Proceedings Paper:

Robles, R, Tovar, E, Lara, M et al. (2017) Network Diversity Multiple Access in Rayleigh Fading Correlated Channels with Imperfect Channel and Collision Multiplicity Estimation. In: 24th Telecommunications Forum (TELFOR) 2016. 24th Telecommunications Forum (TELFOR 2016), 22-23 Nov 2016, Belgrade, Serbia. IEEE. ISBN: 978-1-5090-4086-5.

https://doi.org/10.1109/TELFOR.2016.7818814

Network Diversity Multiple Access in Rayleigh Fading Correlated Channels with Imperfect Channel and Collision Multiplicity Estimation

Ramiro Robles, Eduardo Tovar, Mauricio Lara, Aldo Orozco, Desmond C. McLernon, and Mounir Ghogho

***Abstract*—Network diversity multiple access or NDMA is the family of algorithms with the highest potential throughput in the literature of signal-processing-assisted random access. NDMA uses the concept of *protocol-induced retransmissions* to create an adaptive source of physical (PHY) layer diversity. This adaptive diversity is used to resolve packet collisions (via signal separation) without the explicit need (or as a complement) of a multiple antenna receiver. This paper proposes a further improvement on the modelling of NDMA by considering the effects of imperfect channel and collision multiplicity estimation. In addition, this work considers channel correlation between consecutive retransmissions (i.e., temporal correlation). Conventionally, the analysis of NDMA assumes that any error in the collision multiplicity estimation translates into the loss of all contending packets. This is an optimistic assumption because even when the multiplicity has been correctly estimated, errors can still occur. On the other hand, it is also pessimistic because correct reception can also occur when the multiplicity has been incorrectly estimated. This paper presents a more detailed study of the performance of NDMA considering these more specific detection/reception cases.**

I. Introduction

Cross-layer design is an important tool in future random access networks. Correct reception now depends on physical (PHY)-layer performance, as well as traffic load conditions [1]. A breakthrough in this topic was the work in [2], where collisions were resolved using a new type of diversity based on retransmissions. The algorithm was called network diversity multiple access (NDMA). In NDMA, retransmissions are used to create a virtual MIMO (multiple-input multiple-output) system from which colliding signals can be recovered via source separation. Signals with collisions that can not be resolved immediately are not discarded as in conventional protocols. They are initially used to estimate the collision multiplicity. Based on this information, the base station (BS) requests further retransmissions from the contending terminals in an attempt to create a *full-rank* MIMO system. The BS uses the stored signals to resolve the collision via source separation. A cooperative NDMA protocol was later proposed in [3]. NDMA with multi-packet reception was proposed in [4] with a finite user model. Stability analysis of NDMA with perfect collision multiplicity estimation and packet reception can be found in [5]. A Markov model for NDMA stability analysis was presented in [6].

In NDMA, the collision multiplicity estimation is used to determine the number of retransmissions that are necessary to resolve the collision. Too many retransmissions translates into a waste of resources and throughput degradation. Too few retransmissions means that full-rank conditions will be probably lost rendering the incorrect decoding of signals. Conventional modelling of NDMA is based upon the assumption that any error in the collision multiplicity estimation means the loss of all the contending packets [2]. However, this assumption is both optimistic and pessimistic at the same time. It is optimistic because even in the case of correct estimation of the collision multiplicity, packet decoding errors can still occur. It is also pessimistic because some packets can still be correctly decoded in case of incorrect estimation of the collision multiplicity. In addition, the protocol has only been analysed under the assumption of uncorrelated retransmissions and perfect channel estimation. This paper addresses these issues by reformulating all protocol expressions based on a more accurate model with all the potential cases of correct or incorrect packet reception, and under the assumption of incorrect estimation of the collision multiplicity. The model also includes the effects of channel estimation errors, as well as the effects of correlation between consecutive retransmissions (temporal correlation).

The remainder of this paper is organised as follows. Section II describes the system scenario. Section III deals with the signal model. Section IV presents the performance analysis of the protocol. Section V presents results of the performance of the protocol, and finally Section VI presents the conclusions.

