A Random Access Protocol incorporating Multi-Packet Reception, Retransmission Diversity and Successive Interference Cancellation

Ramiro Robles*

Desmond C. McLernon

Mounir Ghogho

*CISTER Research Center

CISTER-TR-150711

2015/09/03

A Random Access Protocol incorporating Multi-Packet Reception, Retransmission Diversity and Successive Interference Cancellation

Ramiro Robles*, Desmond C. McLernon, Mounir Ghogho

*CISTER Research Center

Polytechnic Institute of Porto (ISEP-IPP)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: rasro@isep.ipp.pt

http://www.cister.isep.ipp.pt

Abstract

This paper presents a random access protocol assisted by a set of signal processing tools that significantly improve the multi-packet reception (MPR) capabilities of the system. A receiver with M antennas is used to resolve collisions with multiplicity K _ M. The remaining unresolved conflicts (with multiplicity K > M) are processed by means of protocol-induced retransmissions that create an adaptive multiple-input multiple-output (MIMO) system. This scheme, also known as NDMA (network diversity multiple access) with MPR, can achieve in ideal conditions a maximum throughput of M packets/time-slot. A further improvement is proposed here, where the receiver attempts to recover the information immediately after the reception of each (re)transmission. This is different from conventional NDMA, where this decoding process only occurs once the adaptive MIMO channel is assumed to become full-rank (i.e., once the estimated number of required retransmissions has been collected). The signals that are correctly decoded at every step of the proposed algorithm are used to mitigate interference upon the remaining contending signals by means of successive interference cancellation (SIC). This allows for improved reception as well as for the reduction of the number of retransmissions required to resolve a collision. Significantly high throughput figures that surpass the nominal rate of the system (T > M) are here reported. To the best of our knowledge this is the first random access protocol that achieves this figure. Correlation between antennas and between retransmissions, as well as imperfections of SIC are also considered. In ideal conditions, the effects of SIC are equivalent to a splitting tree operation. The inclusion of SIC in NDMA-MPR also opens the possibility of backwards compatibility with legacy terminals. The protocol achieves the highest throughput in the literature of single-hop wireless random access with minimum feedback complexity. This is a significant result for future highly dense 5G networks.

A Random Access Protocol incorporating Multi-Packet Reception, Retransmission Diversity and Successive Interference Cancellation

Ramiro Samano-Robles[1], Desmond C. McLernon[2], and Mounir Ghogho[3]

[1] Research Centre in Real-time and Embedded Computing Systems, Porto, Portugal
[2] School of Electronics and Electrical Engineering, University of Leeds, Leeds, UK.
[3] International University of Rabat, Morocco.
emails:[1]rasro@isep.ipp.pt, [2]d.c.mclernon@leeds.ac.uk; [3]m.ghogo@leeds.ac.uk

Abstract. This paper presents a random access protocol assisted by a set of signal processing tools that significantly improve the multi-packet reception (MPR) capabilities of the system. A receiver with M antennas is mainly used to resolve collisions with multiplicity $K \leq M$. The remaining unresolved conflicts (with multiplicity $K > M$ or with decoding errors) are processed by means of protocol-induced retransmissions that create an adaptive multiple-input multiple-output (MIMO) system. This scheme, also known as NDMA (network diversity multiple access) with MPR, can achieve in ideal conditions a maximum throughput of M packets/time-slot. A further improvement is proposed here, where the receiver attempts to recover the information immediately after the reception of each (re)transmission. This is different from conventional NDMA, where this decoding process only occurs once the adaptive MIMO channel is assumed to become full-rank (i.e., once the estimated number of required retransmissions has been collected). The signals that are correctly decoded at every step of the proposed algorithm are used to mitigate interference upon the remaining contending signals by means of successive interference cancellation (SIC). The use of SIC allows for improved reception as well as for the reduction of the number of retransmissions required to resolve a collision. Significantly high throughput figures that surpass the nominal rate of the system ($T > M$) are reported. To the best of our knowledge this is the first random access protocol that achieves this figure. Correlation between antennas and between retransmissions, as well as imperfections of SIC are also considered. In ideal conditions, the effects of SIC are found to be equivalent to a splitting tree operation. The inclusion of SIC in NDMA-MPR also opens the possibility of backwards compatibility with legacy terminals. The protocol achieves the highest throughput in the literature with minimum feedback complexity (identical to the conventional NDMA protocol). This is a significant result for future highly dense and 5G networks.

Keywords: Retransmission diversity, multi-packet reception, successive interference cancellation, cross-layer design

1 Introduction

The demand for wireless connectivity is rapidly increasing, particularly with the advent of the Internet-of-everything (IoE), machine-to-machine (M2M) communications, and 5G networks. However, the scarcity of spectrum resources has reignited interest in random access or decentralized allocation to reduce the signalling load and to facilitate opportunistic resource sharing. Random access protocols have experienced considerable progress over the last few years. A brief review related to the topic of this paper is now given, followed by the motivation for this work and finally we outline the contributions of this paper.

1.1 Background and previous works

In the design of legacy random access protocols, two main assumptions were commonly used: 1) packet collisions *always* lead to the *loss* of all the transmitted information, and 2) transmissions without collision are *always correctly received*. This so called *collision model* was relatively accurate in wire-line settings. However, in wireless scenarios, it shows severe drawbacks [1]. The collision model ignores any effects of the PHY-layer, which in the case of wireless channels is highly inaccurate. For example, collision-free transmissions can fail due to wireless channel fading and interference. Conversely, collisions can be resolved by signal processing or by capture effect. This ability to recover concurrent transmissions is known as multi-packet reception (MPR) [1]. It has been shown that the study of MPR-based random access protocols requires a new approach based on the co-design of physical (PHY) and medium access control (MAC) layers [1]-[2]. The effects of the PHY-layer need to be incorporated accurately in MAC-layer design.

