Retransmission Control in Mobile Radio Slotted ALOHA

Michele Zorzi and Ramesh R. Rao

ABSTRACT. For the slotted ALOHA protocol in a mobile radio environment, a retransmission control scheme is proposed, and a rigorous analysis is developed to prove the stability of the resulting system. This analysis is shown to apply to virtually any model for the spatial distribution of the users. It is also shown that the use of retransmission control is unavoidable if the network performance predicted in the recent literature is to be achieved.

1 Introduction

The use of the slotted ALOHA protocol [1] in mobile radio networks has been considered in recent years. This is due to the fact that in a mobile communication environment one can take great advantage of the so-called "capture effect", which takes place when one packet can be received even in the presence of simultaneous transmissions [2]. In fact, due to the random distribution of the distance of the terminals from the common receiver (near-far effect) and to the random effects in propagation (fading and shadowing), the received power from different users can vary greatly, and therefore there is a good chance that one packet will predominate over the others, "capturing" the receiver.

In fact, many papers have appeared in the recent literature, which analyze the capture probabilities and the average throughput and delay of such networks, under various assumptions about the traffic distribution and the propagation environment, with application to mobile services [3, 4, 5, 6, 7, 8]. While many average analyses have been presented, the issue of stability is considered in very few papers [5, 6]. However, stability is a major concern in random access protocols, since it is possible that the evolution of the system leads to an unbounded number of blocked users (as in the case of infinite user models, in which the issue is instability [9]) or converges towards an undesired stable point (as in the case of finite user models, in which the issue is bistability [10]).

Therefore, a stability analysis of such protocols must be considered, and adequate strategies to combat undesired behaviors must be devised. Even though some papers have considered stability, the retransmission control schemes, used in slotted ALOHA to stabilize systems otherwise unstable [9, 11], have almost never been applied to a mobile radio communications environments. In [12, 13], a rigorous investigation of the stability of a radio mobile network, in the presence of fading and shadowing, has been presented. The criterion given in [14] was used to assess stability, and some sufficient conditions for stability were given in terms of the spatial traffic distribution. Also, the use of the control retransmission policy proposed in [9, 11] was considered, and the corresponding performance was evaluated. It must be noted, however, that the algorithm considered in [9] was shown to be optimum for the geometric capture model, whereas in the present environment it turns out to be suboptimal, as discussed in [13].

In this paper, based on an analysis similar to [9], an optimal decentralized retransmission control algorithm is proposed and analyzed. In particular, the stability of the system is rigorously proven; also, it is shown that such a scheme will achieve values of the throughput arbitrarily close to the maximum possible. The analysis presented is not critically dependent on the functional form of the traffic distribution, and therefore applies to various environments.

In addition, this paper reiterates the need to examine the stability question closely, and demonstrates that efficient retransmission control protocols can be designed and proven to be stable rigorously.

2 Capture Model

The system model is the same as in [12, 13]: the analysis developed yields, for the probability of a successful transmission, conditioned on the number of colliding packets and on the distance from the base station,

\[ P_n(r) = P\text{[user at distance } r \text{ is successful }] = \int_{-\infty}^{\infty} e^{-\frac{s^2}{2 \sigma^2}} \frac{dI}{\sqrt{2 \pi \sigma}} \frac{[I(\xi, r)]^{n-1}}{[1 + \int_{0}^{\infty} \frac{g(s) ds}{1 + e^{s-\xi}(\xi)}]^{n}}, \]  

(1)

where

\[ I(\xi, r) = \int_{-\infty}^{\infty} e^{-\frac{s^2}{2 \sigma^2}} \frac{ds}{\sqrt{2 \pi \sigma}} \frac{\int_{0}^{\infty} \frac{g(s) ds}{1 + e^{s-\xi}(\xi)}}{[1 + e^{s-\xi}(\xi)]^{n}}, \]  

(2)

and \( g(r) \) is the pdf of the distance of the mobiles from the BS, identically zero outside the interval \([0, 1]\). The capture
probabilities can be written as
\[
C_n = \int_0^1 n P_n(r) g(r) dr = \int_{-\infty}^{\infty} e^{-\frac{\xi^2}{2\sigma^2}} d\xi \int_0^1 n[(\xi, r)]^{n-1} g(r) dr. \tag{3}
\]

Given \(g(r), \sigma\) and \(b\), these probabilities can be numerically evaluated for all \(n\). Note that, since \(\xi\) and \(r\) are random variables, \(I(\xi, r)\) is a random variable, and its \(k\)-th moment is the average probability of a successful transmission, \(P_{k+1}\), in the presence of \(k+1\) simultaneous transmissions.

