NETWORK ASSISTED DIVERSITY FOR RANDOM ACCESS WIRELESS DATA NETWORKS

Michail K. Tsatsanis, Ruifeng Zhang and Subrata Banerjee
Electrical & Computer Engineering Dept.
Stevens Institute of Technology
Castle Point on Hudson
Hoboken, NJ 07030
e-mail: mtsatan@stevens-tech.edu

ABSTRACT

Contention based protocols inevitably lead to packet collisions and therefore underutilization of the channel resources. Collided packets however contain mixtures of useful user packets and should not necessarily be discarded. In this paper, a signal processing viewpoint is adopted in the development of alternative random access protocols which (at high SNR) introduce no throughput penalties. The approach is based on utilizing the network to create diversity through selective retransmissions of the corrupted packets. The user packets are then retrieved using conventional source separation techniques. The resulting protocol requires \( k \) time slots to transmit \( k \) collided users and does not affect the channel throughput in a negative way. Performance issues are studied in the paper, related to the implementation of the collision detection algorithm. The protocol's parameters are optimized to maximize the system throughput.

1. INTRODUCTION

The majority of wireless cellular services today are voice oriented and provide a circuit switched, constant bit rate service to each user. In these circumstances, multiplexing various users together may be accomplished with relatively simple TDMA, FDMA or CDMA techniques, (e.g., [7]).

There is, however, an increasing interest in wireless data services and/or multimedia services, where variable bit rate sources have to be multiplexed. In this case, simple TDMA solutions are extremely inefficient and some random access techniques are typically preferred [10]. Simple random access protocols of the ALOHA type offer a relatively straightforward implementation and can accommodate bursty users. They suffer however from a severe throughput penalty and underutilization of the channel resources [3].

Carrier sensing (CSMA) and/or collision detection mechanisms are typically employed in an effort to improve the throughput performance of random access schemes [3]. But in a wireless environment carrier sensing may not be reliably performed due to unpredictability of the wireless channel (e.g., [7]). Data sensing, (DSMA) [5] is usually implemented in cellular data networks. In this scheme, the base station detects collisions and continuously broadcasts a busy/idle signal in a control channel to all users, this way providing the necessary feedback.

The emphasis in the random access literature has been mostly on retransmission schemes which minimize future collisions (e.g., [10]). However, when a collision does occur, the collided packets are typically discarded and no information is exploited from them. It is clear that the throughput penalty incurred by collisions cannot be eliminated unless some way is devised to extract useful information from the collided packets.

2. NETWORK ASSISTED DIVERSITY FOR COLLISION RESOLUTION

Consider a wireless cellular random access system with \( J \) users. The system is slotted and at each time slot \( n \), each user \( i \) may transmit a packet consisting of \( N \) symbols, \( w_{i,n}(t) = [w_{i,n}(1), \ldots, w_{i,n}(N)]^T \), provided that its queue is non-empty (see Fig. 1). The symbol \( w_{i,n}(t) \) is assumed drawn from a finite QAM constellation. The received baseband discrete-time record for packet \( n, y_{n}(N) = [y_{n}(1), \ldots, y_{n}(N)]^T \), (after matched filtering and sampling

![Figure 1. A random access, slotted wireless system](image-url)
at the symbol rate) is

\[ y_N(n) = \sum_{i \in I(n)} a_i(n)w_{N,i}(n) + v_N(n), \]  

(1)

where \( I(n) \) is the index set of users that are active at time slot \( n \), \( v(n) \) is additive noise, and \( a_i(n) \) is the \( i \)-th user's channel gain. A non-frequency selective channel is considered in this paper and therefore the gains are \( a_i(n) = A_i(n)e^{j\varphi_i(n)} \), where \( A_i(n) \) and \( \varphi_i(n) \) denote the amplitude and phase, respectively. The following assumptions will be made regarding the model of (1):

(A1): \( v_n(k) \) is zero mean, complex circular AGN, white in \( n \) and \( k \), with variance \( \sigma_v^2 \);

(A2): \( \varphi_i(n) \) is a uniform \([0, 2\pi]\) phase, i.i.d., in \( i \) and \( n \).