The first complete MAC-PHY design was the study of ALOHA with MPR in [3], where a stochastic MPR matrix was used for throughput optimization under the assumption of a symmetrical infinite population. The extension to asymmetrical scenarios was then proposed in [4]. An extension to the case of decentralized channel state information was provided in [5], and optimization using game theory was proposed in [6]. The last years have witnessed an intensive research using stochastic geometry for random access in multi-hop networks using signal-to-interference-plus-noise ratio (SINR) reception models (e.g., [7]).

A breakthrough in cross-layer design was the work in [8], where collisions are resolved by retransmission diversity. The algorithm was called network diversity multiple access (NDMA). In NDMA, retransmissions create an adaptive MIMO (multiple-input multiple-output) system from which collisions can be resolved via multi-user detection. Stability analysis with perfect detection/separation is given in [9] in asymmetrical settings. Stability analysis of a symmetrical NDMA-MPR was later provided in [10] using imperfect user detection.

1.2 Motivations and objectives

NDMA achieves one of the highest values of throughput in wireless random access by using cross-layer concepts [8]. In NDMA, PHY-layer diversity is explicitly created by adaptive MAC retransmissions. In theory, even large collisions can be mitigated, mainly because enough diversity (retransmissions) can be adaptively created to resolve them. This means that in NDMA signals with unresolvable collisions are not discarded immediately as in ALOHA-type algorithms. Instead, they are stored in memory to be subsequently processed, thus leading to important capacity gains. However, there are still several issues and potential improvements on NDMA to be addressed. In the conventional *training-based* version of NDMA with M antennas at the base station (BS), the number of retransmissions is calculated based on an estimation of the collision multiplicity. If there is a collision of K users, then the BS requests $\lceil \frac{K}{M} \rceil - 1$ retransmissions in an attempt to create a *full-rank* MIMO channel (where $\lceil \cdot \rceil$ is the ceil integer operator) [8][10]. Therefore, the maximum throughput achieved by NDMA-MPR cannot exceed M packets/time-slot. This limit is both optimistic and pessimistic at the same time. It is optimistic because multi-user detection is not perfect under finite SNR (signal-to-noise ratio) scenarios. However, it is also pessimistic because there is a chance (using appropriate decoding and a slight change in the protocol operation) that some of the collisions can be resolved with *less retransmissions* than in conventional NDMA. This opens the possibility for improvements that will be further investigated in this paper. In addition, NDMA has not been yet analysed in realistic settings with temporal and spatial correlation. This paper also addresses these issues.

1.3 Contributions and organization

This paper proposes a significant improvement for NDMA-MPR, where the BS attempts the decoding of the contending signals immediately after the reception of each (re)transmission. This is different from conventional NDMA, where the BS waits until the MIMO channel created by retransmissions is considered to become full-rank before processing information. In our proposed scheme, if some of the signals are correctly recovered at any stage of the algorithm, then they are used in the subsequent decoding steps and retransmissions to mitigate their interference towards the remaining signals by means of successive interference cancellation (SIC). If the remaining contending signals can not be correctly decoded, then the BS does not discard them as in other SIC-based and ALOHA-type algorithms. Instead, further retransmissions are requested that will allow to process all the collected signals considering also the previously correctly decoded signals (via SIC). SIC has therefore two main functions in NDMA-MPR: improving of the probability of correct reception (as in other SIC-based algorithms), and mainly the *reduction of the number of retransmissions required to resolve a collision* (which is particular to NDMA). This leads to important gains allowing more information being correctly transmitted in a reduced number of time slots. SIC has been used in contention binary tree algorithms in [12] to also reduce the number of splitting operations (and therefore reducing the length of the contention resolution period), boosting throughput to a historical maximum (for tree algorithms) of 0.69 packets/time-slot. This paper has been inspired by [12] using SIC to improve NDMA-MPR.

So, our work considers a single-hop network with symmetrical/asymmetrical features. Symmetrical settings help in visualizing the main properties of the algorithm (such as maximum stable throughput or MST), while asymmetrical settings allow for interpretation in realistic scenarios. The protocol is shown to exceed M packets/time-slot, particularly for scenarios with high traffic loads, where a large number of retransmissions enhances the effects of SIC. To the best of our knowledge, this is the first protocol to achieve this throughput figure. The combination of MPR, NDMA, and SIC shows that it is possible to achieve the highest throughput in wireless random access with a minimum expense of feedback, which is shown to be the same as the conventional NDMA. The protocol transmits more packets with a given target value of operational SINR by using less time-slots (retransmissions) than all previous solutions. In addition, the study includes correlated retransmissions, correlated antenna signals, and random phase modulation, which have not been explored before in the context of NDMA. The paper also includes the effects of imperfect SIC. In ideal conditions, the effects of SIC are shown to be equivalent to a splitting operation of packets correctly decoded. Based on our results at full traffic load, retransmissions can be also used as a resource to be allocated in the uplink of centralized networks. Furthermore, the use of SIC also opens the possibility of backwards compatibility with legacy random access. More details follow in the main body of the paper.

This paper is organized as follows. Section 2 describes the scenario and protocol operation. Section 3 presents the metrics to be studied. Section 4 addresses the calculation of the throughput region. Section 5 provides the results of simulation, and finally Section 6 draws conclusions.