It is shown in [14, 15] that the system without retransmission control is stable if and only if \(\lim_{n \to \infty} C_n > 0\). It is possible to show that, under certain conditions, the sequence of the capture probabilities converges to a limit greater than zero, which depends on the capture ratio, \(b\), and on the propagation exponent, \(\eta\). Interestingly, it does not depend on the shadowing parameter, \(\sigma\). A sufficient condition is that \(g(r)\) be "linear near zero", i.e. that \(\lim_{r \to 0} g(r)/r = C_0\), with \(C_0\) a positive real number [12, 13]. On the other hand, if we assume that users are not allowed to be arbitrarily close to the BS, i.e., \(g(r) \equiv 0, r \in [0, \tau]\), for some \(\tau > 0\), the limit of \(C_n\) is zero, as observed in [16], and the system is unstable for any mean input rate, \(\lambda\), and for any distribution of the arrivals, unless a control policy is adopted. In the presence of a large backlog, the number of retransmissions is such that almost every slot contains unsuccessful transmissions, leading to a deadlock. This can be avoided by adaptively varying the retransmission probability in such a way that the global traffic (new arrivals plus retransmissions) is maintained at an optimum level [9, 11]. In the next sections, we analyze such a control policy.

3 Controlled ALOHA

Any backlogged user transmits in slot \(t (t \in Z_+)\) with probability \(f_t\), which depends on the past evolution of the channel output. The new arrivals are Poisson with rate \(\lambda\), and are transmitted in the slot immediately following their arrival. Let \(Z_t\) be the feedback information on slot \(t\), available at time \(t+1\), and taking values in the set \(\{0, 1, \epsilon\}\). The retransmission control algorithm, performed in a decentralized manner, requires the backlogged users to adjust the retransmission probability according to the recursion [9, 11]:
\[
f_{t+1} = a^\gamma(Z_t) f_t \wedge j, \tag{4}
\]
where \(\gamma, \beta, a(j), j = 0, 1, \epsilon\) are positive constants which affect the stability and convergence of the system, and must be appropriately chosen; also, 0 < \(\beta < 1\) and \(\gamma > 0\).

As in [9], we define, for \(j = 0, 1, \epsilon\),
\[
P_j = P[Z_t = j | N_t = n, f_t = f]. \tag{5}
\]
Then, we can show (see Appendix A) that, with \(C_n\) given by (3),
\[
P_0 = e^{-\lambda}(1 - f)^n, \tag{6}
\]
\[
P_1 = \sum_{i=0}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \sum_{j=0}^{\infty} \binom{n}{j} f^j (1 - f)^{n-j} C_{i+j}
\]
\[
= \int_{-\infty}^{\infty} e^{-\frac{\xi^2}{2\sigma^2}} d\xi \int_0^1 g(r) dr \left[ \lambda e^{-\lambda(\xi, r)} (1 - 1\theta(\xi, r)) f^n + e^{-\lambda(\xi, r)} n f (1 - 1\theta(\xi, r))^{n-1} \right], \tag{7}
\]
and
\[
P_0 = 1 - P_0 - P_1, \tag{8}
\]
where \(\theta(\xi, r) = 1 - I(\xi, r)\). Using the local Poisson approximation as in [9, 11], we obtain, for the approximate version of the above probabilities,
\[
P_0 = e^{-\lambda}, \tag{9}
\]
\[
P_1 = \int_{-\infty}^{\infty} e^{-\frac{\xi^2}{2\sigma^2}} d\xi \int_0^1 g(r) dr Ge^{-\lambda(\xi, r)}, \tag{10}
\]
and
\[
P_0 = 1 - \hat{P}_0 - \hat{P}_1, \tag{11}
\]
where \(G = \lambda + e^k\) and \(\psi = \ln([n \in 1])f\). As shown in [17], \(\hat{P}_2 - P_1 \to 0, j = 0, 1, \epsilon\), uniformly as \(n \to \infty\) or \(f \to 0\).

