Wireless Networks 8, 5–11, 2002

Multiple Working Points in Multichannel ALOHA with Deadlines *

DROR BARON ** and YITZHAK BIRK
Electrical Engineering Department, Technion–Israel Institute of Technology, Haifa 32000, Israel

Abstract. This paper addresses the problem of maximizing the capacity of multichannel slotted ALOHA networks subject to a user-specified deadline and a permissible probability of exceeding it. A previous paper proposed to transmit a non-decreasing number of copies of a message in successive rounds until success or deadline. This yielded a low probability of failure due to the large maximum number of copies per message, with only minimal "pollution" due to the small mean number of copies. In this paper, we examine another way of implementing variable resource expenditure in different rounds: the channels are partitioned into groups, one for each round, and the channels used in later rounds are operated with lower offered loads, i.e., at different "working points". The delay-constrained capacity with these Single-Copy Multiple-Working-Point (SC-MWP) policies is shown to be substantially higher than that with conventional ALOHA, but lower than with the optimal Multicopy Single-Working-Point (MC-SWP) scheme. Combining the two to form an MC-MWP scheme only slightly improves capacity over MC-SWP. The SC-MWP approach can be more attractive when using a single transmitter per station because, unlike with multiple copies, transmission time is not prolonged. Therefore, multiple-working-point policies become more attractive when propagation delay is lower (e.g., low orbit satellites).

Keywords: multichannel ALOHA, satellite networks, deadline, multiple working points

1. Introduction

1.1. Background

ALOHA [1] is the simplest access scheme because it does not require channel sensing or collision detection, but performs worse than more elaborate schemes when those are practical. An important use of ALOHA at present is by satellite ground stations, because the long propagation delay precludes timely channel sensing. It is used as the primary access scheme for short messages, and in order to reserve channels for long ones [2].

Another application of ALOHA is cellular networks, wherein the control uplink channels from the cellular phones to base stations are multiple access. A future application for ALOHA may be transmission of short messages over the "upstream" channels of high speed point-to-multipoint terrestrial wireless networks: a central base station could collect transmissions from a large number of users using shared bandwidth. Since round-trip propagation delay for terrestrial stations 10 kilometers apart is on the order of 0.1 milliseconds, even present data rates permit transmission of many thousands of bits during this time, precluding timely channel sensing.

Figure 1 depicts a typical satellite-based ALOHA network. The *stations* transmit data in globally synchronized time slots over contention uplink channels (dashed lines). Successful reception by the *hub* is acknowledged by it immediately over contention-free downlinks (solid lines). The hub can be terrestrial or in space. If several simultaneous transmissions occur over the same channel, they all fail. Stations can only learn about a collision through the absence of an acknowledgment. The time from the beginning of a transmission until the time by which an ACK for it must be received (or else it is considered to have collided) is referred to as a *round*. Unlike time slots, whose boundaries must be synchronized among the stations, a round is "private" and its starting time is decided by each station for itself. A station retransmits packets until they succeed or until a deadline is exceeded. The typical duration of a round is up to several tens of time slots.

In a single-channel ALOHA network, retransmission delay must be randomized to prevent definite repeated collisions [3]. To improve stability, a station must moreover increase the mean back-off time in later rounds. Current satellite networks, however, employ as many as hundreds of channels. When operated with ALOHA, e.g., for small transactions, a station picks a channel at random for each transmission. The hub can receive concurrently over all channels. The randomized retransmission delay is replaced with immediate retransmission over a randomly chosen channel.

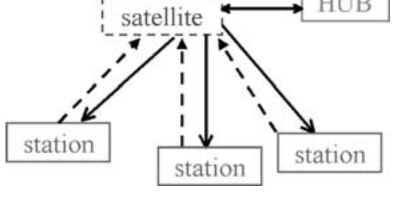

Figure 1. A typical hub-based satellite network.

* This work was supported in part by the Information Superhighway In Space consortium, administered by the office of the Chief Scientist of the Israeli Ministry of Industry and Trade.

** Presently with the ECE Department, University of Illinois at Urbana-Champaign.

Over the years, the bulk of the research on ALOHA and related reservation schemes, e.g. [3], concerned maximizing capacity. Some attention was given to delay-throughput trade-offs and other performance measures. The advent of multi-channel ALOHA networks has given rise to the use of redundant transmissions for performance improvement. For example, [4] studies *multicopy ALOHA*, whereby a station transmits several *copies* of a packet in each round, as a way of improving delay–throughput performance. We refer to the transmission of multiple copies per round as "redundancy" because, unlike retransmission upon failure, some of the transmissions may not be required.