2 System model and protocol description

2.1 Scenario description

Consider the wireless slotted random access network in Fig. 1 with J terminals (each one with one antenna) and one BS with M antennas. The channel between user j and the mth antenna of the BS is denoted by $h_{j,m}$. All channels are assumed to be non-dispersive and Rayleigh distributed: $h_{j,m} \sim \mathcal{CN}(0, \sigma_j^2)$. Note that users are explicitly assumed as statistically different (asymmetrical). Results will be obtained for symmetrical settings as a particular case, and for asymmetrical settings as general case. Channels will be correlated in time and space: $E[h^*_{j,m}(n) h_{j,\tilde{m}}(\tilde{n})] = \rho^{(n,\tilde{n})}_{m,\tilde{m}} \sigma_j^2$, where $(\cdot)^*$ is the complex conjugate operator and $\rho^{(n,\tilde{n})}_{m,\tilde{m}}$ is the correlation coefficient between the signal of antenna m in time slot n with the signal of antenna $\tilde{m}$ in time slot $\tilde{n}$. All channels will be generated using a linear correlation model:

$$h_{j,m}(n) = \sqrt{R - \sum_{(\tilde{m},\tilde{n}) \neq (m,n)} \rho^{(n,\tilde{n})}_{m,\tilde{m}}} Y_{j,m,n}$$

$$+ \sum_{(\tilde{m},\tilde{n})\neq(m,n)} \sqrt{\rho^{(n,\tilde{n})}_{m,\tilde{m}}}\, Z^{(m,\tilde{m})}_{j,n,\tilde{n}}, \qquad Z^{(m,\tilde{m})}_{j,n,\tilde{n}} = Z^{(\tilde{m},m)}_{j,\tilde{n},n}, \tag{1}$$

where the variables $Y_{j,m,n}$ and $Z^{(m,\tilde{m})}_{j,n,\tilde{n}}$ are independent circular complex Gaussian random variables with zero mean and variance σ_j^2, and $R = M\left\lceil \frac{K}{M} \right\rceil - 1$. All terminals are assumed to always have a packet ready for transmission (full-queue or dominant system assumption). The packet transmission will be regulated by a random Bernoulli experiment with parameter p_j, which is the transmission probability of user j. Since collisions are resolved with adaptive retransmissions, the number of time-slots used in the resolution will be described by the random variable l. The period of time used to resolve a collision will be called contention resolution period or *epoch-slot*. The example in Fig. 1 with 5 users ($J = 5$) and two antennas ($M = 2$) shows a scenario with 4 epochs, three with active transmissions and one idle (no user transmits). One of the active epochs has length of two time-slots. This means that retransmission diversity was employed to resolve the collision.

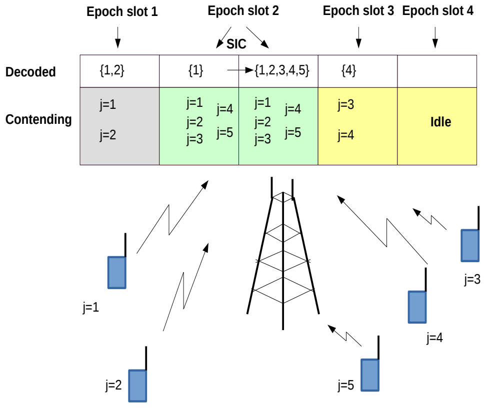

Fig. 1. Random access network with retransmission diversity, multi-packet reception and successive interference cancellation.

2.2 Principles of operation

The proposed protocol exploits *spatial* and *temporal (retransmission)* diversity to resolve collisions. Temporal diversity is adaptively created via retransmissions according to the number of contending signals that require further processing to be resolved. So, at the beginning of any epoch slot, the BS is assumed to know the identity of the contending users, and therefore it has perfect knowledge of the collision multiplicity[4] $K = |\mathcal{T}|$, where $\mathcal{T}$ denotes the set colliding users and $|\mathcal{U}|$ represents the cardinality of any set $\mathcal{U}$. In *conventional non-blind* NDMA, the BS requests $\lceil \frac{K}{M} \rceil - 1$ retransmissions from the contending users to improve rank conditions of the virtual MIMO channel [10]. Once all these retransmissions are received, then the legacy BS is assumed to attempt multi-user decoding. In contrast to this conventional algorithm operation, in our present work *the BS does not wait until all the estimated number of retransmissions have been collected.* Instead, the BS attempts the processing and full decoding of the information immediately after receiving each (re)transmission using a minimum mean square error (MMSE) receiver. Since the channel matrix is in this case *rank deficient*, then several of the contending signals are not likely to be correctly decoded. However, there is a non-zero probability that some of the signals can be actually correctly decoded. All these correctly decoded signals are used in subsequent decoding steps and time-slots to cancel their interference upon the remaining contending signals by using SIC. This scheme is repeated after the reception of each retransmission. The algorithm attempts to recover as much information as possible every time-slot. When it is no longer possible to decode information, further retransmissions are requested that will create further diversity. The algorithm stops requesting retransmissions once all the contending signals are correctly decoded or once the number of retransmissions reaches the value estimated in conventional NDMA ($\lceil \frac{K}{M} \rceil - 1$). The BS uses an ideal and instantaneous binary feedback flag to request retransmissions at the end of each time-slot.

Consider now the example of Fig. 1. The first epoch experiences two contending users, both of which have been correctly decoded in one time-slot because the BS has precisely two antennas and their decoding operation was successful. The second epoch experiences a collision of 5 users, which in the conventional algorithm requires $\lceil 5/2 \rceil - 1 = 2$ further retransmissions for resolution. However, user $j = 1$ has been correctly decoded in the first time slot of the epoch, and thus SIC can be used to remove its interference towards the other contending signals. The BS requests another retransmission, and this time, with the interference created by user $j = 1$ having been removed, the remaining 4 users can be decoded with two time diversity entries and two space diversity sources. Note that the second retransmission (which is necessary in the original NDMA algorithm) *was no longer required.* This illustrates the gains of the algorithm: a total of 5 transmissions were correctly received in *only two time slots*, which leads to an instantaneous rate of $5/2 = 2.5$ packets/time-slot, which exceeds the value of $M = 2$ of the nominal rate of this particular system. Note that in conventional schemes powered by SIC, once channel conditions do not allow for more packets to be correctly decoded, the algorithm stops and the non decoded information is discarded. By contrast, in our scheme retransmissions enable an additional

---

[4] Perfect collision multiplicity estimation can be closely achieved by using cooperative, sequential and multiple antenna processing combining [11][10]. Analysis considering the remaining errors of collision multiplicity estimation for the proposed protocol are part of future research work.

decoding step, thus serving as complement of the SIC cycle and allowing more packets to be correctly decoded without the need of discarding information. The third epoch in Fig. 1 shows a collision of two users with only one user being correctly decoded and no further retransmissions requested, as the collision multiplicity was below the BS antenna reception capability ($K < 1$). In this case, the decoding of the signal of user $j = 3$ has failed. Remaining errors are assumed to be handled by eroor control/correction in upper layers.