Following the approach in [9], it can be shown that
\[
\lim_{n \to \infty \text{ or } f \to 0} E[p_{\psi_1 + 1} - p_{\psi_1} | N_t = n, f_t = f] = \gamma \hat{\eta}(\psi), \tag{12}
\]
where
\[
\hat{\eta}(\psi) = c_0 \hat{P}_0 + c_1 \hat{P}_1 + c_2 \hat{P}_2, \tag{13}
\]
\[
\hat{P}_j = \ln([j]), j = 0, 1, \epsilon. \tag{14}
\]
It is easy to see that, if \(\psi^*\) is a stable point, \(\hat{P}^*_0\) are the corresponding probabilities, and \(G^* = \lambda + e^{\psi^*}\), the choice
\[
c_0 = \frac{1 - \hat{P}_0^*}{1 + \hat{P}_0^*} = \frac{e^{\psi^*} - 1}{e^{\psi^*} + 1}, \tag{15}
\]
\[
c_1 = c_2 = -\frac{\hat{P}_2^*}{1 + \hat{P}_0^*} = -\frac{1}{e^{\psi^*} + 1}, \tag{16}
\]
guarantees that
\[
sgn \hat{\eta}(\psi) = -sgn(\psi - \psi^*), \forall \psi. \tag{17}
\]
In fact, in this case, we have
\[
1 + \frac{\hat{P}_0^*}{\hat{P}_0^*} \hat{\eta}(\psi) = 1 - \frac{\hat{P}_0^*}{\hat{P}_0^*} \hat{P}_0 - \hat{P}_1 - (1 - \hat{P}_0 - \hat{P}_1) = \frac{\hat{P}_0^*}{\hat{P}_0^*} - 1 \tag{18}
\]
which is obviously zero when \(\psi = \psi^*\). Also, since \(\hat{P}_0^*\) is a decreasing function of \(\psi\), we have the above result.

The property (17) can be interpreted in the following manner: the control scheme steers the local traffic towards a given point, \(G^*\). We note that this property of the drift function, \(\hat{\eta}(\psi)\), with the above choice of the \(c_j\)'s, depends only on \(\hat{P}_0\), which in turn is independent of the \(C_n\)'s. In any situation in which \(\hat{P}_0\) is a decreasing function of \(\psi\), (17) will be satisfied, with the above choice of the \(c_j\)'s, regardless of the expression of the capture probabilities. In particular, the system will "converge" to \(\psi^*\) even if \(\psi^*\) is suboptimal. Also, as already observed in [11], the fact that \(c_j = c_0\) means that, as far as the algorithm is concerned, the users need not to distinguish a successful slot from a collision, but are only required to detect the presence of some signal.
3.1 Throughput

Since in our system only one packet per slot can be received successfully, the average throughput, $S$, is given by the probability that $Z_t = 1$, i.e., $P_1$ in (7), and can be approximated, using the Poisson local approximation, by $P_1$. We have, therefore,

$$S(\psi) \simeq \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \frac{d}{\sqrt{2\pi \sigma}} \int_0^1 g(r) dr = \hat{S}(\psi).$$

(19)

Since $\hat{S}$ is zero for $G = 0$ and in the limit as $G \to \infty$, and is always continuous and nonnegative, it has a positive maximum, $S^*$, for some value of $\psi$, $\psi^*$. In the next section, we will prove that the Markov process associated with the controlled system is stable for all $\lambda < S^*$.

4 Stability of the Controlled Scheme

Following the approach in [9], we invoke the Mikhailov’s theorem [18] to prove that the bidimensional Markov process $X(t) = (N_t, f_t)$, with state space $S = \mathbb{Z}_+ \times (0, 1)$, is stable. In order to prove that, we need to prove the following:

A.1) $E[\|X(t+1) - X(t)\|^2 | X(t) = X] < D$, for all $X \in S$, for some constant $D < \infty$.

A.2) The limit $M(X)$ of the mean drift vector function of $X(t)$ exists and the convergence is uniform.

A.3) There exists a stochastic Lyapunov function, i.e., a scalar function $V(X)$ defined over $S$, with the following properties:

A.4) It is always non-negative.

A.5) There exist two finite constants $D_1$ and $D_2$ such that $V(X) \leq D_1 \|X\| + D_2$, for all $X$.

A.6) For any $\epsilon > 0$, there exists $r_0$ such that

$$|LV(X) - M(X)\text{grad}V(X)| < \epsilon$$

(20)

for all $X$ such that $\|X\| > r_0$, with

$$LV(X) = E[V(X(t+1)) - V(X(t))]|X(t) = X].$$

(21)

A.7) There exists an $\epsilon > 0$ such that $M(X)\text{grad}V(X) < -\epsilon$ for all $X \in S$.