1.2. Delay-constrained capacity

Virtually all current applications of ALOHA entail the transmission of single packets, be it for short transactions or to reserve channel resources for the transmission of large amounts of data. Also, the user is typically charged per actual traffic, while the system owner pays for bandwidth (channel) resources. From a user perspective, the key performance criterion is delay, and it is most naturally expressed as a constraint (e.g., deadline). From the system owner's perspective, capacity maximization is the main design goal.

Recently, Birk and Keren [5] proposed an optimization problem that reflects both intuitive user requirements and the desires of network designers: maximization of capacity subject to a deadline and a permissible probability of exceeding it. They proposed a *non-stationary multicopy* transmission policy, whereby a station transmits a monotonically non-decreasing number of copies in successive rounds until successful reception or deadline. Dynamic programming was used to optimize the transmission sequence, resulting in a substantial increase in capacity relative to that attainable with classical ALOHA or even with (fixed) multi-copy ALOHA [4]. The advantage is more pronounced for stricter constraints. They moreover adapted the optimized scheme to the practical situation wherein a station only has a single transmitter that must emulate multiple transmitters. This was done by transmitting a burst of copies in successive time slots over randomly chosen channels, and then waiting to learn their fates before proceeding to the next round. The technique was dubbed *Round Stretching* because the serialized transmission of the copies, along with the need to wait until the fate of the last copy is known before commencing a new round, prolongs the round. Figure 2 illustrates the idea. It can be seen

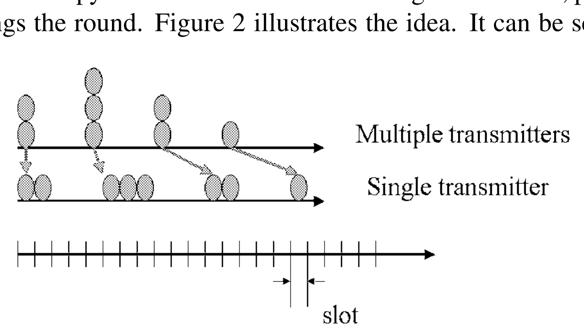

Figure 2. Round Stretching.

that Round Stretching may reduce the permissible number of rounds for a given deadline. Round Stretching was shown to achieve similar capacities to the multi-transmitter scheme in most situations, due mostly to the fact that round-trip propagation time is typically much longer than the duration of a time slot.

The non-stationary use of replication was extended in [5] to general erasure correcting codes for the case of single-round transmissions of multi-slot messages. In [6], such codes are considered for multi-round transmissions, as is the combination of such codes with reservations. The *Coding–Reservation* scheme of [6] raises capacity above $1/e$ by using a received message fragment to also request contention-free channels for the remaining fragment(s). The discussion in this paper is restricted to single-slot messages.

The key idea in the replication-based scheme of [5], which is also employed in this paper, is to permit a large maximum channel-resource expenditure per message while keeping the mean expenditure low. This is done by being more "wasteful" in the later rounds, which are less likely to be required because transmission of a message ceases upon its successful reception. The probability of failure can thus be made very low because a message fails only after the maximum amount of resources has been spent on it; yet, the average resource expenditure per message can also remain low because a message succeeds in the first round with high probability. In [5], the expenditure manifested itself as speculative transmission of multiple copies in late rounds.

In this paper, we propose and study an alternative way of controlling the resource expenditure: the channels are partitioned into groups, one per permissible transmission round, with lower offered loads on the channels used for later rounds. This approach is dubbed *Multiple Working Points* (MWP). We begin by comparing the single-copy-per-round, multiple-working-point (SC-MWP) scheme with the multi-copy, single-working-point (MC-SWP) scheme of [5]. Then, we explore the additional benefit of combining the two approaches (MC-MWP).

The remainder of the paper is organized as follows. In section 2, we present the multiple-working-point approach and the policy design space. In section 3, we present the network model that is subsequently used for performance analysis, and derive some preliminary mathematical relations. Section 4 proves that, when using a deterministic number of copies per round, each round should best use a single working point. Section 5 provides performance analysis of MC-MWP policies, of which SC-MWP is a special case. Section 6 presents numerical results, and section 7 offers concluding remarks.

