On the Merits of Impure Multi-Copy Schemes for Multichannel ALOHA with Deadlines*

Dror Baron† and Yitzhak Birk
Electrical Engineering Department
Technion — Israel Inst. of Technology
Haifa 32000, Israel
birk@ee, drorb@tx.technion.ac.il
June 11, 2000

Abstract

Slotted multichannel ALOHA is used in many satellite-based networks for short messages and as a means of reserving channels for longer ones. For such uses, maximization of channel capacity subject to a user-specified deadline and a (small) permissible probability of exceeding it captures both the user requirements and the system owner’s desires. An optimized multi-copy scheme, whereby a monotonically non-decreasing deterministic number of copies is transmitted in each round until success or deadline, has recently been shown to dramatically increase the deadline-constrained capacity. We show that permitting the number of copies transmitted in each round to be non-deterministic can further improve capacity, but only slightly.

Keywords: Multichannel ALOHA, Multi-copy ALOHA, satellite networks, delay-constrained capacity, impure access schemes.

*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. EE TR 1249 (also CCIT TR 315), June 2000, Technion.
†Presently with the ECE Dept., Univ. of Illinois at Urbana-Champaign, dbaron@ifp.uiuc.edu.
1 Introduction

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]. Other applications include control in cellular networks as well as reservation-making in other terrestrial contention-based systems.

With single-channel ALOHA, retransmission delay (upon failure) must be randomized in order to prevent definite repeated collisions [3]. With multi-channel ALOHA and a large number of channels, the retransmission delay is replaced with a randomized choice of channel for each transmission attempt.

The advent of multichannel ALOHA networks has given rise to the use of redundant transmissions for performance improvement. For example, [4] studies Multicopy ALOHA (MC), 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.

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 deterministic non-stationary Multicopy (MC) 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) MC 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. This was done by transmitting a burst of copies in successive slots over randomly chosen channels, and then waiting for ACKs for all of them before proceeding to the next round. This technique, dubbed Round Stretching, was shown to achieve similar capacities to the multi-transmitter scheme in most situations. (See [6] for an overview of recent work on capacity maximization in multi-channel ALOHA with deadlines.)

One can use pure MC policies, whereby the number of copies transmitted in any given round is deterministic (albeit not the same for all rounds), or impure policies whereby it is randomized. This idea was studied in [7] in the context of optimizing the throughput–delay trade-off with MC ALOHA. The scheme in [5] is pure. In this paper, we explore the merits of relaxing the requirement that the policy be deterministic, and see what if anything can be gained by permitting impure policies.
2 Network model and preliminaries

2.1 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, until the deadline; then, an as-yet unreceived message is declared lost. The number of permissible rounds and the permissible probability of missing the deadline will be denoted by $D_r$ and $P_e$, respectively. (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 [8] have shown this approximation to result in a capacity that is higher by a few percents than true capacity when there are 100 channels per working point and by some 10 percent with 30. With round stretching, the results are even closer. Finally, the finite number of channels affects competing schemes in a similar manner, so the effect on comparative results is substantially lower.

We assume the network to be stable. In practice, ALOHA access schemes can become unstable, in which case an external monitoring mechanism is assumed to detect the problem and rectify it. Stability of ALOHA has been studied, for example in [9].

Because messages may be dropped, albeit with a low probability, a distinction was made in [5] between the generation rate of messages, $S_g$, and the throughput $S$. Specifically, $S = (1 - P_e)S_g$.