2.3 Signal model and protocol steps

The signal received by the mth antenna of the BS in time slot n of an epoch given a set $\mathcal{T}$ of contending users and under SIC operation can be written as:

$$\mathbf{y}_m(n) = \sum_{j\in\mathcal{T}} \check{h}_{j,m}(n)\mathbf{s}_j - \sum_{k\in\mathcal{T}_d(n)} \check{h}_{k,m}(n)\hat{\mathbf{s}}_k + \mathbf{v}_m(n), \tag{2}$$

where $\mathbf{s}_j = [s_j^{(0)}, \ldots, s_j^{(Q-1)}]^T$ is the packet with Q QAM symbols transmitted by user j; $\hat{\mathbf{s}}_k$ is the estimated signal of user k once it has been correctly decoded in the SIC cycle (described later in this subsection); $\mathcal{T}_d(n)$ indicates the subset of contending users that has been correctly decoded up to time slot n of an epoch slot; and $\mathbf{v}_m(n) = [v_m^{(0)}, \ldots, v_m^{(Q-1)}]^T$ is the vector of zero-mean Gaussian noise with variance σ_v^2: $v_m^{(q)}(n) \sim \mathcal{CN}(0, \sigma_v^2)$. In the previous expression, $\check{h}_{k,m}(n) = e^{-i\psi_{j,n}} h_{j,m}(n)$, where $i = \sqrt{-1}$ and $\psi_{j,n}$ is the uniformly distributed random phase used to counteract the effects of time correlation[5]. The second term in (2) corresponds to the operation of subtracting the interference of users correctly decoded in the previous steps from the main received signal. Note that under perfect SIC, once some of the signals are correctly decoded, their interference disappears in subsequent time-slots (via SIC). This is similar to a splitting operation in tree algorithms. This means that a PHY tool mimics a MAC operation with reduced feedback. This type of equivalence between PHY and MAC operations commonly arises in cross-layer design problems. Considering (2), all the collected (re)transmissions create a virtual MIMO system that can be expresed as follows [8] [9]:

$$\mathbf{Y}(n) = \mathbf{A}_{\bar{d}}(n)\mathbf{S}_{\bar{d}} + \xi\mathbf{A}_d(n)\mathbf{S}_d + \mathbf{V}(n), \tag{3}$$

where $\mathbf{Y}(n)$ is the array formed by all received signals across all of the Mn resources obtained from the combination of M antennas and n time-slots; $\mathbf{A}_d(n)$ and $\mathbf{A}_{\bar{d}}(n)$ denote the MIMO channels of the set of correctly decoded and not correctly decoded users, respectively; $\mathbf{S}_d$ and $\mathbf{S}_{\bar{d}}$ are the arrays of stacked packets for both sets of users; ξ is the parameter that will be used to measure different degrees of efficiency of SIC and which depends on several PHY-layer factors such

---

[5] In NDMA, time correlation degrades full-rank condition of the virtual MIMO channel matrix. Random phase modulation aims to partially counteract this effect by explicitly introducing randomness in the transmitted signals at each time-slot [8].

as channel estimation, the SIC stage and modulation format; and finally $\mathbf{V}(n)$ is the noise. The channel matrices can be estimated using a convenient training sequence design (see [8]). In this paper we assume perfect channel estimation. The non-decoded signals can be processed at the BS using a linear decoding matrix $\mathbf{G}(n)$ given by the MMSE criterion: $\mathbf{G}(n) = (\mathbf{A}_{\bar{d}}^H(n)\mathbf{A}_{\bar{d}}(n) + \sigma_v^2\mathbf{I}_{nM})^{-1}\mathbf{A}_{\bar{d}}^H(n)$, where $\mathbf{I}_{nM}$ is the identity matrix of order nM and $(\cdot)^H$ is the hermitian transpose operator. This operation applied to (3) leads to:

$$\hat{\mathbf{S}}_{\bar{d}}(n) = \mathbf{W}_{\bar{d}}(n)\mathbf{S}_{\bar{d}} + \xi\mathbf{W}_d(n)\mathbf{S}_d + \mathbf{G}(n)\mathbf{V}(n), \tag{4}$$

where $\mathbf{W}_d(n) = \mathbf{G}(n)\mathbf{A}_d(n)$ and $\mathbf{W}_{\bar{d}}(n) = \mathbf{G}(n)\mathbf{A}_{\bar{d}}(n)$. The signal for user j in time-slot n in (4) will experience an instantaneous SINR given by:

$$\mathit{\Gamma}_j(n) = \frac{|W_{\bar{d}}^{(j,j)}(n)|^2}{\eta_j(n) + |\mathbf{g}_j(n)|^2\sigma_v^2}, \tag{5}$$

where $W_{\bar{d}}^{(j,k)}(n)$ denotes the entry of matrix $\mathbf{W}_{\bar{d}}(n)$ that corresponds to the row and column of user j and user k, respectively, and $\mathbf{g}_j(n)$ is the row of matrix $\mathbf{G}(n)$ corresponding to user j. The interference term in (5) is given by:

$$\eta_j(n) = \sum_{k\neq j, k\notin\mathcal{T}_d(n)} |W_{\bar{d}}^{(j,k)}(n)|^2 + \sum_{k\in\mathcal{T}_d(n)} \xi^2|W_d^{(j,k)}(n)|^2.$$

A packet is assumed to be correctly received if the SINR exceeds a threshold β (SINR reception model). This threshold depends on the modulation format in use and the operational block error rate (BLER)[6]. Both β and ξ will be obtained via PHY-layer simulation. The decoding operation previously described is repeated every time-slot of an epoch, and it continues until there is no change in the set of decoded users or until the maximum number of retransmissions ($\lceil\frac{K}{M}\rceil - 1$) has been collected. These steps are repeated for subsequent epoch-slots. The steps of this algorithm are summarized in Table 1. Some additional advantages can be observed in the proposed algorithm. The SIC cycle decodes packets sequentially according to SINR conditions. This includes signals that are part of the initial set of users but which do not appear again, or which only appear a few random times during the resolution period. These are typical signals of terminals in ALOHA or binary tree mode or interfering signals. Therefore, the algorithm could potentially incorporate legacy terminals by recovering their signals at the end or at different stages of the decoding cycle, and also automatically reject co-channel interference.