2. Multiple Working Points (MWP)

2.1. Multiple-Working-Point schemes

In [5], later rounds received preferential treatment by transmitting more copies in them over randomly chosen (distinct) channels, thereby increasing the probability of success of the round. The cost was "pollution", since transmitting k copies meant that k channels were prevented from carrying any other transmission successfully, and at most one distinct successful reception could occur over them. The scheme was nonetheless very effective because the probability of reaching a late round is very low.

In this paper we explore a different method of offering preferential treatment to the later rounds, namely dedicating to them a subset of the channels, and keeping those lightly loaded. Thus, even if only one copy of a message is transmitted per round until successful reception or deadline expiration, the probability of success in later rounds is higher. Referring to the offered load on a channel as its *working point*, we refer to this scheme as *Multiple Working Points* (MWP).

MWP is implemented as follows. The channels are partitioned into groups, one per round, and transmissions belonging to any given round are carried over channels that are chosen randomly from among those of the relevant group. The selection of the number of channels for each group and its working point is the result of an optimization, and will be discussed later.

The offered load on a channel is the mean number of transmissions per time slot over it. A single transmission over a lightly loaded channel therefore constitutes a larger fraction of its offered load than a single transmission over a heavily loaded channel, and thus consumes more channel resources than the latter. Consequently, even if a single copy is transmitted in each round, later-round attempts consume more channel resources than those in earlier rounds.

The multi-copy scheme of [5] and MWP represent different mechanisms for implementing the same idea. Our purpose in this paper is to compare their effectiveness and to see whether combining them gives substantial additional benefit.

2.2. Design space

Multi-channel ALOHA access schemes for single-slot messages can be classified along several dimensions:

- *Single/multiple copies* per round (SC/MC).
- *Single/multiple working points* (SWP/MWP).
- *Stationary/non-stationary*. A stationary policy acts identically in every round, whereas non-stationary ones are round-dependent.
- *Pure/impure* [7]. Pure policies are deterministic. Impure policies can entail a probabilistic choice of the number of copies in any given round (with a possibly different probability distribution for each round), as well as a probabilistic choice among several working points in each round (with possibly different probabilities for each round). The randomized selection of a channel among those in the same group does not constitute impurity.

  Impure stationary policies were studied in [7] in the context of optimizing the throughput–delay trade-off with multi-channel ALOHA. An impure variant of the replication-based scheme of [5] was studied [6], and was shown to only yield a negligible increase in capacity.

Having described the various policies and the rationale behind them, we next proceed to optimize them and evaluate their performance.

3. Network model and preliminaries

3.1. Model and definitions

The network comprises ground stations that transmit single-slot messages over randomly chosen channels. A hub monitors all channels and ACKs all successful receptions. The lack of an ACK when it is expected indicates a collision. A station transmits in rounds, waiting for the results of one round before continuing to the next. The duration of a round is thus the sum of transmission time, round-trip propagation delay and any processing delays. A typical round length is 20 time slots. A station ceases to transmit a message upon success or when a deadline expires. Upon reaching the deadline, an as-yet unreceived message is declared lost. (We will consider very small permissible loss probabilities, so "lost" messages may be reissued with a negligible effect.)

We assume an infinite number of stations and a large number of channels. The number of transmissions over any given contention channel in any given time slot is modeled as a Poisson random variable, independent from slot to slot and from channel to channel. With these assumptions, the probability of collision of a packet is only a function of the offered load on the channel over which it is transmitted. (Simulations [6] have shown this approximation to result in a capacity that is higher by a few percents than the true capacity when there are 100 channels per working point, and by some 10 percent with 30 [6]. With Round Stretching, the approximation is even closer.) Finally, the finite number of channels affects competing schemes in a similar manner, so the effect on comparative results is substantially lower.

3.2. Stability

The original ALOHA access scheme is unstable, because a temporary increase in the message generation rate can reduce throughput to well below the mean generation rate. The resulting retransmissions of backlogged packets along with the transmission of new ones cause throughput to drop to zero. Much research has studied this issue along with schemes to rectify the situation, e.g., [8].