The amplitude \( A_i(n) \) may be assumed to be either constant or randomly distributed depending on whether the channel is fading and whether power control is implemented or not. Furthermore, each user is assumed to have an infinite length buffer which holds fixed length packets arriving as a Poisson process of rate \( \lambda \).

There are also some restrictive assumptions in (1). One is the perfect synchronization of all users implicit in (1). That restricts its applicability to cellular systems with synchronization control (users may be synchronized only for one particular receiver location, and therefore the notion of a Base Station (BS) becomes necessary). Furthermore, no multipath effects are taken into account in (1). It is customary to discard a packet \( y_N(n) \) when a collision is detected, and initiate a retransmission schedule. As can be seen from (1), however, \( y_N(n) \) contains information about the transmitted packets and should be exploited.

Let us consider the case where \( k \) users collide in time slot \( n \), and therefore \( I(n) = \{i_1, \ldots, i_k\} \). Then \( y_N(n) \) consists of a mixture of \( k \) sources that need to be separated. From a signal processing viewpoint, this problem may be solved if we are able to create an \( M \)-branch diversity (e.g., use \( M \) antennas) with \( M \geq k \), and collect \( M \) independent mixtures of the signals \( w_{N,i} \) (c.f. Space Division Multiple Access techniques [6]). Here we will show that in the current random access framework, the diversity can be created at the protocol level by using the network resources.

Assume for the moment that all users are aware of (a) whether there has been a collision at time slot \( n \), and (b) its multiplicity \( k \).

Assume furthermore that according to the protocol each collided user will retransmit its information packet \( k - 1 \) more times in the next \( k - 1 \) slots (i.e., in slots \( n + 1, \ldots, n + k - 1 \)). Finally, no other user will initiate a new transmission in the next \( k - 1 \) slots. An example of this procedure for a collision of two users is shown in Fig. 2. With these conventions, the BS will receive a total of \( k \) copies of the collided packets,

\[ \begin{bmatrix} y_N(n), & y_N(n + k - 1) \\ w_{N,i_1}(n), & \cdots, & w_{N,i_k}(n) \end{bmatrix} = \begin{bmatrix} a_{i_1}(n) & \cdots & a_{i_1}(n + k - 1) \\ \vdots & \ddots & \vdots \\ a_{i_k}(n) & \cdots & a_{i_k}(n + k - 1) \end{bmatrix} \]

or equivalently,

\[ Y_{N,k}(n) = W_{N,k}(n)A(n) + V_{N,k}(n), \]

(3)

with obvious definitions for the matrices in (3). If the mixing matrix \( A(n) \) is known or can be estimated, then a simple suboptimal solution is

\[ W_{N,k}(n) = Y_{N,k}(n)A^{-1}(n), \]

(4)

provided that \( A(n) \) has full rank. The latter property of \( A(n) \) is guaranteed with probability 1, under assumption \( (A2) \). More involved maximum likelihood solutions for (3) are also applicable.

Notice that only \( k \) slots are required to make \( A(n) \) square and therefore resolve \( k \) colliding users. Hence, no slots are lost and no throughput penalties are incurred by this collision resolution method.

To satisfy the requirement for all users to be aware of the collision, the BS is in charge of identifying that information and broadcasting it back in a separate control channel. (See, for example the DSMA protocol, [7].) As far as the feedback procedure is concerned, it can be implemented using a broadcast control channel with rate one bit/time slot.

At the end of slot \( n \) the BS may indicate in the control channel whether the next slot is free or busy (i.e., reserved for re-transmission). If slot \( n + 1 \) is indicated as busy all users who transmitted in slot \( n \) will retransmit. If the BS repeats the busy indication for \( k - 1 \) slots all colliding users will be forced to re-transmit \( k - 1 \) times. Of course new users will not be allowed to transmit while the busy signal is on.

