An Efficient Polling MAC for Wireless LANs

Oran Sharon and Eitan Altman

***Abstract*—Polling schemes are an important class of medium access control (MAC) protocols for wireless local area networks (WLANs). A major drawback of these schemes is their inefficiency when only a small number of mobile stations have packets to transmit. This inefficiency is due to the polling of mobile stations with no packets to transmit, which delays the transmissions of mobile stations with packets. In this paper, we suggest a new polling MAC which exploits the capture phenomena and enables simultaneous polling and transmissions of information packets. Mathematical analysis and simulation results show that the new MAC overcomes the above inefficiency considerably, and thus it is more efficient in the sense that it enables higher throughput and a lower access delay. For example, we show scenarios in which the average access delay is reduced by about 30% and the throughput increases by 66%–75%.**

***Index Terms*—Capture phenomena, MAC protocols, polling schemes, wireless LANs.**

I. Introduction

POLLING schemes are an important class of medium access control (MAC) protocols for cell-based wireless local area networks (WLANs) [1]–[4]. A cell is composed of a base station (BS) and several mobile stations (for short, we will only use the term *stations*). When a polling MAC is used in a cell, the BS polls the stations for transmissions. Polling MAC protocols have advantages both in physical aspects and in the services they enable. On one hand, in polling MAC protocols, it is known which station is transmitting, thus helping the BS to overcome the fading and inter-symbol interference phenomena, which are a source for transmission impairments when radio channels are used. This is done by the BS using a special adaptive array of antenna elements which are directed toward the transmitting mobile [3]. On the other hand, with contention-free transmissions under a polling MAC, it is possible to guarantee quality of service (QoS) parameters such as delay, delay jitter, and bandwidth to applications that need these QoS characteristics.

The simplest polling scheme is when the BS polls every station in sequence and checks if it has packet(s) to transmit. This scheme has two major drawbacks. The first is related to the mobility of stations, which prohibits the BS from always knowing which stations are under its coverage. The other drawback is related to the situation in which only a small number of stations have packets to transmit [5]. Since the BS polls every station, it can happen that many stations are polled only to learn that they do not have packets to transmit, thus unnecessarily delaying the stations with packets. However, for fairness reasons, every station must be polled in order to enable it to transmit when it has packets. The above unnecessary delay is composed of the transmission times of the polling/response packets, the processing and turnaround times in the stations and the BS, and the propagation delays.

In this paper, we assume a static cell, i.e., mobile stations always belong to one cell only. An example for such a system arises in a room where one server (the BS) serves several clients (mobiles), e.g., classrooms, study rooms, libraries, etc., where the BS serves as an exit/entrance point to outside servers or information bases. Moreover, we assume that stations transmit only while at a fixed location (portable). The main motivation for a WLAN in such an environment is safety. For example, in a classroom full of children, where computers can be located all over the class, including the middle section, using cable LANs such as Ethernet can result in dangerous obstacles. A WAN is a solution to such a problem. Also, there are some convenience considerations—not having to route cables, plug connectors together, etc. is seen as a significant advantage.

We propose a polling scheme that overcomes the inefficiency encountered when only part of the stations have packets to transmit by taking advantage of the well-known capture phenomena in radio channels [6]–[9]. In the scheme, stations with packets can transmit continuously in a Round Robin fashion, while simultaneously the stations without packets are polled to find out if any of them received packets, and thus shall be included in the transmission cycle. The idea behind the scheme is that stations respond to the polling packets from the BS by a transmission of a low signal, while the information packets are transmitted by a strong signal. Due to the capture phenomena, the BS can successfully receive information packets together with responses to its polling packets. Also, notice that since we assume that such a WLAN is located inside a room, there is no interference between networks operating in adjacent rooms.

Our protocol also uses some mechanisms proposed in the IEEE 802.11 MAC protocol [4], which we list below. In the IEEE 802.11 MAC, two modes of operations are possible: contention and contention-free. The standard suggests transmissions in alternating periods, i.e., contention periods separated by contention-free periods.

In the contention-based mode of operation, stations use the CSMA/CA MAC protocol. This MAC is used for the transmission of information packets and for the transmission of control packets by which stations get associated/de-associated with the BS of the cell. The association with the BS is required before any transmission of information packets. In the contention-free mode of operation, the stations transmit by a polling MAC where the BS polls the stations for transmissions.

Manuscript received July 13, 2000; revised November 30, 2000; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor M. Ammar. This paper was presented in part at the HPCS Conference, Chalkediky, Greece, June 1997.

O. Sharon is with the Department of Mathematics, Haifa University, Haifa, Israel (e-mail: oran@mathcs11.haifa.ac.il).

E. Altman is with INRIA BP93, 06902 Sophia Antipolis Cedex, France (e-mail: Eitan.Altman@sophia.inria.fr).

Publisher Item Identifier S 1063-6692(01)06850-9.

1063–6692/01$10.00 © 2001 IEEE

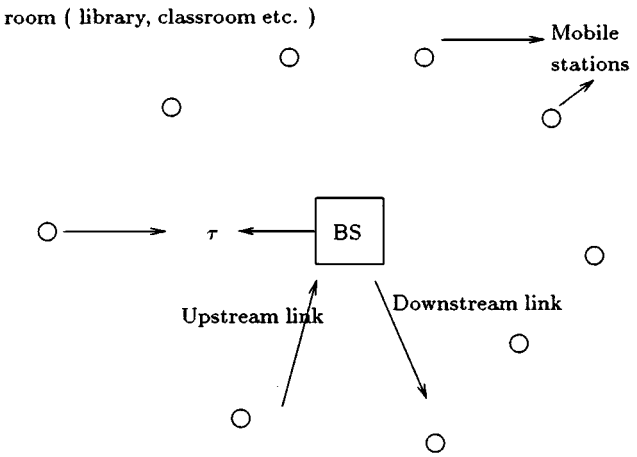

Fig. 1. System model.

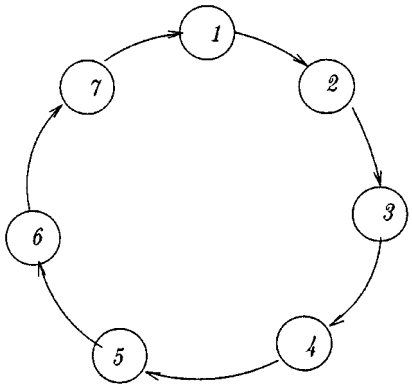

Fig. 2. Example of stations in the System ring.

The BS knows about all the stations under its coverage by the association procedure.

During the operation of the polling MAC, the BS transmits packets to the stations that can have several functions; a packet can simultaneously contain information, an acknowledgment to a previously transmitting station, and at the same time it can also contain a poll query to some other station. In our new protocol, we also assume that the BS maintains a list of stations to poll and also transmits packets with multiple functions.

As mentioned, we assume in this paper a static community of stations in a cell. However, our new MAC can easily be incorporated in a scheme by which stations associate with the BS when entering the cell and de-associate when leaving the cell. A possible way to accomplish this is to use alternating contention and contention-free periods as proposed in the IEEE 802.11 MAC standard. Thus, in the context of the IEEE 802.11 standard, we suggest an improvement to the polling MAC the is used during the contention-free periods.

The rest of the paper is organized as follows. In Section II, we define the system model. In Section III, we describe our new MAC, referred to as simultaneous transmit response polling (STRP), and in Section IV we deal with reliability issues of STRP. In Section V, we compute lower and upper bounds on the packet queuing delay in STRP, and in Section VI we expand the analysis by simulations and compare the simulation results for STRP, for the classical polling scheme and for a modified scheme of STRP that we describe later. Section VII concludes the paper.

II. System Model

We assume a one-cell WLAN that is composed of a BS and of several mobile stations, as depicted in Fig. 1. We assume that the system is static in the sense that stations always belong to the same cell and are connected to the same BS as can be found in classrooms, study rooms, libraries, etc. Moreover, we also assume that stations transmit only while at a fixed location (portable).

There are two transmission links in the system, an Upstream link from the stations to the BS, and a Downstream link from the BS to the stations. The stations transmit to the BS in the Upstream link and the BS transmits to the stations in the Downstream link. The two links are independent. Stations do not hear any transmissions on the Upstream link, and so do not hear any transmissions from the other stations. In other words, they only hear transmissions arriving from the BS on the Downstream link. The stations share the Upstream link, and in the current paper, we propose a new polling based MAC protocol that coordinates the transmissions among the stations on this link. This scenario is again typical in a library, e.g., where readers do not communicate among themselves but only with the BS for outside communication.

Let τ be the maximum propagation delay in the system, i.e., the propagation delay between the far most station to the BS. In our MAC, we assume that τ is bounded and that the minimum transmission time of an information packet is at least 4τ. Notice that since our proposed MAC is designed for a cell, it is reasonable to assume that the distances from the stations to the BS are bounded and predictable, and so τ can be bounded (at least by a standard, as it is the case in Ethernet networks [10]).

As will become clear later, our MAC is based on the capture phenomena. In this phenomena, a receiver—in our case the BS—can receive successfully the transmission of one station, which transmits with a strong signal, even in the presence of a weaker signal transmitted by another station.