3 Performance metrics

The conditional reception probability $q_{\mathcal{S}|\mathcal{T}}(n)$ is defined as the joint probability that all users in $\mathcal{S} \subseteq \mathcal{T}$ experience a value of SINR $\mathit{\Gamma}_j(n)$ in (5) above threshold

---

[6] Under the SINR reception model, the throughput represents the average information transmitted per packet with a given BLER performance. The SINR reception model is commonly used in random access to incorporate enriched PHY-layer information in protocol design. It is also used for the study of SIC in random access (e.g. [13])

```
1. Generate $\mathcal{T}$ using traffic model. Start of an epoch.
2. Set $n = 1$, $\mathcal{T}_d(n) = \mathcal{R}(n) = \emptyset$.
3. while $n \leq \lceil \frac{K}{M} \rceil$ & $|\mathcal{T}_d(n)| < |\mathcal{T}|$ do
       (a) Calculate $\mathit{\Gamma}_j(n)$ in (5) $\forall j \in \mathcal{T}$,
       (b) $\mathcal{T}_d(n) = \{j | \mathit{\Gamma}_j(n) > \beta\}$
       (c) while $|\mathcal{T}_d(n)| > |\mathcal{R}(n)|$ do
               i. $\mathcal{R}(n) = \mathcal{T}_d(n)$;
               ii. Calculate $\mathit{\Gamma}_j(n)$ in (5) $\forall j \notin \mathcal{T}_d(n)$;
               iii. $\mathcal{T}_d(n) = \{j | \mathit{\Gamma}_j(n) > \beta\}$;
           end
       (d) $n = n + 1$
   end
```

Algorithm 1: Algorithm NDMA-MPR-SIC

β conditional on the set of contending users $\mathcal{T}$, and on the epoch length l having reached the value n:

$$q_{\mathcal{S}|\mathcal{T}}(n) = \Pr\{\cap_{j \in \mathcal{S}} \mathit{\Gamma}_j(n) > \beta | \mathcal{T}, l = n\}.$$

Note that this probability is conditional on the previous decoding events in the epoch slot that have lead to a realization of n time slots (i.e., the collision was not resolved in previous slots). The throughput of terminal j is defined as the ratio of the probability of correct reception over all epoch realizations (S_j) to the average length of an epoch ($E[l]$):

$$T_j = \frac{S_j}{E[l]} = \frac{\sum_{n, \mathcal{T}, j \in \mathcal{T}} \Pr\{\mathcal{T}\} \Pr\{l = n | \mathcal{T}\} q_{j|\mathcal{T}}(n)}{\sum_{n, \mathcal{T}} n \Pr\{\mathcal{T}\} \Pr\{l = n | \mathcal{T}\}}, \tag{6}$$

where $\Pr\{\mathcal{T}\}$ indicates the probability of realization of the set of contending users $\mathcal{T}$, and $\Pr\{l = n|\mathcal{T}\}$ indicates the probability mass function (pmf) of the epoch length conditional on the set of colliding users $\mathcal{T}$. These terms can be expressed as:

$$\Pr\{\mathcal{T}\} = \prod_{k \in \mathcal{T}} p_k \prod_{j \notin \mathcal{T}} \bar{p}_j,$$

and

$$\Pr\{l = n | \mathcal{T}\} = \begin{cases} q_{\mathcal{T}|\mathcal{T}}(n) \prod_{t=1}^{n-1} \bar{q}_{\mathcal{T}|\mathcal{T}}(t), & n < \lceil K/M \rceil \\ \prod_{t=1}^{n-1} \bar{q}_{\mathcal{T}|\mathcal{T}}(t) & n = \lceil K/M \rceil \\ 1, & K = 0, n = 1 \end{cases},$$

where $\bar{(a)} = 1 - (a)$, for any a. The last expression indicates that a retransmission is requested whenever the resolution process has failed during the current time slot. It also indicates that the resolution period ends when all the colliding users are correctly decoded or when the maximum number of (re)transmissions ($\lceil K/M \rceil$) of the original NDMA-MPR scheme have been received. All the conditional reception probabilities will obtained via PHY-layer simulation and later used in a MAC-PHY simulation/analysis, as shown later in the section of results. Consider now the vector $\mathbf{T} = [T_1, T_2, \dots T_J]^T$ of stacked *throughput* values, and the vector $\mathbf{p} = [p_1, p_2, \dots p_J]^T$ of transmission probabilities. The *throughput region*, is defined as the union of all achievable values $[T_1, T_2, \dots T_J]$ over all possible transmission policies $(0 \leq p_j \leq 1)$[14]:

$$\mathcal{C}_{\tilde{T}} = \{\tilde{\mathbf{T}} | \tilde{T}_j = T_j(\mathbf{p}), 0 \leq p_j \leq 1\}. \tag{7}$$

The throughput region is one of the main metrics in the study of random access in asymmetrical settings[14].