We next turn our attention on the problem of how the BS may identify the colliding users.

3. COLLISION DETECTION TECHNIQUE

The identification of the set of active users \( I(n) \) is equivalent making a decision for each user, as to whether he is active or not. In order for the BS to discriminate the users, an address field is required in the packet which contains a unique ID sequence for each user. Let us assume without loss of generality that the first \( M \) symbols of each packet of user \( i \) form an identifying vector \( w_i \), (using MATLAB notation,) \( w_i = [w_{N,i}(n); \cdots; w_{N,i}(n+M-1)]. \) Let also \( y(n) = [y_N(n); \cdots; y_N(n+M-1)]. \) Based on equation (1),

\[ y(n) = \sum_{i \in I(n)} a_i(n)w_i + v(n). \]

Equation (5) reveals the fact that the estimation of \( a_i(n) \) from the data \( y(n) \) is a linear regression problem, while the identification of \( I(n) \) is a model (or regressor) selection problem. There are \( 2^M \) different hypotheses that need to be evaluated.

\[ H_m : I(n) = I_m, \quad m = 1, 2, \ldots, 2^M. \]

A simple comparison of the modeling error for each hypothesis is not sufficient since it is well known that the error always decreases as more regressors are added (e.g., [4]). However, most classical model selection procedures are applicable to this problem[2, 1, 8]. Unfortunately, due to the large number of hypotheses \( (2^M) \), all those procedures will be computational demanding and their complexity will increase exponentially with the number of users.

In order to simplify the collision detection procedure, in the sequel we make the following assumption,
(AS 3): \( w_i^H w_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \).

In other words, we design the ID sequences to be orthogonal to each other.\(^1\) Under (AS 3) it can be shown that the joint user detection problem can be decoupled into \( J \) independent single user detection problems.

Indeed, the conditional likelihood function of the data is, (c.f. (AS 1))

\[
f(y(n)|\alpha_i(n), i \in I_m, \ I_m) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp \left( -\frac{||y(n) - \sum_{i \in I_m} \alpha_i(n) w_i||^2}{2\sigma^2} \right)
\]

(7)

If we project \( y(n) \) onto the ID vectors \( w_i \), then

\[
y(n) = \sum_{i=1}^{J} [w_i^H y(n)] w_i + r(n) \quad \text{(8)}
\]

where \( r(n) \) is the difference between \( y(n) \) and its projection onto the space spanned by the ID sequences \( w_i \). Thus \( r(n) \) is orthogonal to all \( w_i \)'s. Substituting (8) into (7) and using (AS 3), (7) can further be written as

\[
f(y(n)|\alpha_i(n), i \in I_m, \ I_m) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp \left( -\frac{||r(n)||^2}{2\sigma^2} \right) \prod_{i \notin I_m} \exp \left( -\frac{||w_i^H y(n)||^2}{2\sigma^2} \right) \prod_{i \in I_m} \exp \left( -\frac{||w_i^H y(n) - \alpha_i(n)||^2}{2\sigma^2} \right). \quad \text{(9)}
\]

It is clear from (9) that the sufficient statistic for detecting user \( i \) is \( z_i(n) = w_i^H y(n) \), i.e., the output of a matched filter matched to the signature \( w_i \). And the \( 2^J \)-ary hypotheses detection problem is reduced to \( J \) independent binary hypothesis ones.

For user \( i \), the conditional likelihood under two hypotheses: \( H_i, 0 \) corresponding to \( i \) in \( I_m \), and \( H_i, 1 \) corresponding to \( i \) not in \( I_m \), are

\[
f(y(n)|H_i, 0) = C \exp \left( -\frac{|z_i(n)|^2}{2\sigma^2} \right), \quad \text{(10)}
\]

\[
f(y(n)|\alpha_i(n), H_i, 1) = C \exp \left( -\frac{|z_i(n) - \alpha_i(n)|^2}{2\sigma^2} \right) \quad \text{(11)}
\]

where \( C \) is a complex constant not depending on \( \alpha_i(n) \). Thus the likelihood ratio is

\[
\Lambda(y(n)|\alpha_i(n)) = \frac{f(y(n)|\alpha_i(n), H_i, 1)}{f(y(n)|H_i, 0)} = \exp \left( \frac{z_i(n)\alpha_i^*(n) + \alpha_i(n)}{2\sigma^2} \right) \exp \left( \frac{|\alpha_i(n)|^2}{2\sigma^2} \right) \quad \text{(12)}
\]