The capture phenomena is well known and feasible, and was previously suggested in many other MAC protocols, e.g., [7]–[9]. Moreover, since our system is limited in size, i.e., τ is bounded, the signal powers required to enable the capture phenomena at the BS can be computed, also for the extreme case where the strong and weak signals are transmitted by the far most and the nearest stations from/to the BS, respectively [11]–[13].

III. The STRP MAC

The STRP MAC utilizes three logical rings at the BS. The first ring, referred to as *System ring*, contains all the stations in the system, which are arranged in a cyclic order. For example, assume that the system contains stations 1–7, which are arranged in the System ring as depicted in Fig. 2. In a regular polling scheme, the BS checks with station 1 if it has packets to transmit, then checks with station 2, and so on.

The other two rings are imposed and defined on the System ring. One ring consists of Active stations and the other Idle stations. Active stations are stations that notified the BS that they

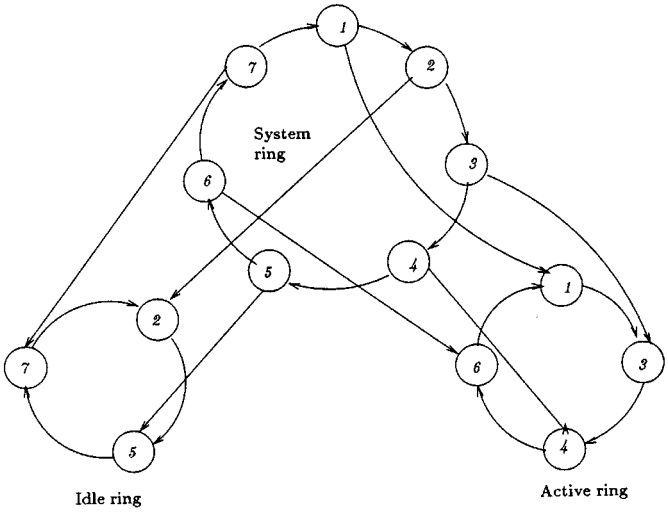

Fig. 3. Example of stations in the Active and Idle rings. The order is imposed by the System ring.

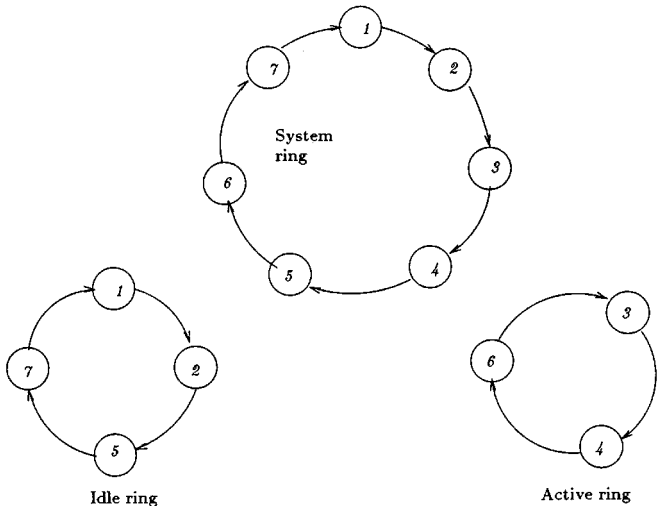

Fig. 4. Idle and Active rings after station 1 is moved from the Active to the Idle ring.

have packets to transmit. Idle stations are stations that notified the BS that they do not have packets to transmit. We refer to the two rings as the *Active ring* and *Idle ring*, respectively.

Consider Fig. 3 and assume that stations 1, 3, 4, and 6 announced to the BS that they have packets to transmit. These stations belong to the Active ring. Stations 2, 5, and 7 announced to the BS that they do not have packets to transmit and so they belong to the Idle ring. The relative order in each ring is the same as in the System ring.

A. The Active Ring

Assume that stations 1, 3, 4, and 6 are in the Active ring and in the order shown in Fig. 3. The BS gives the stations in the Active ring permission to transmit in a Round Robin fashion, according to their order in the ring. Thus, station 1 transmits first, then stations 3, 4, and 6 transmit, and so on. The mechanism by which this is done is described later. Every station in the Active ring, when transmitting, announces to the BS if it wants to stay in the ring for another round, i.e., if it has an additional packet to transmit apart from the one that it is currently transmitting. If the station has another packet for transmission, it stays in the Active ring; otherwise, it is transferred to the Idle ring by the BS. The place in the Idle ring into which the station is inserted is imposed by the System ring. For example, if station 1 in Fig. 3 is moved from the Active ring to the Idle ring, the resulting two rings are as shown in Fig. 4.

B. The Idle Ring

Assume that stations 2, 5, and 7 are in the Idle ring and in the order shown in Fig. 3. As mentioned earlier, this ring contains the stations that announced to the BS that they do not have packets to transmit. The BS polls the stations in the ring in a Round Robin order and asks every station if it has recently received packet(s) to transmit. If a station does not have packets to transmit, it is left in the ring. Otherwise, it is moved to the Active ring. Again, the place in the Active ring into which the station is inserted is imposed by the System ring. For example, if station 2 in Fig. 3 is moved from the Idle ring to the Active ring, the resulting two rings are as shown in Fig. 5.

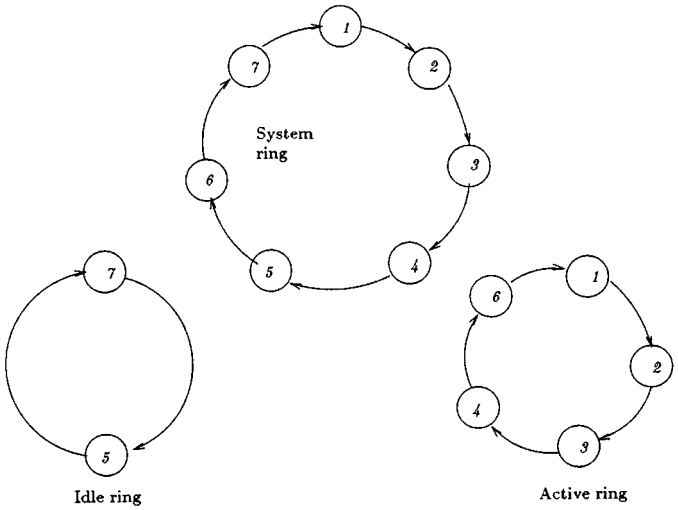

Fig. 5. Idle and Active rings after station 2 is moved from the Idle to the Active ring.

C. The Transmission Pattern

We now describe the method by which the BS enables stations in the Active ring to transmit and how it polls stations in the Idle ring. In our scheme, we utilize the capture effect by which the BS is capable of receiving and interpreting a strong signal in the presence of a weaker signal. We divide the description below into three cases: all the stations are in the Idle ring, all the stations are in the Active ring, and finally, we consider the case in which both the Active and Idle rings are not empty.

1) Case 1: All the Stations are in the Idle Ring: In this case, the BS asks the stations in the Idle ring if they have information packets to transmit. The BS uses a special control packet, *Query* (Q), for this purpose. In *Query*, the BS includes the identity I of the station to which it transmits the packet. I answers the BS with a *No-Packet (NP)* control packet if it does not have packet(s) to transmit. Otherwise, if I has information packet(s), it transmits one packet and announces to the BS if it has more packets to transmit. The announcement is done by a special bit in the information packet.

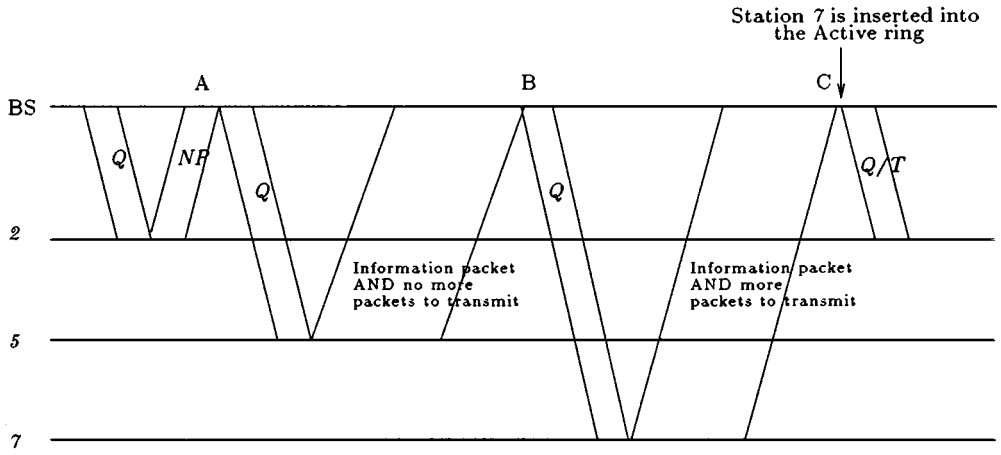

Fig. 6. A transmission pattern for the case when the Idle ring is not empty and the Active ring is empty.

If the BS receives an information packet and an announcement from I that it has more packets to transmit, it transfers I into the Active ring. Otherwise, if an *NP* packet is received or I announces that it does not have any more information packets to transmit, it is left in the Idle ring. Finally, after receiving a *NP* control packet or an information packet, the BS continues and transmits *Query* to the next station in the Idle ring, if such exists.

