*) are very close to each other for the same parameters. Note that we did rarely use approximations for our analysis except for in Eq. (21). In both cases, delays are almost constant for small rates, but monotonically increase, and then go to infinity as the actual total incoming rate λ_{total} $(= \lambda_d + R_u^s)$ goes to $W_{total}^{max}/(E[L_{nr}^1 - 1]T_s)$ ($\approx 8.26e^{-3}$ in the results). Due to the closed-loop behavior of the clients, $R_u^s \leq \lambda_u$. Hence, in the figures, the marginal rates (at which delays become infinite) appear larger for uplink under the same parameters. Note that the uplink delays are larger than the downlink delays by as much as $T_s + E[\hat{M}]/R_u^s$ under the same condition from Eqs. (28) and (28).

Figure 11 compares the simplified protocol with the simplifications **S**1 and **S**2 (marked with *sim1*) and the actual protocol without **S**1 and **S**2 (marked with *sim2*) using the simulation results of the delay-versus-rate relationship. For the actual protocol, $MNRSL = 10$ was used. In both graphs, we observe that delays are smaller for the actual protocol, especially for large rates, since the messages can be transmitted consecutively without the appearance of a reservation if the first message has less than *MNRSL* packets. Consequently, the marginal rates are larger for the actual protocol, implying that the maximum achievable throughput be larger for the actual protocol.

7 Concluding Remarks

In this paper, we have considered a WLAN providing QoS guarantees for heterogeneous traffic in a cell. According to the required QoS, traffic is categorized into class I (real-time) and class II (non-real-time). Class-II traffic is divided into two subclasses according to whether the traffic is delay-sensitive or not. The protocol is based on the framing strategy for class I and a reservation scheme for class II, where class I has priority over class II. When each class-I connection follows a smoothness model, it was shown to be possible to guarantee the delay bound and throughput using the stop-and-go queueing. The admission test for a new class-I connection was also defined. A modified scheme for voice transmissions was presented for efficient link utilization at the cost of some packet losses. When there is no eligible class-I traffic, class-II traffic is transmitted. Uplink class-II transmission reservation and uplink class-I connection establishment were requested using the reservation scheme. We finally analyzed the average delay and throughput of the reservation scheme for class-II traffic, and presented the numerical calculation and simulation results.

A Proof of Eq. (3)

If all (including the new connection requested) of the required slots for class-I connections can be reserved, the new connection can be established. Since M_i slots need to be reserved for connection i of type g, within a type-g frame, a total of M_g^f $(= \sum_{\{T_i = T_g^f\}} M_i)$ slots should be reserved. Note that T_1^f is the least common multiple (LCM) of $\{T_1^f, T_2^f, \cdots, T_G^f\}$. Since each frame repeats itself, it is enough to consider only one type-1 frame of duration T_1^f. There are M_1^{max} $(= T_1^f/(T_s + T_{ms}))$ slots within one type-1 frame. There are also N_g $(= T_1^f/T_g^f)$ type-g frames in one type-1 frame for all g. From $g = G$, we start to reserve first M_G^f slots in each of N_G type-G frames. Next, for $g = G - 1$, we reserve first M_{G-1}^f available slots in each of N_{G-1} type-$(G-1)$ frames among $M_1^{max} - N_G \cdot M_G^f$ available slots. With the same pattern, we reserve up to type-1 connections, i.e., reserve first M_g^f available slots in each of N_g type-g frames among $M_1^{max} - \sum_{g'=g+1}^{G} N_{g'} \cdot M_{g'}^f$ available slots. To reserve all of the required slots, we must satisfy the condition $\sum_{g=1}^{G} N_g \cdot M_g^f \le M_1^{max}$, which is equivalent to Eq. (3). □

References

[1] S. J. Golestani, A framing strategy for congestion management, IEEE Journal on Selected Areas in Communications, 9(7) (1991) 1064–1077.

[2] Kaveh Pahlavan and Allen H. Levesque, *Wireless Information Networks*, Wiley-Interscience, New York, NY, 1995.

[3] Hamid Ahmadi, Arvind Krishna, and Richard O. LaMaire, Design issues in wireless LANs, Journal of High-Speed Networks, 5(1) (1996) 87–104.