4 Optimization

To derive the envelope of the throughput region in (7), a multi-objective optimization of the J objective functions in (6) is proposed:

$$\mathbf{p}_{opt} = \arg\max_{\mathbf{p}} \quad [T_1, T_2, \dots T_J], \qquad 0 \leq p_j \leq 1. \tag{8}$$

Since this vector optimization usually lacks a unique solution [15], the concept of Pareto optimality is commonly employed. A Pareto optimal solution can be loosely defined here as the point that is at least optimum for one or more of the elements of the vector objective function (see [15] for a complete definition). The multi-objective optimization problem in (8) can be transformed into a single objective optimization problem using the method of scalarization [15]:

$$\mathbf{p}_{opt} = \arg\max_{\mathbf{p}} \sum_{k=1}^{J} \mu_k T_k, \quad 0 \leq p_j \leq 1, \tag{9}$$

where μ_k is the relative weight given to the kth objective function. Differentiating the objective function in (9) we obtain a set of equations given by $\sum_{k=1}^{J} \mu_k \frac{\partial T_k}{\partial p_j} = 0, \quad k, j = 1.., J$. The solution of this set of linear equations, independent from the values of the weighting factors μ_k, can be proved, in our context, to be equivalent to setting the following Jacobian determinant to zero [16]:

$$|\mathbf{J}| = 0, \qquad 0 \leq p_j \leq 1, \tag{10}$$

where $J_{j,k} = \frac{\partial T_k}{\partial p_j}$ is the (j, k) element of the Jacobian matrix $\mathbf{J}$. The optimum transmission policy can be obtained by solving the Jacobian determinant equation in (10) using the expressions of throughput. A closed-form solution of this problem is in general difficult to obtain. This paper proposes a method that provides a solution in closed-form by considering that the desired solution is a deviation from the solution of an equivalent collision model protocol. The solution for the optimization of the throughput region of random access protocols under the collision model results in Jacobian $\tilde{\mathbf{J}}_a$ square matrices (following the lines of the derivation of the expression in (10)) that have the following property:

$$\tilde{J}_a(j, k) = \begin{cases} x_j, \, k = j \\ y_j, \, k \neq j \end{cases} \tag{11}$$

which means that all the elements of a row j are all the same except for the element of the main diagonal. Under this structure, the Jacobian determinant $|\tilde{\mathbf{J}}_a|$ has been proved in ([16]) to be equal to

$$|\tilde{\mathbf{J}}_a| = 1 - \sum_j \left\{ \frac{y_j}{x_j - y_j} \right\}.$$

The structure of the Jacobian matrix for the MPR cases in general does not have the same structure as in the collision model. However, the elements can be arranged in a way that is quasi-symmetrical or slightly approximate to a collision model matrix. We can then propose a complement that produces the desired quasi-symmetrical property. This can be mathematically expressed as follows:

$$\tilde{J}_a(j,k) = J_a(j,k) + \dot{J}_a(j,k),$$

where $\dot{J}_a(j,k)$ is the element of Jacobian matrix $\dot{\mathbf{J}}_a$ that complements the original Jacobian matrix $\mathbf{J}_a$ to acquire the desired symmetrical property defined in (11). The Jacobian determinant can now be obtained (using the well known co-factors formula) as the determinant of the symmetrical collision model matrix component $\tilde{\mathbf{J}}_a$ minus the deviation component that can be obtained by analysing each one of the components (co-factors) of the complement matrix $\dot{\mathbf{J}}_a$. This can be mathematically expressed as follows:

$$|\mathbf{J}_a| = |\tilde{\mathbf{J}}_a| - \sum_{j=1}^{J} (-1)^j \{ J_a(1,j)(|\tilde{\mathbf{J}}_a^{1,j}| - |\mathbf{J}_a^{1,j}|) + \dot{J}_a(1,j)|\tilde{\mathbf{J}}_a^{1,j}| \}, \qquad (12)$$

where $\mathbf{B}^{k,j}$ denotes the submatrix that is formed by removing the k-th row and the jth column of matrix $\mathbf{B}$. The expression in (12) can be solved via different numerical methods to obtain the boundary of the throughput region, as shown in the next section.

5 Results

This section presents the results that show the gains of the proposed protocol. Symmetrical and asymmetrical results will be presented. Fig. 2 and Fig. 4 show the throughput versus the transmission probability ($p_j = p, \forall j$) for a symmetrical system. The results have been calculated in a scenario with $J = 16$ users experiencing an average SNR of $\frac{\sigma^2}{\sigma_v^2} = 3$ dB ($\sigma^2 = \sigma_j^2$, $\forall j$) with noise variance $\sigma_v^2 = 1$. The results were obtained using different values of spatial and temporal correlation, and using two antenna elements ($M = 2$). For simplicity, we have considered that all antennas experience the same spatial correlation between each other (ρ_s), and also all retransmissions experience the same temporal correlation (ρ_t). Note that in a realistic scenario retransmissions and antennas farther apart in time and space from each other must experience lower correlation than those being close to each other. Therefore, our assumption of symmetrical correlation represents the worst case scenario in the correlation model in (1).

The values of β, ξ and all the conditional reception probabilities of the theoretical formulae obtained in previous sections have been obtained via PHY-layer simulation with perfect channel estimation, using a packet length of $Q = 100$ BPSK symbols, and considering a packet is correctly decoded when the number of symbol errors is below $N_e = 10$. The effects of different SIC stages have also been considered when obtaining ξ. This means that packets with a given BLER contain some symbol errors that can contribute to the incorrect decoding of other packets when their signals are used to subtract interference via SIC (i.e., error propagation). The value ξ is therefore used to model the residual interference caused by the symbols containing some errors in the set of users considered as correctly decoded.

The algorithm with and without random phase modulation is tagged, respectively, NMSP and NMS. The acronyms indicate the combination of NDMA, MPR, SIC, and phase modulation. For purposes of comparison, the figure also includes the results of the conventional NDMA-MPR with and without phase modulation and for ALOHA MPR (labelled AM). The results show the high performance gains of the proposed algorithm and how random phase modulation helps to alleviate, although not completely counteract, the effects of temporal correlation in both the proposed algorithm and in the conventional NDMA-MPR (both algorithms show some improvement). We recall the reader that the random phase modulation helps in inducing full-rank conditions in scenarios with high temporal channel correlation. Spatial correlation is also shown to reduce throughput due to lack of diversity gains, particularly for ALOHA. Note that in all cases the proposed algorithm has surpassed M packets/time-slot, which means the scheme achieves, to the best of our knowledge, one of the highest values of throughput in the literature of random access. This means that the protocol exploits as much as possible the available resources, dynamically adapting to the different collision sizes by enabling retransmissions to complement the baseline multi-packet reception capabilities of the system, and using SIC to refine detection and to eliminate those retransmissions that are not necessary to resolve the collision. The algorithm also improves access delay and energy consumption, mainly because the use of SIC reduces the average number of time-slots required to resolve a collision.