The optimal detector for the \( i \)-th user depend on the statistical property of the \( \alpha_i(n) \). According to different channel models, \( \alpha_i(n) \) falls into four cases :\(^2:\)

---

\(^1\)The unit norm assumption implied in (AS 3) can be made without loss of generality.

\(^2\)Of the following four cases, the first one violates (AS 1). It is included, however, just for completeness.

---

ideal AWGN channel: \( \alpha_i(n) \) is deterministic constant but unknown;

non-fading channel with power control but arbitrary phase:

the amplitude of \( \alpha_i(n) \), \( A_i(n) \), is constant (maybe unknown) while the phase \( \varphi_i(n) \) is random and uniformly distributed in \([0, 2\pi]\);

Rayleigh fading channel: the phase \( \varphi_i(n) \) is uniformly distributed in \([0, 2\pi]\), while the amplitude \( A_i(n) \) is Rayleigh distributed with parameter \( \sigma_A \) and \( \lambda_i(n) \) are independent;

Rician fading channel: the phase \( \varphi_i(n) \) is uniformly distributed in \([0, 2\pi]\), while the amplitude \( A_i(n) \) is Rician distributed with parameter \( A \) and \( \sigma_A \), and \( \lambda_i(n) \) are independent;

For the first case, the optimal optimal detector is the generalized maximum likelihood test, which is estimating \( \alpha_i(n) \) by maximizing the likelihood function, then plugging the estimate into (12), and comparing the result with a threshold. For the other three cases, the optimal detector is the mean maximum likelihood test, which is averaging the conditional likelihood ratio \( \Lambda(y(n)|\alpha(n)) \) with respect to \( \lambda_i(n) \) and \( \varphi_i(n) \), and comparing the result with a threshold.

Despite the different distributions of \( \alpha_i(n) \), it turns out that the optimal detectors under all the four channel environments are the same [10]:

\[
|z_i(n)| \xrightarrow{H_i,0} T, \quad \text{(13)}
\]

i.e., to compare the amplitude of the output of the receiver matched filter matched to the ID sequence \( w_i \) with a predetermined threshold \( T \). And the probability of false alarm \( P_F \) under the four channel environments are also the same:

\[
P_F = \int_T^\infty \frac{x}{2\sigma^2} \exp \left( -\frac{x^2}{2\sigma^2} \right) dx = \exp \left( -\frac{T^2}{2\sigma^2} \right). \quad \text{(14)}
\]

The probability of detection \( P_D \) under every channel environment is listed below:

ideal AWGN channel:

\[
P_D = \int_T^\infty \frac{x}{2\sigma^2} \exp \left( -\frac{x^2 + A_i^2(n)}{2\sigma^2} \right) I_0 \left( \frac{A_i(n)x}{\sigma^2} \right) dx;
\]

non-fading channel with power control but arbitrary phase:

\[
P_D = \int_T^\infty \frac{x}{2\sigma^2 + \sigma_A^2} \exp \left( -\frac{x^2 + A_i^2(n)}{2(\sigma^2 + \sigma_A^2)} \right) dx = \exp \left( -\frac{T^2}{2(\sigma^2 + \sigma_A^2)} \right); \quad \text{(17)}
\]

Rayleigh fading channel:

\[
P_D = \int_T^\infty \frac{x}{\sigma^2 + \sigma_A^2} \exp \left( -\frac{x^2 + A_i^2(n)}{2(\sigma^2 + \sigma_A^2)} \right) I_0 \left( \frac{A_i(n)x}{\sigma^2 + \sigma_A^2} \right) dx.
\]