In our case, the access scheme is never unstable, provided that the mean packet generation rate is below capacity, because messages that miss the deadline are discarded by the sender. However, it is bistable in certain load ranges: a momentary increase in packet generation rate can push the network into the negative-slope range of throughput versus offered load, resulting in a higher missed-deadline probability than possible for a given throughput. Nonetheless, because the backlog has a "short memory" and cannot accumulate, a momentary underload brings the network back to the efficient operating range. Moreover, a network hub can detect such situations and "push" the network into the "good" operating region. Our focus in this paper is on the potential of novel access schemes to improve performance. The details of handling bistability with the new schemes are left for future research or implementation. The analysis in the rest of this paper applies to periods during which the network is in the good operating range.

3.3. Delay constraints

A user-specified deadline is expressed in time units. For facility of exposition, we define this to be the time from the first transmission until the time of the latest transmission that would still arrive by the deadline. With fixed-size time slots, D_s denotes the deadline in time slots. D_r denotes the maximum permissible number of rounds. P_e denotes the permissible probability of missing the deadline.

When Round Stretching [5] is used, let T_A denote the number of time slots from single-slot transmission until an ACK is received or its absence indicates that the next round may begin. (T_A represents the propagation delay and processing time.) Then,

$$N_{\max} \leqslant D_s - (D_r - 1)T_A, \tag{1}$$

where $N_{\max}$ is the maximum total number of transmitted copies of any given message. (Note that D_s, being the actual delay constraint, is the independent variable. Given D_s and T_A, there is a trade-off between the number of rounds and the maximum number of copies per message, and the transmission policy optimization is charged with optimizing this trade-off.) When $T_A \gg 1$, D_r is not affected much by $N_{\max}$, and Round Stretching hardly changes performance. For small T_A, the effect varies.

3.4. Useful relations

This paper deals with multi-channel ALOHA. However, for convenience and in order to facilitate the use of well-known single-channel relationships, all variables and parameters in this paper are per-channel unless stated otherwise.

Delay-constrained operation implies that messages may be dropped, albeit with a low probability. A distinction was therefore made in [5] between the generation rate of messages, S_g, and the throughput S. Specifically,

$$S = (1 - P_e)S_g. \tag{2}$$

For pure MC-SWP policies and single-slot messages [5],

$$G = S_g \cdot E[N], \tag{3}$$

where G denotes offered load and $E[N]$ denotes the mean number of transmitted copies per message until success or deadline. Capacity is thus

$$S = S_g(1 - P_e) = \frac{G(1 - P_e)}{E[N]}. \tag{4}$$

The total number of copies transmitted per message is $N = \sum_i n_i \leqslant \sum_{i=1}^{D_r} n_i \leqslant N_{\max}$, where n_i denotes the number of copies transmitted in round i. The probability of collision is $P_c = 1 - e^{-G}$. Since $P[\text{reach round } i] = (P_c)^{\sum_{j=1}^{i-1} n_j}$,

$$E[N] = n_1 + \sum_{i=2}^{D_r} n_i (P_c)^{\sum_{j=1}^{i-1} n_j}. \tag{5}$$

For systems with multiple channel groups operating at different working points, we define the capacity as the aggregate (summed up over the channels of all groups) capacity, divided by the total number of channels. Also, though new messages are only generated over the channels of the first-round group, we define the mean per-channel generation rate as the aggregate generation rate of distinct messages over the channels of the first-round group, divided by the total number of channels. Consequently, the relation $S = S_g(1 - P_e)$ remains valid for the system as a whole.

Consider channel-group m with W_m channels and offered load G_m (per channel). The aggregate load of this group is then

$$x_m = W_m \cdot G_m. \tag{6}$$

Also, ignoring integer constraints on W_m, an incremental increase of x_m, say by Δx, would increase W_m by

$$\Delta W_m = \frac{1}{G_m} \cdot \Delta x. \tag{7}$$

4. Optimality of a single working point per round

Numerical results have shown that, with multi-round MC-SWP policies, a probabilistic number of copies can increase capacity, but the increase is minute; also, for a single round, a pure policy appears to be optimal [6]. In view of this, we only consider a deterministic number of copies in each round, denoted n_i, and prove that it is best to use a single working point for each round.

Theorem 1. Any optimal policy that transmits a deterministic number of copies in each round and at each working point used by it uses a single working point for any given round.

Proof. Recalling the assumptions that were made, whereby the fate of a transmitted copy is only influenced by the offered load (working point) of its channel, it suffices to consider a single round. The proof is by contradiction.