[4] D. Ferrari and D. C. Verma, A scheme for real-time channel establishment in wide-area networks, IEEE Journal on Selected Areas in Communications, 8(3) (1990) 368–379.

[5] S. J. Golestani, Congestion-free communication in high-speed packet networks, IEEE Trans. on Communications, 39(12) (1991) 1802–1812.

[6] A. K. Parekh and R. G. Gallager, A generalized processor sharing approach to flow control in integrated services networks: the single-node case, IEEE/ACM Trans. on Networking, 1(3) (1993) 344–357.

[7] D. D. Kandlur, K. G. Shin, and D. Ferrari, Real-time communication in multi-hop networks, IEEE Trans. on Parallel and Distributed Systems, 5(10) (1994) 1044–1056.

[8] Q. Zheng and K. G. Shin, On the ability of establishing real-time channels in point-to-point packet-switched networks, IEEE Trans. on Communications, 42(2/3/4) (1994) 1096–1105.

[9] J. F. Kurose, M. Schwartz, and Y. Yemini, Multiple-access protocols and time-constrained communication, ACM Computing Surveys, 16(1) (1984) 43–70.

[10] Nicholas Malcolm and Wei Zhao, Hard real-time communication in multiple-access networks, Real-Time Systems, 8(1) (1995) 35–77.

[11] A. S. Acampora and M. Naghshineh, Control and quality-of-service provisioning in high-speed microcellular networks, IEEE Personal Communications, 1(2) (1994) 36–43.

[12] S. Tasaka and Y. Ishibashi, A reservation protocol for satellite packet communication— a performance analysis and stability considerations, IEEE Trans. on Communications, 32(8) (1984) 920–927.

[13] D. Tsai and Jin-Fu Chang, Performance study of an adaptive reservation multiple access technique for data transmissions, IEEE Trans. on Communications, 34(7) (1986) 725–727.

[14] David J. Goodman, Cellular packet communications, IEEE Trans. on Communications, 38(8) (1990) 1272–1280.

[15] F. A. Tobagi and L. Kleinrock, Packet switching in radio channels: Part II-the hidden terminal problem in CSMA and busy-tone solution, IEEE Trans. on Communications, 23(12) (1975) 1417–1433.

[16] Sunghyun Choi and Kang G. Shin, Centralized wireless MAC protocols using slotted ALOHA and dynamic TDD transmission, Performance Evaluation, 27 & 28 (1996) 331–346.

[17] W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw-Hill, New York, NY, 1989.

[18] Arnold M. Michelson and Allen H. Levesque, *Error-control techniques for digital communication*, Wiley, New York, NY, 1985.

[19] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, Kluwer Academic Publishers, Boston, MI, 1988.

[20] S. M. Ross, *Stochastic Processes*, Wiley, New York, NY, 1983.

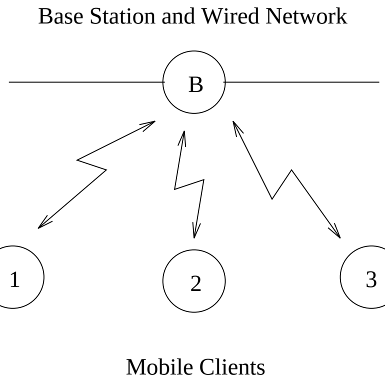


Figure 1: A centralized wireless network with a base station.

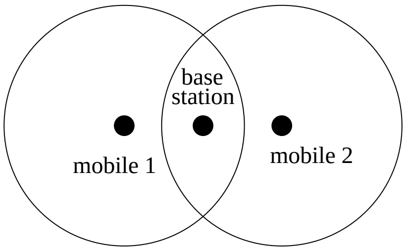


Figure 2: The hidden terminal situation.