Rician fading channel:

\[
P_D = \int_T^\infty \frac{x}{\sigma^2 + \sigma_A^2} \exp \left( -\frac{x^2 + A_i^2(n)}{2(\sigma^2 + \sigma_A^2)} \right) I_0 \left( \frac{A_i(n)x}{\sigma^2 + \sigma_A^2} \right) dx.
\]
In the above formulas, \( L_0(\cdot) \) is zeroth order modified Bessel function of the first kind. In the next section, we will show the role of \( P_D \) and \( P_P \) in the expression of the system throughput, and optimize them to maximize the system throughput.

4. THROUGHPUT ANALYSIS

In section 2 we explained that the proposed method does not require extra slots to resolve a packet collision and therefore introduces no throughput penalties. This conclusion however is correct only under the assumption that the BS makes no errors in identifying the number of active users in every slot. For example, if two users collide but the BS incorrectly concludes that only one user is present, no retransmission will be requested and the collision will not be solved. It is therefore evident that in order to more accurately assess the performance of the proposed method we need to study the probability of incorrect BS decisions and their effects on the system’s throughput.

We will consider the worst case scenario, in order to provide a unified analysis framework. We assume that:

(i) The number of active users is determined at the first transmission by the BS, and the BS is not allowed to correct its original decision using subsequent transmissions;

(ii) Every incorrect decision by the BS results in the loss of all packets involved in that transmission epoch.

Finally, we will not be concerned here with packets which are resolved but lost due to excessive bursts of errors in the payload portion of the packets. These losses are not due to the collision access protocol but due to inadequate error correcting coding.

It will be instructive for our analysis purposes to view the traffic in the channel as a flow of epochs with random length (measured in time slots). The epochs include idle epochs (which consist of only one idle time slot) or busy epochs (during which some packets are under transmission). The length of a busy epoch is the number of time slots the channel takes to serve the currently active users. From a throughput viewpoint it is important to further distinguish busy epochs into useful epochs (meaning absence of detection errors) and corrupted epochs (meaning presence of detection errors).

The system throughput is defined as

\[
R = \frac{\text{average length of useful epoch}}{\text{average length of (busy or idle) epoch}}. \tag{19}
\]

And if we denote by \( P_e \) the probability of a user’s queue being empty at the beginning of an epoch (see Fig. 3), we have the following results.

Proposition 1: The system’s throughput can be expressed as

\[
R = \frac{J(1 - P_e)}{J(1 - P_e) + P_e} P_D [(1 - P_e) P_D + P_e (1 - P_P)]^{J - 1}. \tag{20}
\]

The proof is omitted in this paper. Readers may refer to [12] for detailed proof.

As \( SNR \to \infty \), we expect \( P_D \to 1 \) and \( P_P \to 0 \). Hence we obtain the following corollary.

Corollary 1: As \( SNR \to \infty \), the throughput

\[
R \to \frac{J(1 - P_e)}{J(1 - P_e) + P_e}. \tag{21}
\]

In order to express (20) in terms of \( \lambda \), we need to relate the probability of empty queue \( P_e \) with the input traffic rate \( \lambda \). In the flow of epochs, if we focus on the beginning instants of each epoch, and specify as our state variable the number of packets in the queue of a user at that time instant, then the state forms an embedded Markov chain. If we let \( q_n \) be the state variable that is, the number of data packets in a user’s queue at the beginning of the \( n \)-th epoch, then we seek expressions for \( \lim_{n \to \infty} P_r(q_n = k) \) or its characteristic function. As a special case \( P_e = \lim_{n \to \infty} P_r(q_n = 0) \).

Proposition 2: \( P_e \) is the unique solution in \([0, 1]\) of the equation

\[
\lambda P_e + (1 + \lambda), P_e - (1 - \lambda) = 0. \tag{22}
\]

For a proof see [12].