2.2 Useful relations

This paper only addresses access schemes whereby all channels operate with the same offered load. These are referred to as single working point (SWP) schemes. For pure multi-copy single-working-point (MC-SWP) policies and single-slot messages [5],

$$G = S_g \cdot E[N], \quad (1)$$

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

$$S = S_g(1 - P_e) = \frac{G(1 - P_e)}{E[N]}, \quad (2)$$
The total number of copies transmitted per message is \( N = \sum_i n_i \leq \sum_{i=1}^{D_i} n_i \leq N_{\text{max}} \), where \( n_i \) denotes the number of copies transmitted in round \( i \). The probability of collision is \( P_c = 1 - e^{-G} \). Since \( P[\text{reach round } i] = (P_c)^{\sum_{j=1}^{i-1} n_j} \),

\[
E[N] = n_1 + \sum_{i=2}^{D_i} n_i (P_c)^{\sum_{j=1}^{i-1} n_j}.
\] (3)

### 3 Impure MC-SWP policies

We consider multi-copy, single working point policies for single-slot messages. Impure policies entail the transmission of a non-deterministic number of copies in each round. An impure policy can be considered as a quantization of a policy using real numbers of copies. It can also be viewed as an interpolation among pure policies. One mathematical representation considers a set of pure policies \( \{\alpha_k\} \) chosen among with probabilities \( \{p_k\} \) upon message generation. Another representation assumes that, being in round \( i \) after having transmitted \( j \) unsuccessful copies in previous rounds, the probability for transmitting \( k \) copies in the current round is \( p_{i,j,k} \). These representations are equivalent, because any transmission trajectory is possible. It should be noted that, because all channels are assumed to operate with the same offered load, and because we assume independence among the fates of different transmissions, the assignment of specific channels to each of the pure policies that make up the impure policy would make no difference. Accordingly, we assume that each transmission takes place over a randomly chosen channel.

#### 3.1 Analysis

Relations for pure SWP policies for single-slot messages were developed in Section 2.2. Here, they are used as a foundation for the derivation of \( P_c \) and \( S_g \) for Impure policies, and for the analysis of interpolation policies, which are a subset of impure policies.

Consider an impure policy that is specified as a set of pure policies \( \{\alpha_k\} \), chosen among with probabilities \( \{p_k\} \) upon message generation. Once we select the policy, the pure mechanism is used, so the error probability is

\[
P_c = \sum_k p_k P_{c_k},
\] (4)

where \( P_{c_k} \) is the error probability when using policy \( \alpha_k \). To derive the generation rate, note that the offered load \( G \) is unchanged, but \( E[N] \) is now the mean number of copies over all the pure policies. Therefore (1)

\[
S_g = \frac{G}{\sum_k p_k E[N_k]} = \frac{1}{\sum_k p_k \frac{1}{S_{\alpha_k}}},
\] (5)

where \( E[N_k] \) is the expected number of copies when using \( \alpha_k \).
**Interpolation policies.** These impure policies represent the most straightforward approximation of a pure policy that is capable of transmitting non-integer numbers of copies. Specifically, the number of copies that are transmitted in any given round is the result of a probabilistic choice between two consecutive integers.

An interpolation policy can thus be specified as follows. In round \(i\), having previously (actually) transmitted a total of \(j\) copies, we should transmit either \(n_{i,j}\) or \(n_{i,j} + 1\) copies with probability \(1 - p_{i,j}\) and \(p_{i,j}\), respectively. (For pure policies, \(p_{i,j} = 0\).)

Given \(i\) and \(j\), the expected number of copies transmitted is

\[
E[n_{i,j} | i, j] = n_{i,j} + p_{i,j}.
\]

The probabilities of all round-\(i\) copies colliding are

\[
\begin{align*}
P[\text{all collide} | i, j, n_{i,j}] &= (P_c)^{n_{i,j}} \\
P[\text{all collide} | i, j, n_{i,j} + 1] &= (P_c)^{n_{i,j} + 1}
\end{align*}
\]

Therefore,

\[
P[\text{all collide} | i, j] = (P_c)^{n_{i,j}} (1 - p_{i,j} + p_{i,j} P_c).
\]

Define \(Q_{i,j}\) as the probability that a message reaches round \(i\) after \(j\) copies were transmitted in previous rounds and all of them collided. All messages use round 1 (without any previous transmissions), so \(Q_{1,0} = 1\). For larger \(i\) and \(j\), \(Q_{i,j}\) can be derived recursively by

\[
Q_{i,j} = \sum_{k=0}^{j-1} Q_{i-1,k} \cdot (\delta(j - k) \cdot (1 - p_{i-1,k}) + \delta(j - k - 1) \cdot p_{i-1,k}) \cdot P_c^{j-k},
\]

where \(\delta(m) = 1\) if and only if \(n_{i-1,k} = m\) and is zero otherwise.