Consider two copies of a packet, transmitted over channels belonging to different groups (different working points); the probabilities of collision are denoted by P_{c_1} and P_{c_2}, respectively. We will show that transmitting them at a single working point, such that the probability that both collide is unchanged (equals $P_{c_1} \cdot P_{c_2}$), would reduce channel resource requirements.

Let Δx denote the (absolute, not per-channel) additional load corresponding to the transmission of a single copy. Consider two channel groups whose channels operate at working points (offered loads) G_1 and G_2, respectively. It follows from (7) that the total number of additional channels required in "support" of the transmission of a single copy in each of the channel groups (one per group) is

$$\Delta W = \frac{\Delta x}{G_1} + \frac{\Delta x}{G_2}. \tag{8}$$

The same probability of error could instead be obtained by transmitting two copies at a single working point with collision probability $\widetilde{P}_c = \sqrt{P_{c_1} P_{c_2}}$. According to (7), $\Delta\widetilde{W} = 2\Delta x/\widetilde{G}$ additional channels would be required, where $\widetilde{G}$ is the offered load corresponding to $\widetilde{P}_c$. It suffices to show that $\Delta\widetilde{W} \leqslant \Delta W$.

Define $R(P_c) \triangleq -\ln P_c$, then $R(\cdot)$ is an additive measure, and the channel requirement ΔW can be calculated using R as

$$\Delta W(R) = -\frac{\Delta x}{\ln(1 - e^{-R})}. \tag{9}$$

It is clear that $2 \cdot \widetilde{R} = R_1 + R_2$, hence it suffices to show that $\Delta W(R)$ is convex. It suffices to show that $G(R) = 1/\Delta W(R)$, the offered load on channels with some R, is convex. But

$$\frac{\partial^2 G}{\partial R^2} = \frac{e^{-R}}{\Delta x (e^{-R} - 1)^2} > 0. \tag{10}$$

Therefore, the channel requirement is convex in R, so merging the two channel groups (working points) into a single group with a working point whose probability of collision is equal to $\widetilde{P}_c$ and transmitting the two copies over channels of the consolidated group would lower channel requirements. □

5. Capacity of pure MC-MWP schemes

Based on theorem 1, we use a single working point, WP_i, in each round, with an offered load G_i and probability of collision P_{c_i}. We use n_i to denote the number of copies transmitted in round i. The probability of a message failing to meet the deadline is

$$P_e = \prod_{i=1}^{D_r} (P_{c_i})^{n_i}. \tag{11}$$

The generation rate for a channel operating at WP_i is

$$S_{g_i} = \frac{G_i}{n_i}. \tag{12}$$

Consider W_1 channels used for WP_1 (first round). The aggregate generation rate of distinct messages in the network is $S_{g_1} \cdot W_1$. The rate of messages entering round 2 is $S_{g_1} \cdot W_1 \cdot (P_{c_1})^{n_1}$. However, it is also equal to $S_{g_2} \cdot W_2$, so

$$W_2 = W_1 \frac{S_{g_1}}{S_{g_2}} (P_{c_1})^{n_1}. \tag{13}$$

Similarly, $S_{g_1} \cdot W_1 \cdot \prod_{k=1}^{i-1} (P_{c_k})^{n_k}$ messages enter round i, so

$$W_i = W_1 \frac{S_{g_1}}{S_{g_i}} \prod_{k=1}^{i-1} (P_{c_k})^{n_k}, \quad i \geqslant 2. \tag{14}$$

Noting that new messages are only generated over channels of WP_1, the overall mean (over all network channels) per-channel generation rate of distinct messages is

$$S_g = \frac{W_1 \cdot S_{g_1}}{\sum_{i=1}^{D_r} W_i} = \frac{W_1 S_{g_1}}{W_1 + \sum_{i=2}^{D_r} W_1 (S_{g_1}/S_{g_i}) \prod_{k=1}^{i-1} (P_{c_k})^{n_k}}. \tag{15}$$

According to (12) and purity in round i, and since $P_{c_i} = 1 - e^{-G_i}$,

$$\frac{1}{S_g} = \frac{1}{S_{g_1}} + \sum_{i=2}^{D_r} \frac{1}{S_{g_i}} \prod_{k=1}^{i-1} (P_{c_k})^{n_k} = \frac{n_1}{G_1} + \sum_{i=2}^{D_r} \frac{n_i}{G_i} \prod_{k=1}^{i-1} \left(1 - e^{-G_k}\right)^{n_k}, \tag{16}$$

and the capacity is $S = S_g(1 - P_e)$.