In Fig. 6, we show how the BS queries the stations in the Idle ring, based on the assumption that only stations 2, 5, and 7 are in the system. The BS transmits a Q packet to station 2 which responds with a *NP* packet (point A). Thus, station 2 is left in the Idle ring. The Active ring is still empty and the BS, still in case 1, transmits a Q packet to the next station in the Idle ring which is station 5. Station 5 has an information packet and it announces to the BS that it has no more packets to transmit. Thus, station 5 is left in the Idle ring (point B). Then, the BS transmits a Q packet to station 7. Station 7 has packets to transmit. It transmits a packet and announces to the BS that it has more packets to transmit. Thus, station 7 is moved to the Active ring (point C). Now, both the Active and the Idle rings are not empty and the system is as in case 3. In this case, to be described later, the BS transmits another type of control packet, *Query/Transmit* (Q/T), to the next station in the Idle ring, which is again station 2. The way by which station 2 responds to this packet is also described in case 3.

2) Case 2: All the Stations are in the Active Ring: In this case, the BS enables the stations in the Active ring to transmit information packets according to their order in the ring. The BS uses a special control packet, *Transmit* (T), for this purpose. In this packet, the BS includes the identity I of the station to which it enables transmission. I transmits its information packet after receiving the *Transmit* packet. Also, I announces to the BS if it has more packets to transmit. If I has more packets, it is left in the Active ring. Otherwise, it is moved to the Idle ring. Finally, after detecting the end of an information packet, the BS enables the next station in the Active ring, if the Active ring is not empty, to transmit.

In Fig. 7, we show how the BS queries the stations in the Active ring, based on the assumption that only stations 1, 3, 4, and 6 are in the system. The BS transmits a T packet to station 1 which transmits an information packet and signals that it has more packets to transmit. Thus, station 1 remains in the Active ring (point A) and the Idle ring is still empty. The BS, still in case 2, transmits a T packet to the next station in the Active ring, which is station 3. Station 3, upon receiving the T packet, transmits an information packet and signals that it does not have more packets to transmit. When the BS receives the packet, it transfers station 3 to the Idle ring (point B). Now, both the Active and the Idle rings are not empty and the system is in case 3 which will be described later. In this case, the BS transmits another type of control packet, Q/T, to the next station in the Active ring, which is station 4. The way in which station 4 responds to this packet is described in case 3.

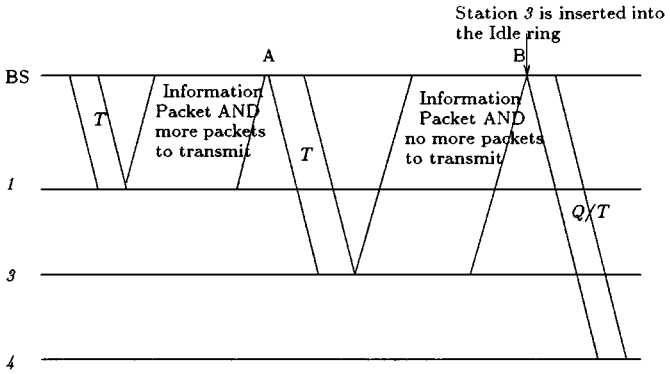

Fig. 7. A transmission pattern in the case when the Active ring is not empty and the Idle ring is empty.

3) Case 3: The Active and Idle Rings are Not Empty: When both rings are not empty, the BS enables a station in the Active ring to transmit an information packet and polls a station in the Idle ring by the same control packet. This packet is denoted by *Query/Transmit* (Q/T). It includes the identity I of the station in the Active ring that receives the right to transmit. It also contains the identity J of the station in the Idle ring to which the BS transmits a query.

When I receives the *Q/T* packet, it begins to transmit its information packet. When J receives the *Q/T* packet, it waits 2τ time units and then begins to transmit a *Jam* if it has packet(s) to transmit and wants to join the Active ring. Otherwise, if J does not have packet(s) to transmit, it remains quiet. Station I

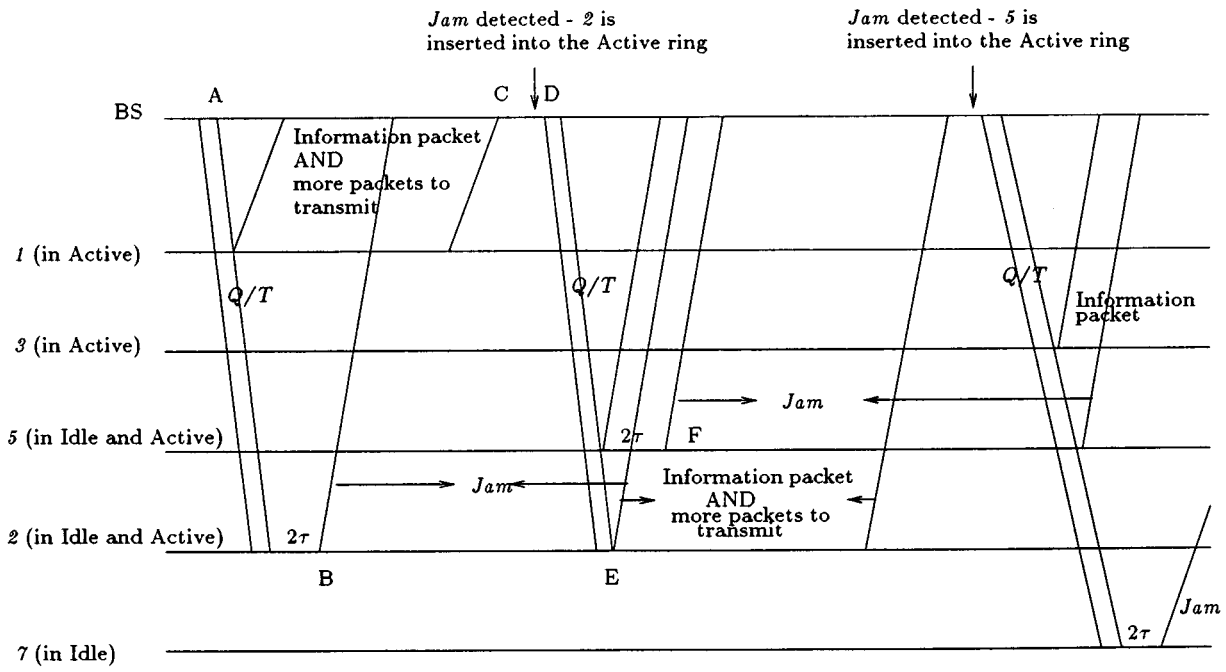

Fig. 8. A transmission pattern in the case when the Active and Idle rings are not empty.

transmits its packet by a strong signal while J transmits its *Jam* by a weaker signal. Notice that the *Jam* from J collides with the information packet from I. However, by the capture effect, since J transmits the *Jam* by a weaker signal than that of I, the BS succeeds in receiving the packet from I. As mentioned, this mechanism, which exploits the capture phenomena by scheduling transmissions of different signal powers, is well known in the literature [7]–[9], [11]–[13].

I transmits until it finishes transmitting its information packet and it signals the BS if it wants to stay in the Active ring. After detecting the end of the packet from I, the BS continues to listen to the Upstream link. If it continues to detect a signal, it detects the *Jam* from J, meaning that J has packet(s) to transmit and shall be moved to the Active ring. Otherwise, if no signal is detected, this means that station J does not have packets to transmit. The BS now transmits a new control packet, according to the aforementioned three cases. When J receives the next control packet from the BS, it terminates the transmission of the *Jam* if it is transmitting one. Notice that if the next control packet from the BS triggers another station K in the Idle ring to transmit a *Jam*, then since K waits 2τ time units after receiving the control packet from the BS and before starting the *Jam* transmission, the BS always receives at most one *Jam* at a time and can receive successfully any information packet.

A last important remark is that if the length of the information packets is fixed and known in advance, we can omit the use of the capture phenomena by enforcing a station in the Idle ring to compute, after receiving the *Query/Transmit* packet, when the transmission of the information packet is finished and then to transmit a *NP* packet or an information packet, as described in case 1.

In Fig. 8, we show how the BS transmits to the stations in the Active and Idle rings of Fig. 3. The BS transmits a Q/T packet to stations 1 and 2 (point A). Station 1 begins to transmit its packet and station 2, assuming that it has packet(s) to transmit, transmits a *Jam* after waiting 2τ time units (point B). After the BS detects the end of the packet from station 1 (point C) and the *Jam* from station 2 (point D), station 1 is left in the Active ring and station 2 is moved to the Active ring. The BS now transmits a Q/T packet to stations 2 and 5. When this packet is detected by station 2, it stops transmitting the *Jam* (point E) and begins to transmit its information packet. Station 5, assuming that it has a packet to transmit, waits 2τ time units and then begins to transmit a *Jam* (point F). Notice that the BS enables station 2 to transmit a packet immediately after it is inserted into the Active ring because it appears after station 1 in the System ring, and hence, is now the first station in the Active ring. The rest of Fig. 8 shows the continuation of the activity in the system according to the same rules.

IV. Reliability

In this section, we deal with possible packet losses and station failures, and their impact on the protocol operation.

A. Loss of Packets

In the case when the Active ring is empty, i.e., all the stations are in the Idle ring, the loss of a *Query*, *No-Packet*, or an information packet from a station J can be detected by the BS by a time-out mechanism. The BS can then re-transmit a *Query* packet to J or move on to the next station in the Idle ring. No station is moved to the Active ring in this case but this has no effect on the proper operation of the MAC. Also, if an information packet from J is lost its recovery is taken care of, if necessary, by higher layers in the protocol stack, e.g., in the LLC or transport layers.