Note that A.1 and A.2 are verified as in [9] (the proof depends on the probabilities $P_j$, $j = 0, 1, \epsilon$ only through their uniform convergence to the Poisson local approximation). Note also that, choosing $P_j = z_j + \alpha \ln \cosh(z_j - \psi^*)$, with $\alpha > 0$, A.3 to A.5 are automatically verified. Moreover, the proof of A.6 given in [9] is based on the properties of $V(X)$ and on the fact that $X(t)$ satisfies A.1 and A.2, and therefore applies in the present context. Finally, Proposition 5.1 in [9], showing that, for $\lambda < S^*$ and $\gamma > 0$, there exists an $\epsilon > 0$ such that

$$M(X)\text{grad}V(X) < -\epsilon$$

(22)

still holds, since it is based on the continuity of $\hat{S}(\psi)$ and on the fact that $\hat{m}(\psi)(\psi - \psi^*) \leq 0, \forall \psi$, and therefore A.7 is also verified. Therefore, according to Mikhailov’s criterion, the following result can be assessed.

Theorem: If $\lambda < S^*$ and $\gamma > 0$, then the Markov process $X(t) = (N_t, f_t)$ is stable.

We remark that the above result has been obtained under very general conditions. In particular, $g(r)$ can be arbitrary (provided that the probability of a user to be at distance $r = 0$ is zero, which is reasonable). In [16], it was observed that the ALOHA system with capture, when a uniform distribution of the traffic (i.e., $g(r) = 2r$, $r \in [0, 1]$) is assumed, can be studied analytically, but suffers from two major limitations, namely the unrealistic assumption that users can be arbitrarily close to the BS and the fact that, in reality, the distribution of the traffic is not uniform since backlogged users are more likely to be far from the BS. In the above analysis, these two difficulties have been overcome, because $g(r)$ can be identically zero for $r < r_0$, for some $r_0 > 0$, and can have any functional form of practical interest. Therefore, the proposed control scheme is very powerful, and applies to any mobile radio environment in the presence of fading.

5 Discussion and Conclusions

In the above, we rigorously analyzed the stability properties of the Slotted ALOHA protocol with capture and retransmission control. In the recent literature, the expression (19) of the throughput has been computed, for some specific distributions, and its maximum has been claimed to be the “capacity” or the “maximum achievable throughput” of Slotted ALOHA. It must be noted, however, that in reality, with the protocol implementations considered in those papers, this is not true, and the system will be unstable or, in the best situation, will achieve a throughput arbitrarily close to the limit of the capture probabilities, $C_0$, which can be substantially smaller than $S^*$. Therefore, the claims about the achievable throughput and the network performance turn out to be conceptually incorrect. This is true not only from a theoretical point of view, but also from an engineering standpoint, since such systems can not work as claimed, and sometimes they can not even work at all. Rigorous analysis shows that the predicted performance can only be achieved by using a dynamic retransmission scheme. Also, we believe that an analysis similar to the above is feasible for a broad variety of systems, and is the only way to correctly approach the problem.

This paper confirms that the performance promised in the literature can really be achieved (but this was by no means obvious, a priori), and points out how this can be done only by implementing a retransmission control scheme, which guarantees stability and enhances the traffic performance. Also, we strongly underline the difference between the global character of the Poisson approximation for the attempted traffic, which relies on a flawed assumption, and its
local (i.e., conditional) character in each slot, which is the correct approach. One of our main goals in writing this paper was to raise awareness about these fundamental issues, which, in the recent literature on mobile radio networks, have been underestimated, if not completely neglected.

Future research is needed to investigate the delay performance, not considered in this paper. In particular, even though the stability is always guaranteed by the above analysis, and the delay experienced by a packet is therefore bounded, a quantitative analysis of delay, which would require study of the higher moments of the drift, has not been done yet. The choice of some parameters (e.g., $\beta$ and $\gamma$) may be suggested by an analysis of this sort.

Appendix

A Computation of the $P_j$'s

Since the arrival process and the backlog are assumed to be independent, the probability $P_0$ of having an empty slot is just the product of the probability of no new arrivals, equal to $e^{-\lambda}$, and the probability of no retransmissions, equal to $(1 - f)^n$.

To compute $P_1$, we observe that:

$$\sum_{i=0}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \sum_{j=0}^{\infty} \binom{n}{j} f^j (1 - f)^{n-j} y^{j+i} = e^{-\lambda(1-y)} (1 - (1 - y)f)^n \overset{\Delta}{=} \Theta(y)$$

and

$$\sum_{i=0}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \sum_{j=0}^{\infty} \binom{n}{j} f^j (1 - f)^{n-j} (i+j)y^{j+i-1} = \frac{d\Theta}{dy}.$$

Replacing $y$ by $I(\xi, r)$ and averaging over $\xi$ and $r$, we obtain (7).

Since the events 0, 1, $\epsilon$ are disjoint and exhaust all the possibilities, $P_0 + P_1 + P_\epsilon = 1$. \(\square\)

References