6. Numerical results

In order to compare the performance of MWP policies with SWP policies, a computer program that, given (n_i) and P_e, optimizes $\{G_i\}$ according to (16), was written. This was combined with an exhaustive search over (n_i) for the best values. (Note that, because of the independence assumptions and the derivation of results per channel, the total number of channels does not matter.)

We first consider the case of an unlimited number of transmitters per station (but note that even then the required number of transmitters is not large). We begin by comparing the use of multiple working points (SC-MWP) with the conventional multichannel ALOHA (SC-SWP). Next, we compare the two mechanisms for controlling resource expenditure, namely multiple working points (SC-MWP) and multiple copies (MC-SWP) [5] and offer some insights. Then, we discuss the MC-MWP hybrid. Finally, results are presented for the practical single-transmitter situation, in which Round Stretching is employed whenever there are multiple copies per round.

6.1. Multiple transmitters per station

The results are summarized in table 1.

Single-copy policies. Columns 3 and 6 depict the performance of SC-SWP and SC-MWP for several (P_e, D_r) constraints. The use of multiple working points is seen to yield a major performance boost.

Table 1
The capacity of MWP and SWP policies.

D_r	P_e	SWP			MWP		
		SC	Optimal MC		SC	Optimal MC	
		S	(n_i)	S	S	(n_i)	S
3	10^{-2}	0.190	1, 2, 4	0.279	0.233	1, 2, 4	0.281
	10^{-3}	0.095	2, 3, 7	0.247	0.158	1, 2, 6	0.248
	10^{-4}	0.045	2, 3, 10	0.233	0.110	2, 3, 9	0.234
5	10^{-2}	0.306	1, 1, 1, 2, 3	0.340	0.335	1, 1, 1, 2, 2	0.342
	10^{-3}	0.217	1, 1, 1, 2, 5	0.321	0.296	1, 1, 1, 2, 5	0.324
	10^{-4}	0.145	1, 1, 2, 3, 8	0.313	0.264	1, 1, 2, 3, 7	0.314

Multiple copies vs. multiple working points. In columns 5 and 6, we see that MC-SWP [5] outperforms SC-MWP, so controlling the number of copies per round is superior to controlling the working points. This can be explained as follows.

With optimal MC-SWP policies [5], $P_e = (P_c)^{N_{max}}$, so

$$N_{max} = \frac{\ln P_e}{\ln P_c}. \tag{17}$$

Given a working point, N_{max} thus increases logarithmically with a decrease in P_e. Moreover, a message does not always utilize all the rounds, so $E[N]$ is even smaller and thus increases at most logarithmically with P_e. The cost, in terms of channels, needed to maintain a low error probability, is not very high.

With SC-MWP policies, the offered load on late-round channels is low, so

$$G \approx 1 - e^{-G} = P_c, \qquad G \ll 1. \tag{18}$$

According to (6), the number of channels required for each "late" round is therefore roughly inversely proportional to the probability for collision in that round. The cost, in terms of channels, needed to maintain a low error probability, is quite significant.

MC-MWP policies. Permitting the use of multiple working points in conjunction with the optimal MC-SWP transmission sequence (n_i) of [5] slightly increases capacity. Joint optimization of (n_i) and the working points yields an additional small increase in capacity. Derivation of an optimal MC-MWP policy requires a search over the range of (n_i) in conjunction with the equations derived in the previous section.

Columns 7 and 8 of table 1 show the optimal transmission sequence and the resulting capacity of MC-MWP. The improvements in capacity over that of MC-SWP (column 5) are below 1%. Moreover, the effect of a finite number of channels and their subdivision into groups may more than offset this advantage. Consequently, if capacity is the main design goal, MWP policies may not justify the higher implementation complexity.

6.2. MWP with Round Stretching

With Round Stretching, the permissible number of rounds D_r and the maximum number of copies N_{max} are related through (1). Consequently, especially for values of D_s that barely permit another round, one must decide whether to increase D_r at the cost of significantly reducing N_{max} or stay with one fewer round and slightly increase N_{max}.