In the case where the Idle ring is empty, i.e., all the stations are in the Active ring, then the loss of a *Transmit* or an information packet from a station I can again be detected by the BS by a

time-out mechanism. The BS can re-transmit a *Transmit* packet to I or move on to the next station in the ring. Again, if an information packet from I is lost, its loss shall be recovered, if necessary, by higher layers in the protocol stack. Notice that all the packet losses described so far have no impact on the ongoing proper operation of the protocol.

When there are stations in the Active and Idle rings, the BS transmits a *Query/Transmit* packet, say to stations J and I in the Idle and Active rings, respectively. If this packet is lost then the BS can detect this loss by a time-out mechanism and by not receiving any packet from I. If the *Query/Transmit* packet is not lost but the information packet from I is lost, again the BS can detect this loss by a time-out mechanism. Notice, however, that it can happen that the BS will interrupt the lost packet as a *Jam* from station J in the Idle ring.

If station J is transmitting a *Jam* anyway, no harm is done. Otherwise, J is moved to the Active ring when it does not have any packets for transmission. However, at the time when it is polled in the Active ring it can transmit a *No-Packet* to the BS and it will be moved back to the Idle ring.

Finally, in the case that the *Jam* is lost, i.e., the BS does not detect the *Jam* from J, J remains in the Idle ring and it does not receive a permission to transmit information packets. However, the next *Query* or *Query/Transmit* packets that it will receive from the BS will enable it to move to the Active ring and eventually to transmit. The loss of the *Jam* therefore results only in a higher access delay that J encounters.

B. Stations Failure

Clearly, if the BS fails, there are no transmissions in the system ,and this event shall be detected by the management system of the physical network to which the BS is connected.

If a station J in the Idle ring fails then J will not respond to *Query* or *Query/Transmit* packets. In the latter case the BS will not detect any *Jam* signal from J, will conclude that J does not have packets to transmit and will leave J in the Idle ring. In the case of a *Query* packet, the failure of J can be detected by the BS not receiving a *No-Packet* or an information packet from J within a time-out interval. The BS can then try to transmit to J several times, later to conclude that probably J is not operational and it can return to poll J after a pre-defined time interval.

If a station I in the Active ring fails, then its failure can be detected by the BS when it does not receive information packets from I in response to *Transmit* or *Query/Transmit* packets. The BS can try to send these packets to I several times, in order to cope with possible packet losses, before it concludes with high probability that I failed. The BS may then transfer I to the Idle ring and when I is operational again, it will respond to *Query* or *Query/Transmit* packets as a regular station in the Idle ring.

V. Computation of Lower and Upper Bounds on the Queueing Delay in STRP

A. Overhead Computation

In the following, we denote the usual polling scheme in which the BS polls every station in a Round Robin order by U-poll. The two quantities that influence the efficiency of STRP and U-poll are the transmission times of the information packets and the overhead due to the polling. We first analyze the overhead in U-poll and STRP, then we compute lower and upper bounds on the queuing delay in STRP.

In U-poll, the BS polls every station in a Round Robin fashion. The overhead is divided into two parts: the overhead due to the polling packets and the overhead due to the transmission of the response or information packets. The overhead due to a polling packet consists of its transmission time, propagation delay, and the processing time in the mobile station. We denote the sum of these three components by OH1 (OverHead type 1).

If the station does not have information packets to transmit, it transmits a response control packet to the BS. The associated overhead with this transmission is OH1 also. If the station transmits an information packet, the overhead is the propagation delay of the packet and the processing time in the BS. We denote this overload by OH2 and OH2 $=$ OH1 $-$ {transmission time of a control packet}.

Consider STRP now. When the BS transmits a Q packet to a station in the Idle ring, the associated overhead is OH1. Now, we assume that while waiting for a response from the station, the BS can compute which of the next stations are in the Idle and Active rings, either if the polled station stays in the Idle ring or whether it is moved to the Active ring. When the response from the station arrives, the BS decides if the station stays in the Idle ring or is moved to the Active ring. In either case, it has the correct structure of each ring from its advanced computation. Therefore, the computation of the next stations in the Idle and Active rings is not considered an additional overhead.

If the polled station in the Idle ring responds with an *NP* control packet, the associated overhead is OH1. Otherwise, if the station responds with the transmission of an information packet, the associated overhead is OH2. Notice that we neglect the transmission time of the one additional bit in the information packet which signals to the BS whether the station does or does not have more information packets to transmit. If the BS transmits a T control packet to a station in the Active ring, the associated overheads with this transmission and the transmission of the active station are OH1 and OH2, respectively.

Consider now the case when the BS transmits a Q/T control packet. The overhead associated with its transmission is OH1. The overhead associated with the transmission of the information packet from the station in the Active ring is OH2. However, after the BS receives this packet, it does not move to the next polling immediately but continues to listen and detect if there is a Jam in the Upstream link, i.e., it continues to be in a receiving mode and after detecting whether there is a Jam on the channel or not, it moves to the next polling. We denote the overhead associated with the response to a Q/T packet by OH3 where OH3 $=$ OH2 $+$ {detection time}.

B. Analysis Model

We now introduce the model and notations used later in our analysis.

- The system is composed of N sources (stations), denoted $0, 1, \ldots, N-1$.

- The arrival process of packets to the queue of station i, $i = 0, 1, \ldots, N-1$, is Poisson with intensity $\lambda_i = \lambda/N$, independent of the other stations.
- Stations have infinite packet queues.
- Time is divided into slots. A slot starts when a control packet is sent by the BS. A control packet can be of type Q, T or Q/T. Recall that a slot that starts with a Q packet can be of size $2 \cdot$ OH1 (if there is no information packet transmission) or OH1 + OH2 + (transmission time of an information packet). A slot that begins with a T packet takes OH1 + OH2 + (transmission time of an information packet). A slot that begins with a Q/T packet takes OH1 + OH3 + (transmission time of an information packet).

  Notice that slots with transmissions are of two lengths—they both contain the transmission time of an information packet. In addition, one type contains OH1 + OH3 time units and the second contains OH1 + OH2 time units. For simplicity, as will become clear later, we assume that both types of slots are of the same length, and let it be b time units. This means that OH3 equals approximately OH2, or that they are both negligible w.r.t. information packet transmission time plus OH1. Define $\rho = \lambda b$.
- If the Active ring is empty then it takes $2 \cdot$OH1 time units, where $2 \cdot \text{OH1} \leq b$, to poll a station in the Idle ring if that station is empty.

We define a slot to be in one of three possible states.

- *1) Active state:* A slot during which there is a transmission by a station in the Active ring.
- *2) Idle state:* A slot during which there is a transmission by a station in the Idle ring.
- *3) Off state:* A time slot with no information packet transmission. In this case, all the stations are in the Idle ring and none of them have packets to transmit, or at least one station has a packet to transmit. There is no transmission, since the stations with packets have not yet been polled.

If the system is in state 3, it is said to be on vacation. When transmission starts in the Active ring after it was empty, we say that a busy period starts. We say that a busy period ends when the Active ring empties. Thus, the busy periods contain all the slots with transmissions from stations in the Active ring i.e., all the slots in state 1). Note that transmissions may occur also in the nonbusy periods, i.e., the transmissions from the Idle ring in the slots in state 2). Slots in state 2) are neither in busy nor in vacation periods.

Recall now that packets are transmitted from the Active ring according to a Round Robin discipline, one from each station. Parallel to this, if the Idle ring is not empty, it is also polled in a Round Robin way. If a packet is found in a station in the Idle ring during a busy period, this station joins the Active ring.

If a station I in the Idle ring has a packet to transmit while the Active ring is empty, its packet is transmitted; however, the station joins the Active ring only if it has more than one packet to send. A busy period, therefore, starts if the Active ring was empty and station I has more than one packet to transmit.

C. Workload Computation

We now proceed to compute the distribution of the workload[1] in the system. This performance measure is not only of interest on its own, but it also allows us to obtain performance measures of the queue sizes and delays in the system.

We begin by citing some assumptions and a decomposition Theorem from [14]. We assume that the process of total workload in the system has an equilibrium distribution (the condition $\rho < 1$ is clearly necessary). Under the assumptions of our model, we have the following.

Theorem 1 ([14, Th. 2.1]): The steady-state amount of workload in the system is distributed as the sum of the following two independent random variables:

i) the steady-state amount of workload V in the corresponding M/G/1 system (which has no vacations);
ii) the steady-state amount of workload Y present in the original system at a vacation interval.

Remark: Theorem 1 holds under the assumption that the transmission time, i.e., the duration of time slots during which transmissions occur, does not depend on whether we transmit the information packet from the Active ring or from the Idle ring. Therefore, we use the simplification in the duration of time slots with transmissions, which is denoted by b. Note that the *combined* effect of transmissions from the Idle ring and from the Active ring contributes to the term i) in Theorem 1, so we do not have to consider these two separately. The main problem is to compute the term ii), which is obtained by considering nontransmission periods (vacations).