It can be observed that at full traffic load ($p \to 1$), the gains of the algorithm are particularly attractive. At full load the algorithm resembles a centralized resource allocation problem, where retransmissions can be used for resource allocation in the time domain. This means that in centralized uplink systems users can be scheduled to share the exact same transmission resource over different time-slots, mimicking a collision that can be resolved via the proposed algorithm combining multiple antennas, retransmissions and SIC.

Fig. 3 shows results using realistic uncoded packets with BPSK modulation and with $Q = 100$ symbols. The SINR model is not used in this case. All packets are generated and processed according to the description of the algorithm. All the other settings are identical to the settings used in Fig. 2. The results prove that the algorithm performs very similar to the case studied in Fig. 2 with a performance that exceeds the barrier of M packets/time-slot. It can be observed in both cases that at full traffic load, the gains of the algorithm are particularly attractive.

In the previous two figures, we have assumed imperfect SIC operation. This means that remaining symbol errors in some of the packets considered as correctly decoded can cause error propagation when using SIC. The results in Fig. 4 assume perfect operation of SIC. The objective is to assess the potential gains of SIC. All the simulation settings are identical to the previous two cases, except for the value of ξ, which in this case was set deliberately to zero ($\xi = 0$). We can observe in the results that the gains can double the nominal rate of the system ($T > 4$), which is a significant result. This opens the possibility of designing an appropriate modulation and coding scheme that can achieve nearly perfect SIC and thus boost throughput figures close to the upper limit shown in Fig. 4.

Fig. 5 shows the throughput region for a system with two subsets of users, with J_1 and J_2 users each, where $J_1 = J_2 = 8$. A Newton-Raphson algorithm was used to solve (12) for different values of transmission probabilities and then obtain the boundary of the throughput region, which is also the Pareto frontier of the optimization problem in (8). The average SNR of both sets is set to 3 dB and 5 dB, respectively, with all other settings identical to the previous cases. The outer curve represents the boundary of the NDMA-MPR-SIC protocol, whereas the inner curve represents the NDMA-MPR protocol. SIC enlarges the throughput region exceeding (in most of the cases) the nominal rate of M packets/time slot. An enlarged throughput regions brings several benefits to a random access protocol, not only in terms of individual capacity figures, but also in the design of low-complex linear resource allocation policies within the boundaries of the region. It has been also proved in previous works (i.e., [14]), that convex throughput regions allow protocols also to achieve a better trade-off between different metrics such as fairness, power consumption, and sum-throughput.

6 Conclusions

This paper has presented a significant improvement over the NDMA-MPR protocol by using SIC. Unlike the conventional protocol, the proposed scheme allows the BS to attempt the decoding of the colliding signals immediately after the reception of each (re)transmission, rather than waiting until the created (virtual) MIMO channel is assumed to become full-rank. If some of the signals are correctly decoded, then an additional decoding cycle based on SIC is used to improve reception of the remaining signals and to also reduce the number of retransmissions required in the original algorithm. The result is a protocol that: (i) considerably improves the performance of previous algorithms using minimum feedback complexity (highest throughput in literature); (ii) is potentially backwards compatible with legacy random access, and (iii) that is also a potential solution for centralized systems. Finally, the extension to semi-centralized and multi-hop systems using stochastic geometry and imperfect collision multiplicity detection is an attractive future research topic that is currently under progress.

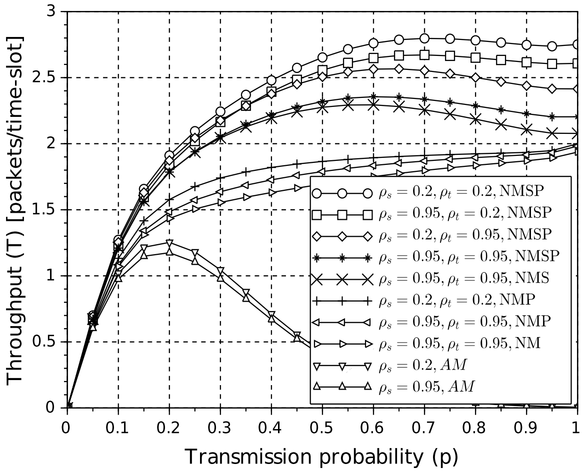

Fig. 2. Throughput T vs. Tx. probability p using $M = 2$, and $\frac{\sigma_j^2}{\sigma_v^2} = 3\text{dB}$, with real uncoded packets, parameterised on ρ_s, ρ_t and algorithm type (NMSP, NMS, NM and AM). Note that $p = p_j\ \forall j$.

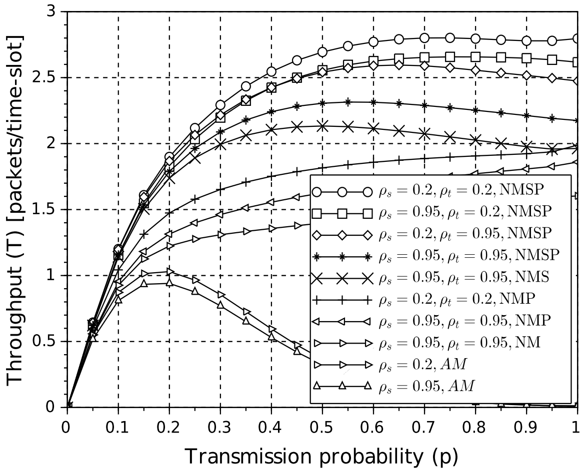

Fig. 3. Throughput T vs. Tx. probability p using $M = 2$, and $\frac{\sigma_j^2}{\sigma_v^2} = 3\text{dB}$, with real uncoded packets, parameterised on ρ_s, ρ_t and algorithm type (NMSP, NMS, NM and AM). Note that $p = p_j\ \forall j$.