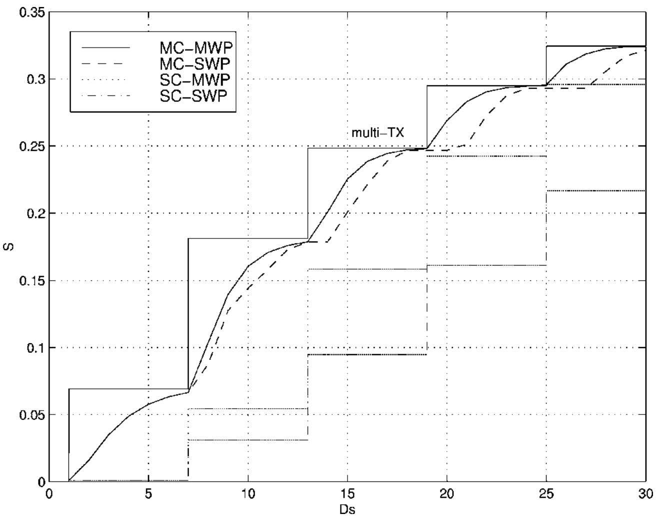

Figure 3. Capacity with Round Stretching. $P_e = 10^{-3}$; $T_A = 5$.

Figure 3 depicts capacity versus D_s with Round Stretching for the policies considered in this paper. Results for MC-MWP with an unlimited number of transmitters and for classical ALOHA (SC-SWP) are shown for reference. As D_s is increased and permits an additional round, the capacity with multiple transmitters rises immediately, whereas that with Round Stretching stays flat until such value of D_s for which an increase in D_r is warranted. Then, capacity rises sharply and approaches that with multiple transmitters per station. The MWP policies, due to their ability to use "clean" last rounds rather than increasing the number of copies and stretching the round, cope well with constraints on N_{max}. Therefore, when optimized, they elect to use an additional round earlier (at smaller values of D_s) than MC-SWP. For this reason, MWP policies can save several time slots of delay in Round Stretching. Note that (with $T_A = 5$) MC-MWP and MC-SWP policies have identical performance up to $D_s = 7$, because a single round is used and the same (single) WP is thus chosen. Also, observe that capacity rises quickly with an increase in the number of rounds. Finally, though not shown in the figure, the performance with Round Stretching becomes closer to that with multiple transmitters as propagation delays become longer (large T_A).

MC-MWP policies only provide slightly higher capacity than MC-SWP policies. However, the optimal use of MC-MWP requires a smaller N_{max}. With Round Stretching, it follows from (1) that reducing N_{max} for any given number of rounds permits a reduction in the permissible delay without altering performance. Alternatively, if the permissible delay is fixed, the permissible number of rounds may increase. It is therefore useful to evaluate the reduction in N_{max} brought about by using MC-MWP relative to the optimal MC-SWP policy for a given value of D_r and P_e while attaining at least the same capacity as the latter policy [5]. Table 2 shows that

Table 2
MWP time slot savings vs. optimal MC-SWP policies [5].

D_r	P_e	MC-SWP			SC-MWP			Savings
		(n_i)	N_{max}	S	(n_i)	N_{max}	S	
3	10^{-2}	1, 2, 4	7	0.279	1, 2, 3	6	0.280	1
	10^{-3}	2, 3, 7	12	0.247	1, 2, 5	8	0.247	4
	10^{-4}	2, 3, 10	15	0.233	2, 3, 7	12	0.233	3
5	10^{-2}	1, 1, 1, 2, 3	8	0.340	1, 1, 1, 1, 2	6	0.341	2
	10^{-3}	1, 1, 1, 2, 5	10	0.321	1, 1, 1, 2, 3	8	0.323	2
	10^{-4}	1, 1, 2, 3, 8	15	0.313	1, 1, 2, 2, 5	11	0.313	4

MWP policies can provide significant savings in N_{max}. When stricter delay constraints are used, N_{max} rises (for both policies), as does the savings in time slots.

Remark. An increase in N_{max} when D_s is reduced appears to contradict (1), but does not. Rather, reducing D_s while requiring the same P_e also forces a reduction in the number of rounds D_r. Furthermore, the capacity goes down, and the high N_{max} reduces the offered loads.