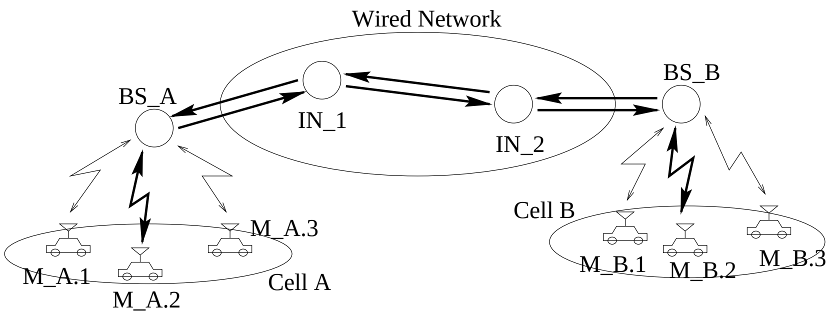


Figure 3: The end-to-end transmission from a sending mobile to a receiving mobile via a wired network.

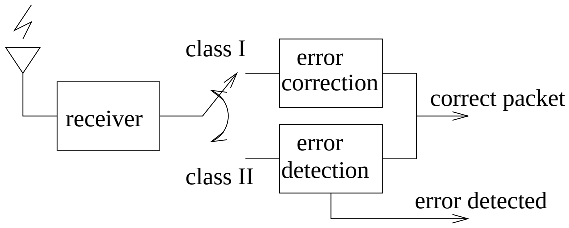


Figure 4: A dual-mode receiver equipped with both error correction (for FEC) and error detection (for ARQ) decoders.

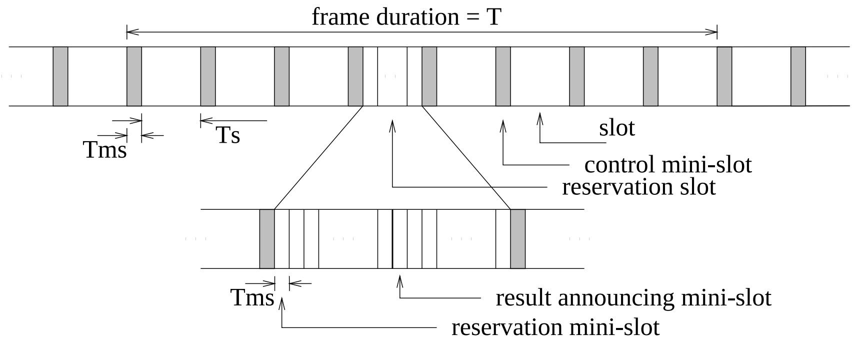


Figure 5: Dividing the time-axis into mini-slots, slots, and reservation slots. A frame includes a number of slots.

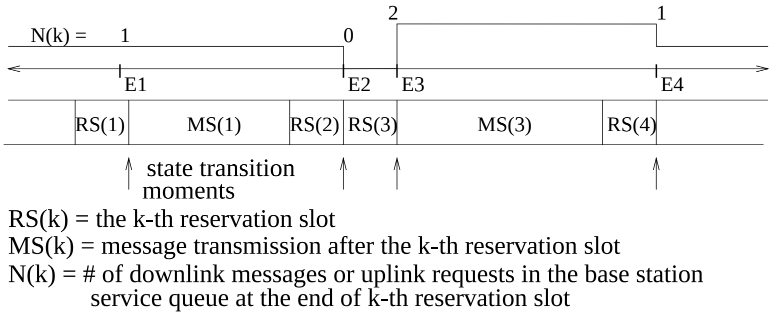


Figure 6: Timing diagram of class-II communications in the absence of class-I traffic.

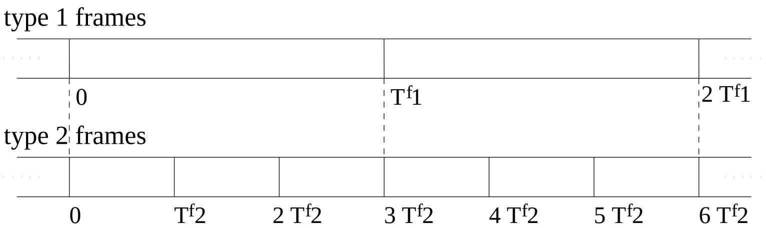


Figure 7: Two frames with duration $T_1^f = 3T_2^f$.