In what follows, we shall focus on each of the terms in Theorem 1. First notice that the Laplace transform of V in i) of Theorem 1 is (e.g., [14])

$$V^*(s) = \frac{(1-\rho)s}{s - \lambda + \lambda B^*(s)} \tag{1}$$

where $B^*(s)$ is the Laplace transform of the service (transmission) time, i.e.,

$$B^*(s) = e^{-sb}. \tag{2}$$

We now proceed to obtain Y that appears in ii) of Theorem 1.

Consider an arbitrary time t during a vacation period. Assume that this vacation starts at a time T_n (T_n is defined to be the time when the nth slot starts since the system started to operate. We assume that the system is initially in steady state), that $t \in [T_m, T_{m+1})$, $m \geq n$, and that station j is polled at time T_m. Note that $T_{m+1} - T_m = 2 \cdot$ OH1, since there is no packet to transmit (otherwise t would not correspond to a vacation period)—see Fig. 9. In the rest of the section, we will use d to denote $T_{m+1} - T_m$. The workload at time t is composed of the sum of two independent terms:

i) the workload that arrived during $[T_m, t)$, denoted by $\hat{Y}$;
ii) the workload at time T_m, denoted by $\tilde{Y}$.

We begin by computing i) above. The past recurrent time $X = t - T_m$ is uniformly distributed over the interval $[0, d]$, since

[1]The workload at time t is defined to be the total duration of the slots, during which transmission of information packets present in the system at time t, will occur. We include in the workload the remaining time of the current slot, if it contains a transmission of an information packet.

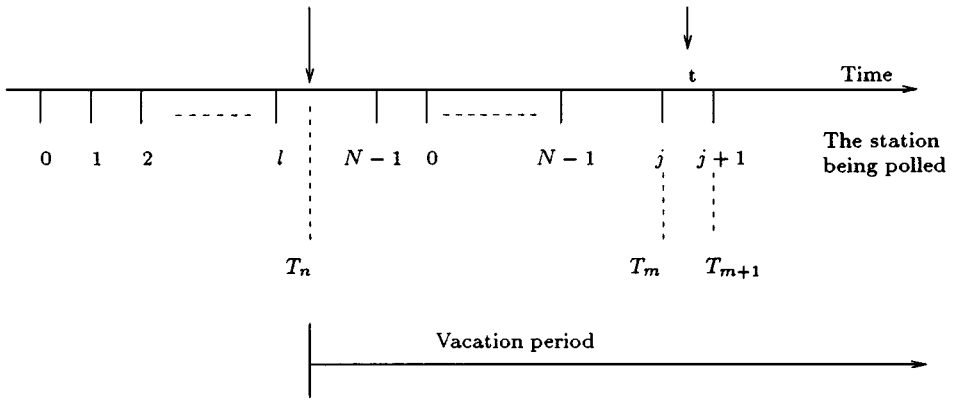

Fig. 9. Description of a vacation period.

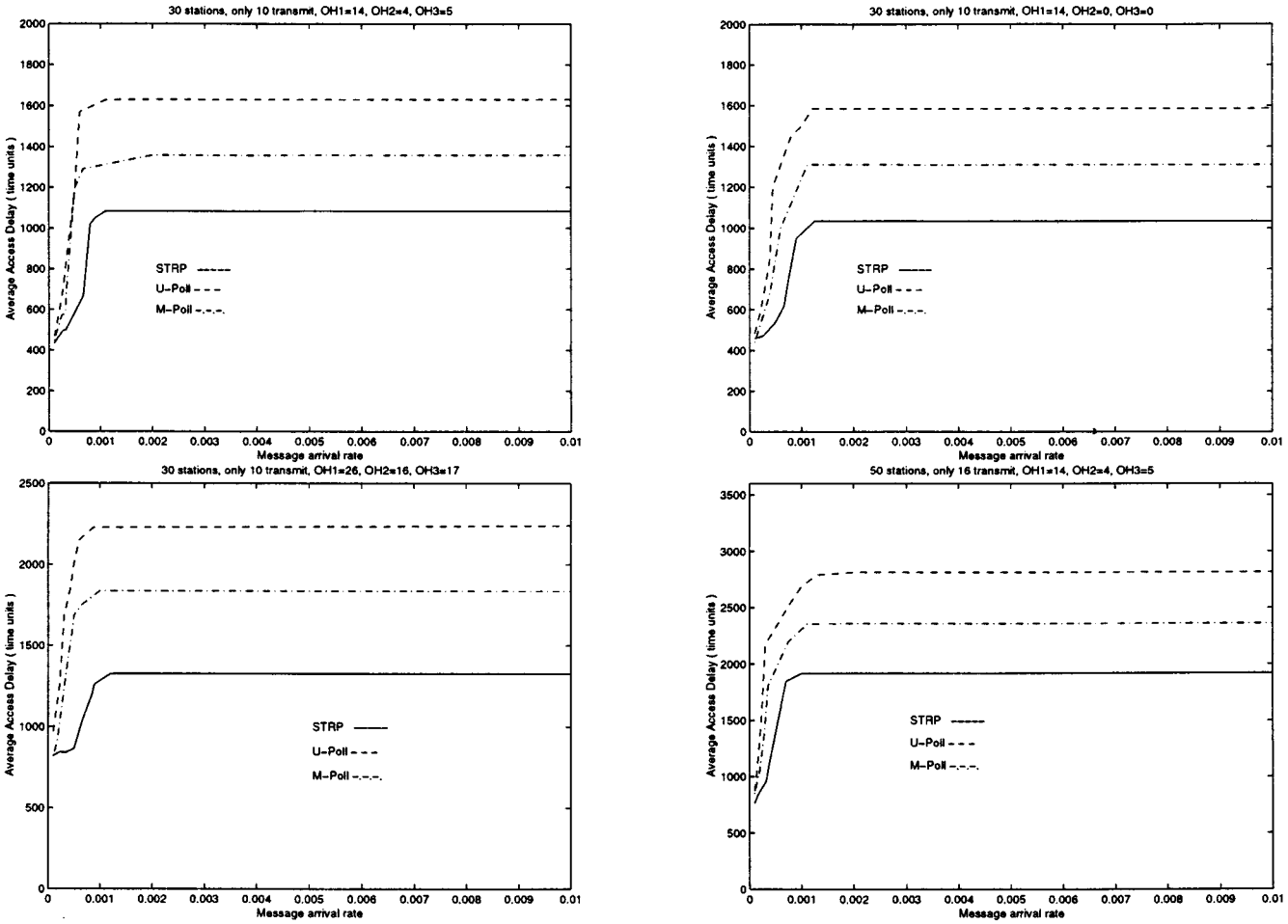

Fig. 10. Average access delay versus message arrival rate for the case when only a part of the stations transmit.

the Idle ring polling slots are assumed to be constant ($=d$). The Laplace transform and expectation of X are thus

$$X^*(s) = E[e^{-sX}] = \frac{1}{d}\int_0^d e^{-sx}\,dx = \frac{1}{sd}(1 - e^{sd}) \quad (3)$$

$$E[X] = d/2. \quad (4)$$

We use the following well known relations.

Lemma 1: Consider a Poisson arrival process with rate λ. Let $A(T)$ be a RV corresponding to the number of arrivals during a time interval of length T. We assume that T and the arrival process are independent. Then, the probability generating function of $A(T)$ is

$$G(z) = E\left[z^{A(T)}\right] = T^*(\lambda - \lambda z) \quad (5)$$

where $T^*(s)$ is the Laplace transform of T. Let $\hat{V}$ be the workload that arrives during interval T. The Laplace transform of the