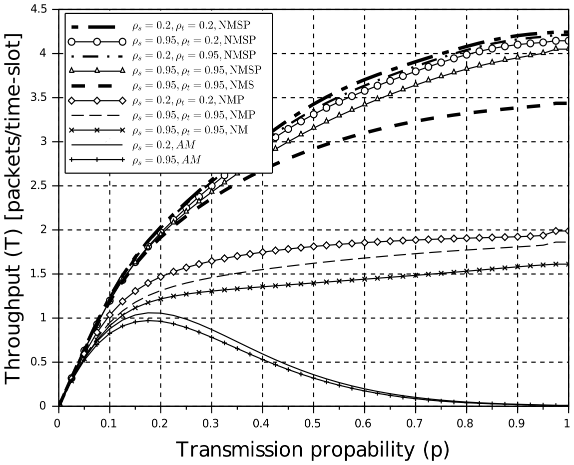

Fig. 4. Throughput T vs. Tx. probability p using $M = 2$, and $\frac{\sigma_j^2}{\sigma_v^2} = 3$dB, with real uncoded packets, parameterised on ρ_s, ρ_t and algorithm type (NMSP, NMS, NM and AM). Note that $p = p_j\ \forall j$.

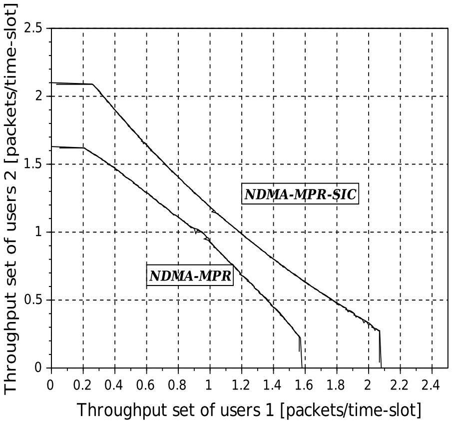

Fig. 5. Throughput region for a system with two subsets of users, $J_1 = J_2 = 8$, with SNRs of 3 dB and 5dB, respectively

References

1. G. Dimic, N.D. Sidiropoulos, and R. Zhang "Medium access control-physical cross-layer design," *IEEE Sig. Proc. Mag.,* vol 21, no. 5, pp. 40-50, Sep. 2004.
2. V. Srivastaya and M. Montani "Cross-layer design: a survey and the road ahead," *IEEE Commun. Mag.*, vol. 43, no. 12, pp.112-119, Dec. 2005,
3. S. Ghez, S. Verdu and S. Schwartz, "Stability properties of slotted Aloha with multipacket reception capability," *IEEE Transactions on Automatic Control,* vol. 33, no. 7, pp. 640-649, July 1988.
4. V. Naware, G. Mergen, and L. Tong, "Stability and delay of finite-user slotted ALOHA with multipacket reception," *IEEE Transactions on Information Theory,* vol. 51, no. 7, pp. 2636-2656, Jul. 2005.
5. M.H. Ngo, V. Krishnamurthy, and L. Tong, "Optimal hannel-Aware ALOHA Protocol for random access in WLANs with multipacket reception and decentralized CSI," *IEEE Trans. on Sig. Processing,* vol. 56, no.6, pp. 2575-2588, June 2008.
6. M.H. Ngo and V. Krishnamurty, "Game theoretic cross-layer transmission policies in multipacket reception wireless networks," *IEEE Trans. on Sig. Proc.* , vol. 55, no. 5, pp. 1911-1926, May 2007.
7. M.K. Hanawal, E. Altman, and F. Baccelli, F., "Stochastic geometry based medium access games in wireless Ad Hoc networks," *IEEE Journal on Selected Areas in Comm.,* vol.30, no.11, pp.2146-2157, Dec. 2012
8. M.K. Tsatsanis, R. Zhang, and S. Banerjee, "Network-assisted diversity for random access wireless networks," *IEEE Transactions on Sig. Proc.*, vol. 48, no. 3, pp. 702-711, Mar.2000.
9. G. Dimic, N.D. sidiropoulos, and L. Tassiulas, "Wireless networks with retransmission diversity access mechanisms: stable throughput and delay properties," *IEEE Trans. on Sig. Proc.*, vol. 51, no. 8, pp. 2019-2030, Aug. 2003
10. R. Samano-Robles and A. Gameiro "Stability properties of network diversity multiple access protocols with multiple antenna reception and imperfect collision multiplicity estimation," *Journal of Computer Networks and Commun.,* vol. 2013, Article ID. 984956, 10 pages.
11. R. Samano-Robles, M. Ghogho, and D.C. McLernon, "Cooperative and sequential user detection for wireless network diversity multiple access protocols,," *3rd Int. Symposium on Wireless Per. Comp., ISWPC 2008.* pp.189,193, May 2008.
12. Y. Yingqun and G.B. Giannakis, "High-throughput random access using successive interference cancellation in a tree lgorithm," *IEEE Transactions on Info. Theory,* vol. 53, no. 12, pp. 4628-4639, Dec. 2007.
13. M. Mollanoori, and M. Ghaderi, "On the performance of successive interference cancellation in random access networks," *IEEE CS Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON),* pp.461,469, June 2012.
14. J. Luo and A. Ephremides "On the throughput, capacity, and stability regions of random multiple access," *IEEE Trans. on Info. Theory,* vol. 52, no. 6, pp. 2593-2607, June 2006.
15. S. Boyd and L. Vandenberghe, *Convex optimization,* Cambridge University Press, 2004.
16. R. Samano-Robles, M. Ghogho and D.C. McLernon "Wireless networks with retransmission diversity and carrier sense multiple access," *IEEE Transactions on Sig. Proc.*, vol. 57, no. 9, pp. 3722-3726, Sept. 2009.