5. THRESHOLD SELECTION

According to Proposition 1, the throughput depends on \( P_D \) and \( P_P \) which in turn depend on the detector’s threshold \( T \). It is therefore worthwhile to appropriately select \( T \) so that the throughput is maximized. The solution is simply to set the derivative of (20) equal to zero,

\[
\frac{dR}{dP_P} = 0. \tag{23}
\]

Of course \( P_D \) is a function of \( P_e \) as given by the ROC curves of the detector. After some tedious but straightforward differentiation we can obtain a solution in terms of \( \frac{dP_D}{dP_P} \).

\[
\frac{dP_D}{dP_P} = \frac{J - 1}{J - 1 - \frac{P_e}{P_D}}. \tag{24}
\]

Depending on the particular fading conditions assumed and therefore the particular ROC expressions given in the previous section, (24) can be solved for the optimal \( P_P \), and further for the optimal threshold \( T \).

At high SNR, we can simplify (24) by approximating \( \frac{dP_D}{dP_P} \) with 1. Then (24) becomes

\[
\frac{dP_D}{dP_P} = \eta, \quad \eta = \frac{J - 1}{1 + J - 1 \frac{P_e}{P_D}}. \tag{25}
\]

Notice that the right hand side of (25) depends only on the traffic characteristics and not on the \( SNR \) or receiver parameters.

We close this section with an example of how (25) can be solved for the case of Rayleigh fading. In that case, \( P_D \) and \( P_P \) are given by (17) and (14). We can also see that

\[
P_D = \frac{P_P}{P_P + \frac{\sigma^2}{\sigma_e^2}}, \quad SNR = \frac{\sigma^2}{\sigma_e^2}. \tag{26}
\]

and therefore

\[
\frac{dP_D}{dP_P} = \frac{1}{1 + SNR \cdot P_P + \frac{\sigma^2}{\sigma_e^2}}. \tag{27}
\]
Substituting (26) in (25) and solving for $P_F$ we obtain

$$P_{F,\text{opt}} = \left[\eta (1 + SNR)^{\	ext{SNR}}\right]^{1+\text{SNR}}. \quad (28)$$

By further approximating at high $\text{SNR}$: $1 + \text{SNR} \approx \text{SNR}$, (28) becomes

$$P_{F,\text{opt}} = \frac{1}{\eta \text{SNR}}. \quad (29)$$

Finally, by substituting (29) into (14) we can solve for the optimal threshold

$$T_{\text{opt}} = \sqrt{2\sigma^2 \log (\eta \text{SNR})}. \quad (30)$$

Notice that the optimal threshold monotonically increases with $\text{SNR}$ and with the traffic load parameter $\eta$.

In fact, (29) has no meaning when $\eta \text{SNR} < 1$. If so, $P_{F,\text{opt}} = 1$, correspondingly $T = 0$ is the optimal solution.

6. SIMULATIONS

We tested the proposed random access method on a simulated slotted data communication system. The total number of users is $J = 16$, and users’ ID sequences are selected from a J-th order Hadamard matrix. All users maintain packets queues which are fed by Poisson sources with density $\lambda$, generating data packets to contribute to the network traffic. The system is run 10000 time slots under each set of parameters. Fig. 4 shows the variation of the system throughput against total traffic load $\lambda J$ for $\text{SNR} = \infty$, $16 dB$, $13 dB$, and $10 dB$. Fig. 5 shows the variation of the system delay against the same parameters. Under all circumstances, the probability of false alarm $P_F$ is controlled under the level of 0.001. We can see that the throughput performance in Fig 4 is in good agreement with the analytical expressions for a range of moderate to high $\text{SNR}$ values met in wireless communication systems. The delay performance of the proposed method is satisfactory even under very heavy load, and is only slightly affected by the $\text{SNR}$.

REFERENCES


Figure 4. Throughput vs Traffic load

In above two figures the curves from up correspond $\text{SNR} = \infty$, $16 dB$, $13 dB$ and $10 dB$, respectively.)

Figure 5. Delay vs Traffic load