7. Conclusions

Capacity maximization of multi-channel slotted ALOHA networks for simgle-slot messages subject to a deadline and a permissible probability of failing to meet it is a goal that faithfully represents user requirements and designer goals for the current uses of ALOHA. This paper explored the use of different working points in different rounds as a means of implementing a non-stationary expenditure of network resources in order to achieve low probabilities of failure while holding down the mean per-message resource expenditure. Through numerical results as well as some analytical insight, this Multiple-Working-Point approach was shown to be significantly superior over the conventional SC-SWP approach. However, it is generally inferior to controlling the number of copies per round. An MC-MWP hybrid offers only a slight advantage when a station is equipped with multiple transmitters, but this advantage increases in the case of a single transmitter per station and Round Stretching, especially when the permissible delay and the permissible probability of failure are small. Smaller propagation delays (fewer time slots per round) that still preclude efficient channel sensing, e.g., low orbit satellites, also increase the attractiveness of using multiple working points.

Directions for future research include the use of MWP policies for multislot messages, the combination of Coding–Reservation schemes [6] with MWP policies, and multiple service categories. Yet another direction is consideration of stability within the core access scheme.

Finally, we note that the results of this paper serve as yet another example of the benefits gained from the judicious use of redundancy for performance enhancement.

References

[1] N. Abramson, Development of the ALOHANET, IEEE Transactions on Information Theory 31(3) (March 1985) 119–123.

[2] R. Rom and M. Sidi, *Multiple Access Protocols* (Springer-Verlag, New York, 1990).

[3] S.S. Lam, Packet broadcast networks – A performance analysis of the R-ALOHA protocol, IEEE Transactions on Computers C-29(7) (July 1980) 596–603.

[4] E.W.M. Wong and T.S.P. Yum, The optimal multicopy Aloha, IEEE Transactions on Automatic Control 39(6) (June 1994) 1233–1236.

[5] Y. Birk and Y. Keren, Judicious use of redundant transmissions in multichannel ALOHA networks with deadlines, IEEE Journal on Selected Areas in Communications 17(2) (February 1999) 257–269.

[6] Y. Birk and D. Baron, Capacity maximization in multichannel slotted ALOHA with deadlines – an overview, EE Technical Report No. 1248 (also CC-pub 314), Technion, Haifa, Israel (June 2000).

[7] Y.W. Leung, Generalised multicopy ALOHA, Electronics Letters 31(2) (January 1995) 82–83.

[8] W.A. Rosenkrantz and D. Towsley, On the instability of the slotted ALOHA multiaccess algorithm, IEEE Transactions on Automatic Control AC-28 (October 1983) 994–996.

Dror Baron received the B.Sc. (summa cum laude) and M.Sc. degrees from the Technion–Israel Institute of Technology, in 1997 and 1999, both in electrical engineering. From 1992 to 1994, he developed software at the Israel Defense Forces. In 1997 he was a teaching assistant in the Electrical Engineering Department at the Technion. From 1997 to 1999 he worked at Witcom Ltd. in modem design. Since 1999 he has been a research assistant at the University of Illinois at Urbana-Champaign, where he is pursuing his Ph.D. His research interests include information theory, data compression, communications systems, signal processing, and hardware design.
E-mail: dbaron@uiuc.edu

Yitzhak Birk has been on the faculty of the Electrical Engineering Department at the Technion since October 1991, and heads its Parallel Systems Laboratory. He received the B.Sc. (cum laude) and M.Sc. degrees from the Technion in 1975 and 1982, respectively, and the Ph.D. degree from Stanford University in 1987, all in electrical engineering. From 1976 to 1981, he was a project engineer in the Israel Defense Forces. From 1986 to 1991, he was a Research Staff Member at IBM's Almaden Research Center, where he worked on parallel architectures, computer subsystems and passive fiber-optic interconnection networks. From 1993 to 1997, he also served as a consultant to Hewlett Packard Labs in the areas of storage systems and video servers. Dr. Birk's research interests include computer systems and subsystems, as well as communication networks. He is particularly interested in architectures for information systems, including communication-intensive storage systems (e.g., multimedia servers) and satellite-based systems, with special attention to the true application requirements in each case. The judicious exploitation of redundancy for performance enhancement in these contexts has been the subject of much of his recent work. Dr. Birk is a member of the Institute of Electrical and Electronics Engineers.
E-mail: birk@ee.technion.ac.il