The mean number of transmitted copies is (6)

\[
E[N] = \sum_{i=1}^{D_r} \sum_{j} Q_{i,j} (n_{i,j} + p_{i,j}).
\]

To derive \(P_e\), we calculate the probability of reaching round \(D_r\) and failing this last round of transmissions.

\[
P_e = P[\text{failed round } D_r] = \sum_k Q_{D_r,k} P[\text{error}[D_r,k]] = \sum_k Q_{D_r,k} (P_c)^{n_{D_r,k}} (1 - p_{D_r,k} + p_{D_r,k} P_c).
\]

Using these formulae, the connections among \(G\), \(P_c\), \(P_e\), and \(S_g\) can be calculated. The capacity is \(S = S_g(1 - P_e)\).
3.2 Results

Optimal non-stationary pure SWP policies were derived in [5], where $S$ was plotted versus $P_c$. Different pure policies are optimal in different ranges of $P_c$ (for smaller $P_c$, policies transmitting more copies are used), but the capacity of any specific pure policy is continuous in $P_c$. Therefore, there are values of $P_c$ for which two pure policies, $\alpha_{k1}$ and $\alpha_{k2}$, have the same performance, i.e., $P_{e1} = P_{e2}$ and $S_1 = S_2$, and there are discontinuities in the derivative of the graph. The following examples show that Impure policies can sometimes improve performance near these values.

**Example 1** Let $(n_1, n_2, n_3)(\alpha_1) = (2, 3, 8)$ and $(n_1, n_2, n_3)(\alpha_2) = (2, 3, 9)$. When $P_c = 3.19935 \cdot 10^{-4}$, $S_{g1} = S_{g2} = 0.23923$ (both are optimal at this $P_c$). However, an Impure policy with $p_1 = 0.40258$ and $P_c = 0.55150$ achieves $S_{g}^I = 0.23929$. The improvement in $S_{g}$ (and capacity) are 0.022%.

**Example 2** Let $(n_1, n_2, n_3)(\alpha_1) = (1, 2, 3)$ and $(n_1, n_2, n_3)(\alpha_2) = (1, 2, 4)$. When $P_c = 1.18195 \cdot 10^{-2}$, $S_{g1} = S_{g2} = 0.28443$ (both are optimal at this $P_c$). However, an Impure policy with $p_1 = 0.38893$ and $P_c = 0.50668$ achieves $S_{g}^I = 0.28457$. The improvement in $S_{g}$ (and capacity) are 0.048%.

These examples strongly suggest that pure policies are near optimal among SWP policies. However, it is possible that further study of Impure policies (possibly interpolating among non-optimal pure policies) would provide better performance. This is beyond the scope of this work, but we conjecture that it is not the case.

**Remark 1** When pure SWP policies are considered, there are discontinuities in the derivative of the “envelope” graph (the curve depicting the capacity of the best pure policy at any given value of $P_c$), and the capacity is not convex in $P_c$. When Impure SWP policies are considered, the performance is slightly improved, but it is unclear whether the capacity is convex in $P_c$. However, Impure MWP policies can allocate different WPs to different optimal pure policies, hence it seems that their capacity is convex.

In the single-round case for multiple-working-point schemes, numerical results suggest that pure policies are optimal [10].

4 Conclusions

Capacity maximization of multi-channel Slotted ALOHA networks for single-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. Numerical results suggest that, under this performance measure, pure policies are near optimal among SWP policies, and in the single-round case pure policies are optimal. Nonetheless, we have established that the optimal policy is not necessarily pure.
References