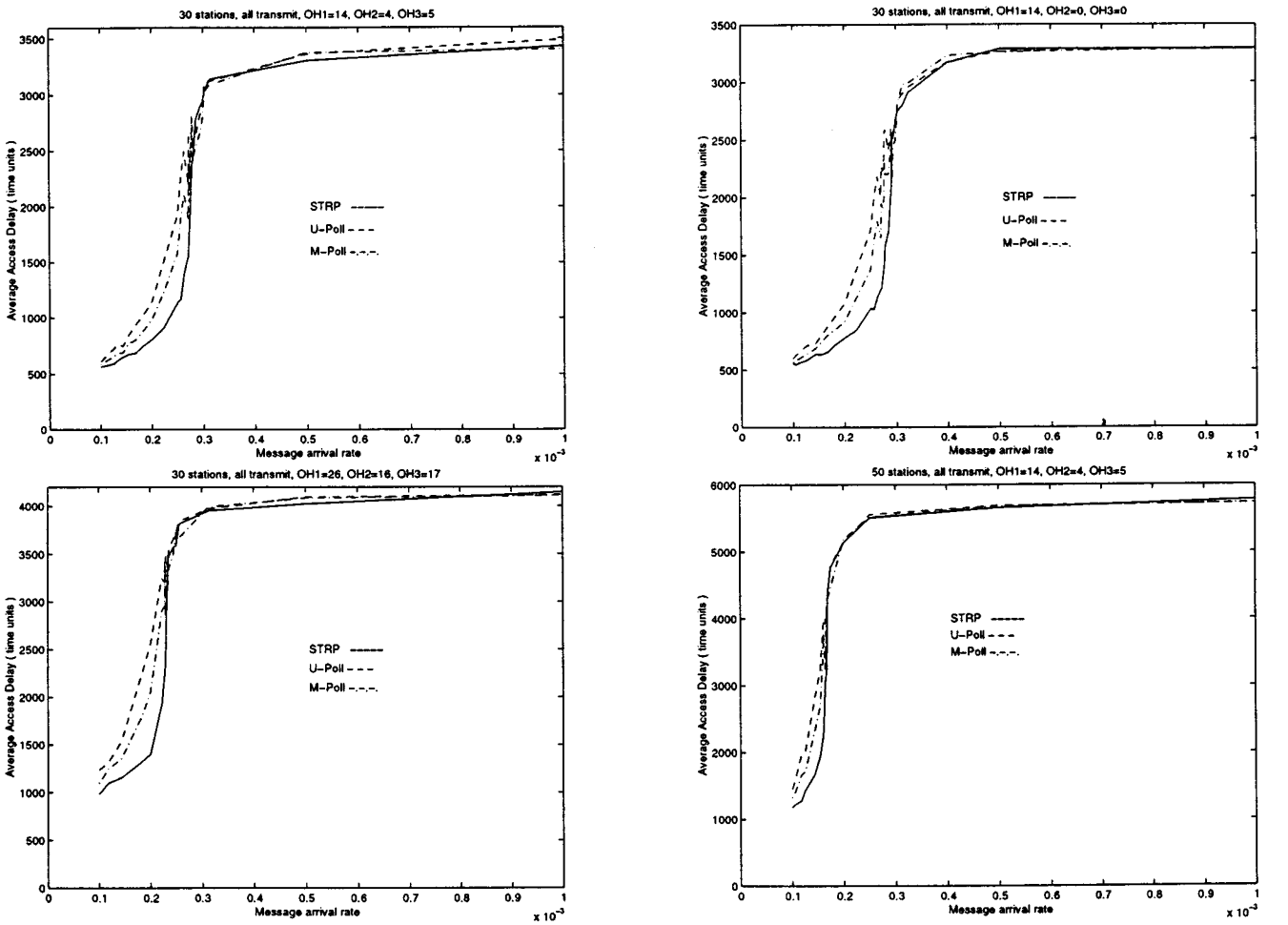

Fig. 11. Average access delay versus message arrival rate for the case when all the stations transmit.

workload that arrives during the interval T is

$$\hat{V}^*(s) = E\left[e^{s\hat{V}}\right] = G(B^*(s)) = T^*(\lambda - \lambda B^*(s)) \quad (6)$$

where $B^*(s)$ is the Laplace transform of the service time.

From the above, we get that the expected number of packets that arrive during T is $\lambda E[T]$, and the expected amount of workload that arrives during that time is $\lambda b E[T]$.

Notice now that in our case, $B^*(s) = e^{-sb}$. With substituting the expression (3) instead of $T^*(s)$ in (6), we get the Laplace transform of $\hat{Y}$:

$$\hat{Y}^*(s) = \frac{1}{(\lambda - \lambda e^{-sb})d}\left(1 - e^{(\lambda - \lambda e^{-sb})d}\right). \quad (7)$$

We now proceed to compute ii), the workload $\tilde{Y}$ due to packets that are in the system at time T_m.

Let Q_m be the number of packets in the system at time T_m. Note that $\tilde{Y} = Q_m b$. Q_m is the sum over $k = 1, 2, \ldots, N-1$ of $Q_{j,(j-k) \bmod N}$, where $Q_{j,(j-k) \bmod N}$ is the number of packets queued at station $(j-k) \bmod N$ at time T_m.

We now prove the following Lemma.

Lemma 2: $Q_{j,(j-k) \bmod N}$, $k = 1, \ldots, N-1$ are *upper bounded* by the number of packets that arrived at station $(j-k) \bmod N$ during $[T_{m-k}, T_m)$.

.

Theorem 2: Under the assumptions of this section, the amount of workload in the system in steady state is:

A) stochastically lower bounded by the random variable V whose Laplace transform is given in (1);

B) stochastically upper bounded by the sum of the independent RVs $V + \hat{Y} + V'$ whose Laplace transform are given in (1), (7), and (12), respectively.

D. The Lower and Upper Bounds on the Queuing Delay

Using Theorem 2, we now obtain the following bounds on the expected queue length and queuing delay.

Theorem 3: The expected number of packets in the system in steady-state, including the one being transmitted is:

A) lower bounded by the expectation of Q, where Q is the average queue in an M/G/1 queue without vacation, i.e.,

$$E[Q] = \frac{E[V]}{b} = \frac{\lambda^2 b^2}{2(1-\rho)} + \rho \quad (13)$$

B) upper bounded by $E[\hat{Q}] + E[Q] + \lambda d/2$, where $E[\hat{Q}]$ and $E[Q]$ are given in (10) and (13), respectively.

The expected queuing delay is:

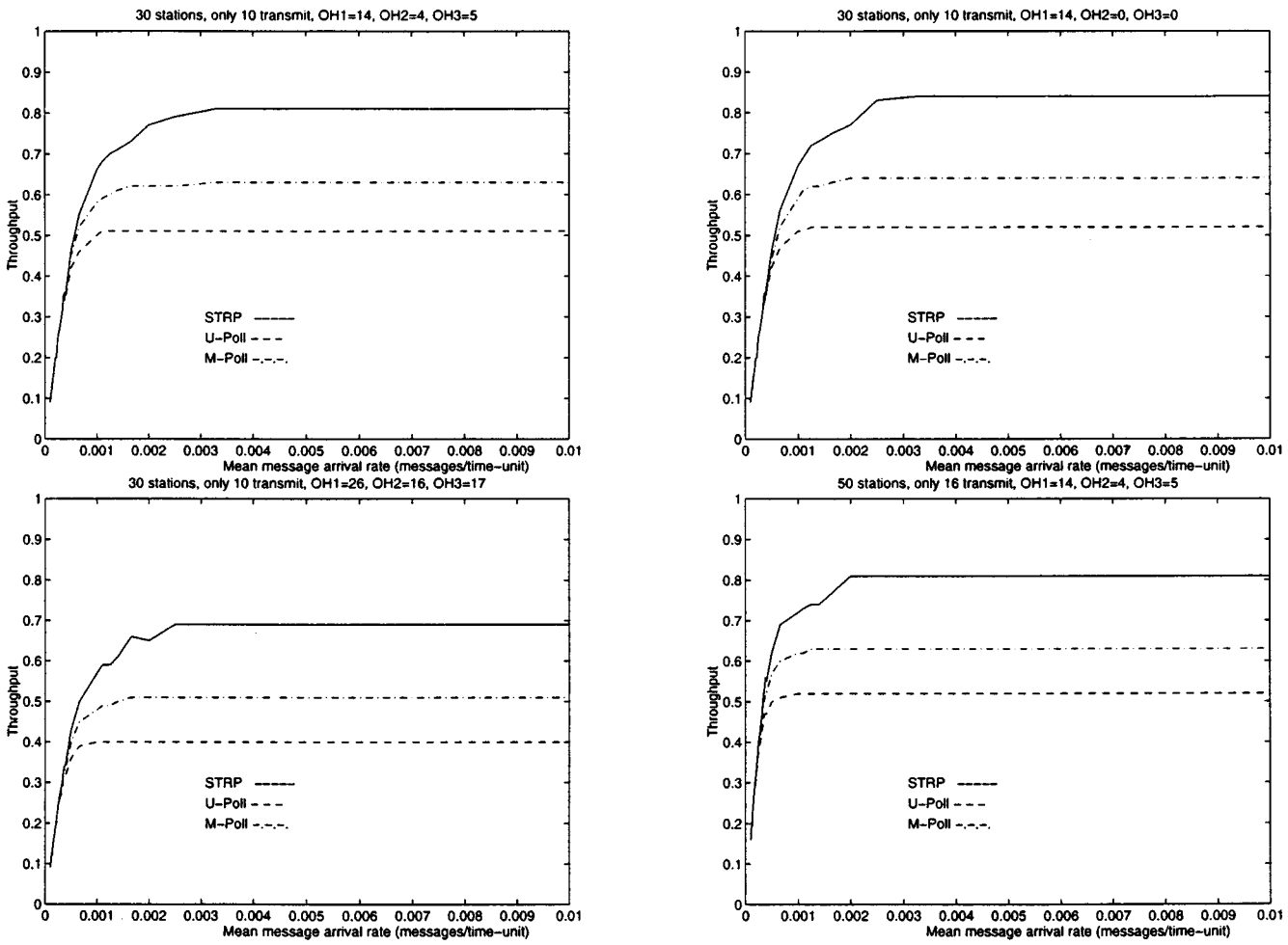

Fig. 12. Throughput versus message arrival rate for the case when only a part of the stations transmit.

C) lower bounded by

$$E[D] = \frac{\lambda b^2}{2(1-\rho)} + b \tag{14}$$

D) upper bounded by $E[\hat{D}] + E[D] + d/2$ where $E[\hat{D}]$ is given by

$$E[\hat{D}] = \frac{(N-1)b}{2}. \tag{15}$$

Proof: Obtained directly from Theorem 2. The workload is by definition the amount of time (including overhead) required to transmit all the packets in the system. Since the time to transmit a packet is assumed to be b, the expected workload in the system is b times the expected number of packets in the system, in steady state. The exact expressions for A) and B) can be obtained by taking derivatives in (1) and (12), respectively, and by substituting $s = 0$ to get the expectations for the corresponding terms of the workload, and then divide by b. But it can also be obtained directly: A) is known to be the expected queue length in an M/G/1 queue with no vacations, and B) already obtained in (10) from a similar derivative of the PGF in (9). The term $\lambda d/2$ is obtained by combining (4) with the last statement of Lemma 1.