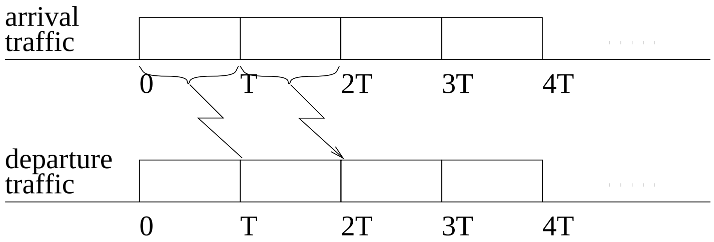


Figure 8: When packets arriving on each frame become eligible for transmission.

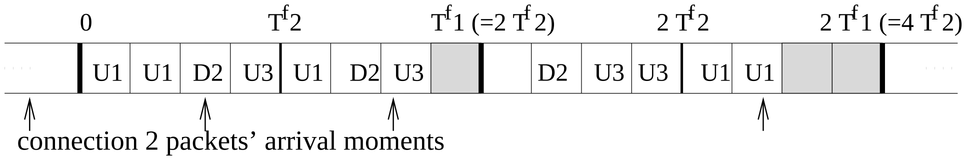
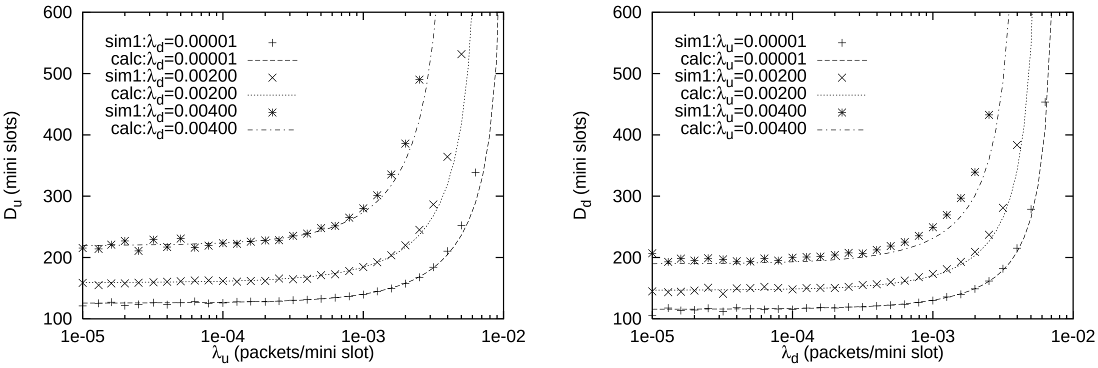
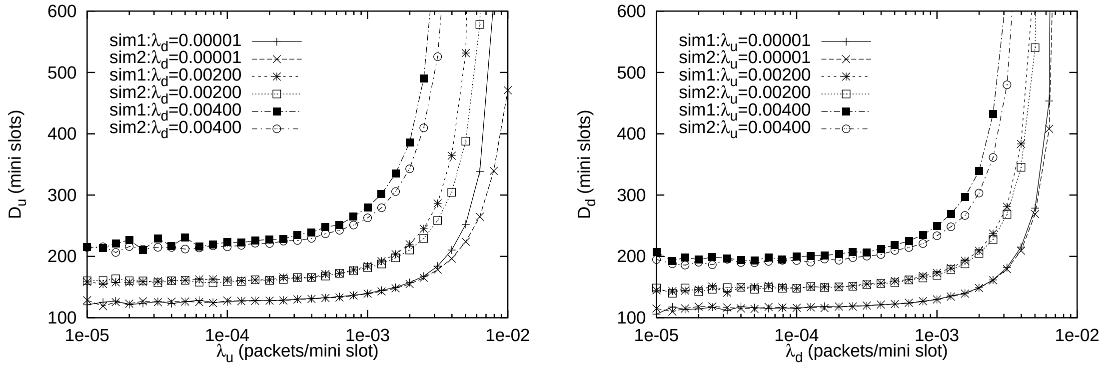


Figure 9: An example of stop-and-go queueing with $G = 2$, three connections.



(a) Uplink; D_u vs. λ_u (b) Downlink; D_d vs. λ_d

Figure 10: Comparison between analytical calculations (marked with *cal c*) and simulations (marked with *sim1*).



(a) Uplink; D_u vs. λ_u (b) Downlink; D_d vs. λ_d

Figure 11: Comparison between the simplified protocol (marked with *sim 1*) and the actual protocol (marked with *sim2*).