The expected queuing delay is obtained from Little's law, according to which it equals the expected queue length divided by the arrival rate λ. □

VI. Simulation Results

In our simulations, we checked the performance of STRP and U-poll. However, notice that it is also possible to add the mechanism by which a station notifies the BS if it has more packets to transmit to the U-poll scheme. We add this mechanism to U-poll and a station that notified the BS that it does not have packets to transmit is not polled in the next round. It is polled again in the second-to-next round. We denote this scheme by modified poll (M-poll).

We simulate systems with 30 and 50 stations. In one simulation scenario, only part of the stations generate and transmit packets—10 stations in the case of 30 stations, and 16 in the case of 50 stations. Notice, however, that all the stations are polled because the BS does not know that only part of the stations actually generate packets and transmit. This scenario can happen, for example, when part of the pupils in a class are doing computer work that evolves communication, while the other pupils are not using their computers, but their computers are turned on. In the second simulation scenario, all the stations generate and transmit packets.

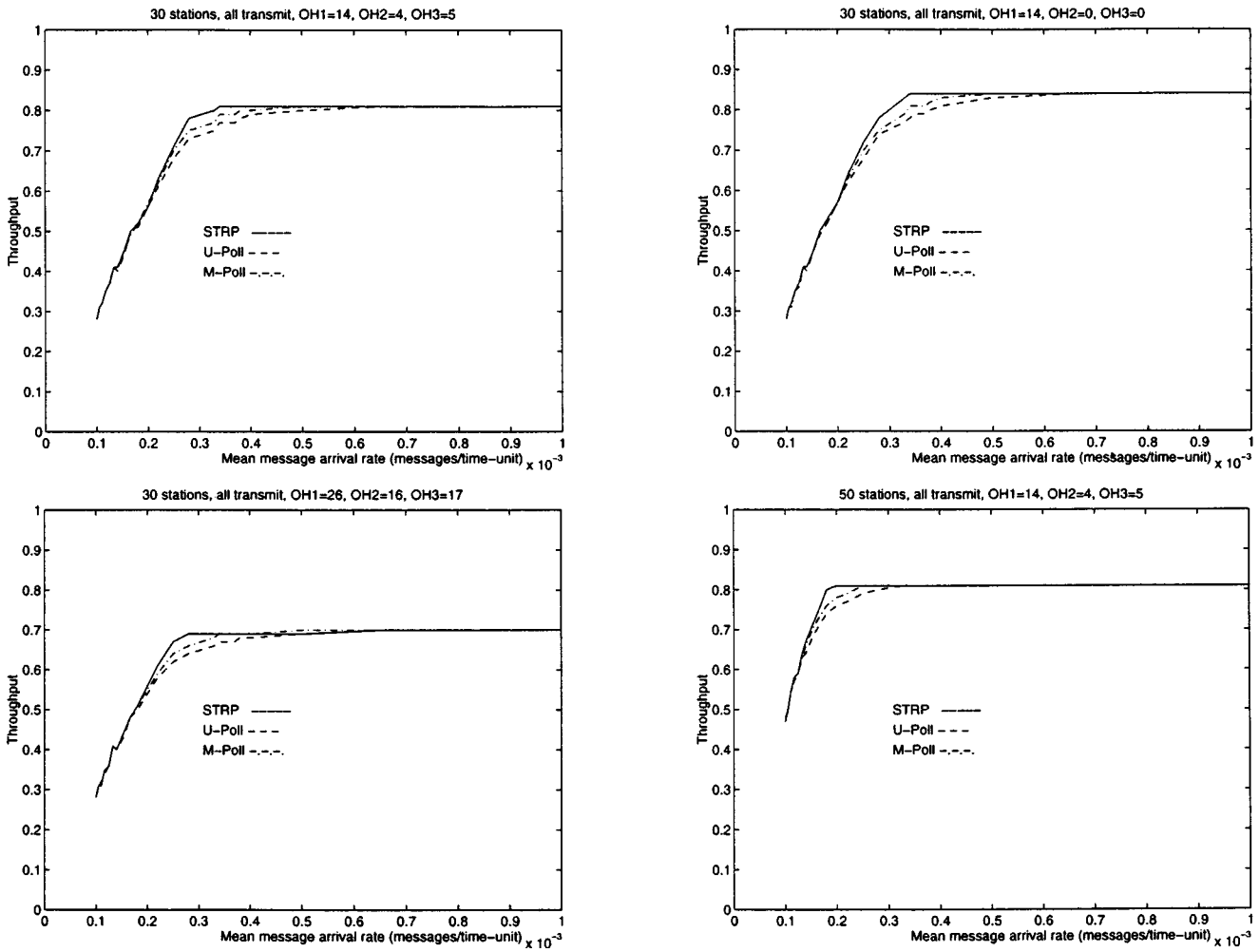

Fig. 13. Throughput versus message arrival rate for the case when all the stations transmit.

In both system models, all the stations are at the same, furthest distance from the BS, i.e., τ, and they all have infinite packet queues. As in Section V, we assume that the stations that generate and transmit packets in each of the above scenarios, generate fixed length information packets according to a Poisson arrival process. We measure three quantities: first, the throughput of the system, second, the mean access delay, i.e., the mean time that elapses from the time a packet arrives at the head of the transmission queue and until its transmission begins. Finally, we measure the mean queuing delay, i.e., the mean time that elapses from the time a packet arrives at the transmission queue and until its transmission ends.

In the simulation, we do not refer to specific transmission times of information and control packets, nor to specific processing times or propagation delays. We are only interested in the *relations* between these quantities. Therefore, we set the transmission time of an information packet to be 100 time units and then set OH1, OH2, and OH3 to different sizes accordingly.

The first set that we checked is (OH1, OH2, OH3) = (14, 4, 5). This set represents for example a channel of 500 kb/s, 500-bit information packets, 50-bit control packets, a processing time of 40 μs and a detection time of 10 μs. The detection time is relatively long. However, it can happen that the BS will listen to the Upstream link for a relatively long time before deciding if it indeed detects a *Jam* or not because of possible noise in the link. The propagation delay is negligible.

Clearly, the STRP scheme has the potential to outdo U- and M-poll when only part of the stations transmit or when the traffic load is moderate. Consider the graphs in Figs. 10 and 11, which show the average access delay versus the mean message arrival rate per station. In Fig. 10, only part of the stations generate and transmit packets, while in Fig. 11 all the stations transmit. First consider the upper, lefthand graph in these Figures. It corresponds to the case when only 10 stations, out of 30, transmit. STRP out-does U- and M-poll significantly because of the polling of the quiet stations, i.e., those stations that do not transmit generate a very low cost (detection time) in STRP because it is done together with the transmissions of the stations with packets. On the other hand, with U-poll, the polling of the quiet stations prohibits transmissions of information packets, and in M-poll, the notification of the quiet stations that do not have packets to transmit reduces the waste of time due to the polling of these stations.

In the case where all the stations generate packets and transmit (Fig. 11), STRP out-does the other schemes only when the traffic load is moderate. This is clear since, when the traffic is light, there are not many active stations, and so the polling is mainly done on the Idle ring, similar to the case in U- and

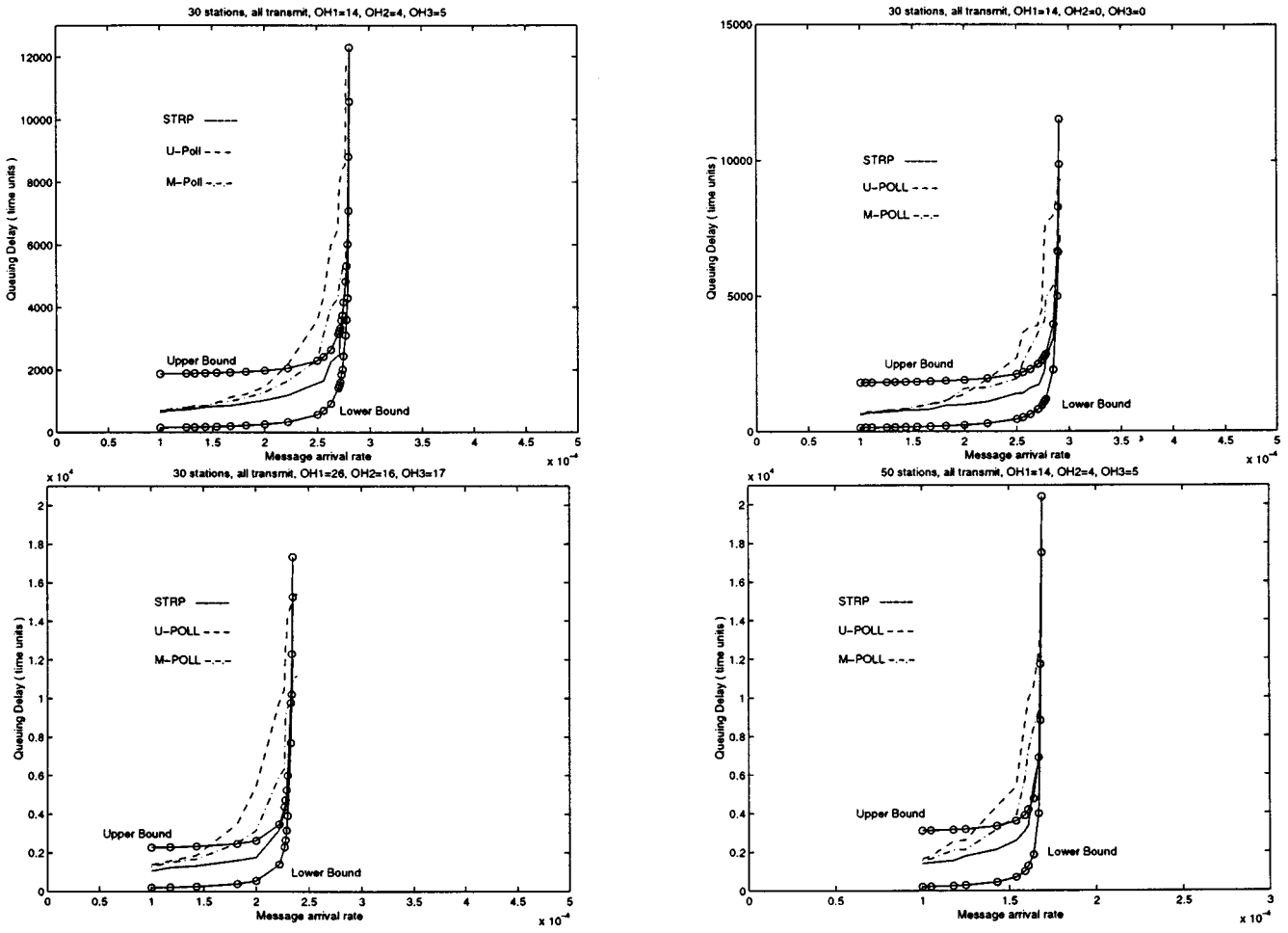

Fig. 14. Queueing delay versus message arrival rate for the case when all the stations transmit.

M-poll. When the traffic is heavy, most of the stations are in the Active ring, and in this case, the polling is mainly on the Active ring, again similar to the case in U- and M-poll. When the traffic is moderate, part of the stations are in the Idle ring and part are in the Active ring. Here, the polling of the stations in the Idle ring is meaningful because it does not delay stations with packets, in comparison to M- and U-poll where stations with packets are delayed.

Notice that when the system is overloaded, the average access delay is about 3300–3400 time units. This is because we assume that the transmission time of information packets is 100 time units. Thus, if the system is loaded, i.e., every station always has packets to transmit, a round of transmissions is 3000 time units long. Together with the overhead (transmission of control packets, processing time etc.) we receive the above mentioned average access delay of about 3300–3400 time units.

Finally, consider the corresponding graphs in Figs. 12 and 13 which show the throughputs of the schemes when only part of the stations transmit and when all the stations transmit respectively. The results show a similar relative behavior to that in Figs. 10 and 11.

We also checked the sets (OH1, OH2, OH3) = (14, 0, 0) and (OH1, OH2, OH3) = (26, 16, 17). Compared to the first set, in the set (14, 0, 0) the processing and detection times are negligible and equal to 0. In the set (26, 16, 17) the processing time is longer, and so is the length of the control packet. The relative detection time is smaller now. These changes do not change the relative results, as can be seen from the upper righthand graphs and the lower, lefthand graphs in Figs. 10–13.

Next, as mentioned, we also simulated a system with 50 stations. We assume in Figs. 10 and 12 that only 16 stations generate packets and transmit, while in Figs. 11 and 13, all 50 stations transmit. The results of this simulation appears in the lower right-hand graphs of every figure. We observe the same relative results as in the previous three examples. Notice that in the case when only one station needs to transmit in the system, STRP enables this station to transmit almost continuously, where its transmissions are separated only by the polling from the BS by the Q/T packets. This is similar to random access MACs such as IEEE 802.3 CSMA/CD, where a station can transmit almost continuously, with some separation between continuous transmissions in order to enable other stations the chance to transmit [10]. On the other hand, U- and M-poll do not enable this feature. In this case, the advantage of STRP is undoubtedly clear.

Notice also that STRP is never worse than U-poll because the only overhead that STRP has, compared to U-poll, is the detection time for a jam. However, this is done only when the BS polls a station and enables a transmission to another station in parallel. The saving of the *Query* packet transmission time in this case compensates for this overhead. Also notice that STRP per-

forms as U-poll when all the stations have packets to transmit, as is shown in Figs. 11 and 13, or when stations never have more than one packet to transmit when they are polled, and thus always stay in the Idle ring.

Finally, in Fig. 14 we show the queuing delay measures for the three schemes, together with the lower and upper bounds for STRP, as computed in Section V. These measures correspond to the case where all the stations transmit. Again, STRP outperforms U- and M-poll and the measured queuing delay for STRP falls between the lower and upper bounds curves.

VII. Conclusion

We proposed a new, efficient MAC protocol for Wireless LANs, referred to as STRP. STRP overcomes the inefficiency of the polling schemes when only a part of the stations transmit or when the traffic load is moderate. Also, STRP keeps all the good characteristics of the polling schemes, such as bounded access delay and the ability to implement priorities in an easy way.

Notice that in STRP, the BS knows which stations have packets to transmit. This has the potential for further improvement in the case where the information packets are of equal length, since the BS can give the right to transmit to several stations together, by using the same control packet. In that control packet, it also defines the order of transmissions. This ability can clearly enlarge the throughput of the system. However, it can delay the transfer of stations from the Idle to the Active ring. This idea shall be further investigated.

References

[1] K. C. Chen and C. H. Lee, "RAP—A novel medium access control protocol for wireless data networks," in *Proc. IEEE GLOBECOM*, 1993, pp. 1713–1717.

[2] ——, "Group randomly addressed polling for wireless data networks," in *Proc. IEEE ICC*, 1994, pp. 913–917.

[3] Z. Zhang and A. S. Acampora, "Performance of a modified polling strategy for broadband wireless LAN's in a harsh fading environment," in *Proc. GLOBECOM*, Phoenix, AZ, Dec. 1991, pp. 1141–1146.

[4] *Wireless LAN medium access control (MAC) and physical layer (PHY) specification*, IEEE Std. 802.11, 1997.

[5] J. L. Sobrinho and J. M. Brazio, "Proposal and performance analysis of a multiple-access protocol for high-speed wireless LANs," *Comput. Networks ISDN Syst.*, vol. 28, pp. 283–305, 1996.

[6] H. Ahmadi, A. Krishna, and R. O. LaMaire, "Design issues in wireless LANs," *J. High Speed Networks*, vol. 5, pp. 87–104, 1996.

[7] J. J. Metzner, "On improving utilization in ALOHA networks," *IEEE Trans. Commun.*, vol. COM-24, pp. 447–448, Apr. 1976.

[8] K. Mutsuura, H. Okada, K. Ohtsuki, and Y. Tezuka, "A New control scheme with capture effect for random access packet communications," in *Proc. IEEE GLOBECOM*, 1989, pp. 938–944.

[9] S. Kubota, K. Mutsuura, O. Akizuki, and S. Ooshita, "A random access micro-cellular sustem," *IEICE Trans. Fundamentals*, vol. E80-A, no. 7, pp. 1241–1248, July 1997.

[10] *Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications*, ANSI/IEEE Std. 802.3, 1998.

[11] B. Lathi, *Modern Digital and Analog Communication Systems*: Holt, Rinehart, and Winston, 1983.

[12] G. Jessop, *VHF/UHF Manual*, 4th ed: Radio Society of Great Britain, 1983, ch. 4.

[13] L. Roberts, "ALOHA packet system with and without slots and capture," *Comput. Commun. Rev.*, vol. 5, 1975.

[14] O. J. Boxma, "Workloads and waiting times in single-server systems with multiple customer classes," *Queueing Syst.*, vol. 5, pp. 185–214, 1989.

[15] A. Jean-Marie and Z. Liu, "Stochastic comparisons for queuing models via random sums and intervals," *Adv. Appl. Prob.*, vol. 24, pp. 960–985, 1992.

Oran Sharon received B.Sc., M.Sc., and Ph.D. degrees in computer science from the Technion-Israel Institute of Technology, Haifa, Israel, in 1986, 1989, and 1994, respectively.

In 1995, he was a Research Engineer at HP Laboratories, Bristol, U.K. Since 1996, he has been a Lecturer at the Mathematics and Computer Science Department,Haifa University, Haifa, Israel, and a Research Engineer at ECI Telecom, Petach Tikva, Israel. His current research work is in the area of MAC protocols, TCP/IP protocols, wireless LANs, ATM, and optical networks.

Eitan Altman received the B.Sc. degree in electrical engineering and the B.A. degree in physics in 1984 and the Ph.D. degree in electrical engineering in 1990, all from the Technion-Israel Institute, Haifa. In 1990, he received the B.Mus. degree in music composition from Tel-Aviv University, Tel Aviv, Israel.

Since 1990, he has been the National Research Institute in Informatics and Control (INRIA), Sophia-Antipolis, France. His current research interests include performance evaluation and control of telecommunication networks, stochastic control, and dynamic games. In recent years, he has applied control theoretical techniques in several joint projects with France Telecom.