II. System model

A. System scenario and channel model

Consider the slotted random access network with retransmission diversity depicted in Fig. 1 with a set of J buffered one-antenna terminals and one central node or base station (BS) with one receiver antenna. The channel between terminal j and the BS in time-slot n is denoted by $h_j(n)$. All channel envelopes are assumed to be non-dispersive with Rayleigh

Funded by FCT/MEC (Fundação para a Ciência e a Tecnologia), ERDF (European Regional Development Fund) under PT2020, CISTER Research Unit (CEC/04234), and by ARTEMIS/0004/2013-JU grant nr. 621353 (DEWI, www.dewi-project.eu)

Ramiro Robles and Eduardo Tovar are with Research Centre in Real-time and Embedded Computing Systems, Porto . emails: {rasro,emt}@isep.ipp.pt
Mauricio Lara and Aldo Orozco are with the Centro de Investigación y de Estudios Avanzados (CINVESTAV), México. emails {mlara,aorozco}@cinvestav.mx
Desmond C. McLernon is with the School of Electronics and Electrical Engineering, University of Leeds, Leeds, UK. emails: d.c.mclernon@leeds.ac.uk
Mounir Ghogho is with the International University of Rabath, Morocco. email: m.ghogho@leeds.ac.uk

statistics: $h_j(n) \sim \mathcal{CN}(0, \gamma)$. Signals experience identical correlation across (re)transmissions (i.e., temporal correlation). This means that $E[h_j^*(n) h_j(\tilde{n})] = \rho_r \gamma$, where ρ_r is the temporal correlation coefficient, $(\cdot)^*$ is the complex conjugate operator, and $E[\cdot]$ is the statistical average operator. For simplicity in analysis, time-slot notation n in all variables will be dropped in subsequent derivations.

		e=1		e=2		e=3			e=4	
Contending terminals	$\mathcal{T}$	$\{1,3\}$	$\{1,3\}$	$\{3,8,9\}$	$\{3,8,9\}$	$\{1,8\}$			$\{5,7\}$	$\{5,7\}$
Detected terminals	$\hat{\mathcal{T}}$	$\{1,3\}$	$\{1,3\}$	$\{3,8\}$	$\{3,8\}$	$\{1,3,8\}$	$\{1,3,8\}$	$\{1,3,8\}$	$\{5,7\}$	$\{5,7\}$
Correctly decoded	$\mathcal{T}_d$		$\{1,3\}$		$\{3\}$			$\{1,8\}$		$\{5\}$
Collision multiplicity	K	2		2		2			2	
Estimated multiplicity	$\hat{K}$	2		2		3			2	
Feedback signal	ξ	1	0	1	0	1	1	0	1	0

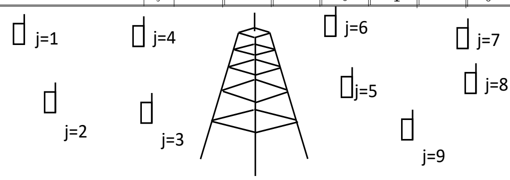

Fig. 1. Random access network assisted by retransmission diversity.

B. Protocol overview

Since NDMA exploits the time domain to create diversity, the number of time-slots used to resolve any packet collision will be described by a random variable denoted here by l. This period of time used to resolve a packet collision will be called contention resolution period or *epoch-slot* (see Fig. 1). At the beginning of every epoch-slot each terminal will be assumed to attempt a packet transmission provided a packet is available for transmission in its queue. All terminals will be assumed to experience a Poisson-distributed random arrival process described by the parameter λ. The probability of transmission at the beginning of each epoch-slot will be denoted by p. The set of contending terminals during the first time-slot of any epoch-slot will be denoted by $\mathcal{T}$.

At the beginning of every epoch-slot, the BS proceeds to estimate the collision multiplicity as described in detail in Section III-B. The BS obtains an estimation $\hat{K} = |\hat{\mathcal{T}}|$ of the collision multiplicity $K = |\mathcal{T}|$, where $\hat{\mathcal{T}}$ indicates the set of terminals detected as active and $|\cdot|$ is the set cardinality operator when applied to a set variable. Once this information has been obtained, the BS estimates the number of retransmissions required to resolve the collision. Since the BS has one antenna, the number of transmissions (i.e., initial transmission plus retransmissions) required in the non-blind version of NDMA is given by $\widehat{K}$ [6]. This means that the number of diversity sources must be greater than or equal to the estimated collision multiplicity.

In NDMA, having more diversity sources than contending signals is necessary to maximize the probability that the channel matrix is *full-rank*, which in turn improves the probability of success of the source separation technique to be used [2]. To request a retransmission for diversity purposes, the BS simply indicates with an ideal and instantaneous binary feedback flag $\xi \in \{0, 1\}$ at the end of each time-slot to all the contending terminals that retransmission is required in the next time slot. The feedback binary flag is kept *on* ($\xi = 1$) until all necessary retransmissions have been collected. These protocol steps are repeated for subsequent epoch-slots.

To further illustrate the mechanism of the NDMA protocol, in Fig. 1 we can observe four realizations of epochs. In the first epoch-slot ($e = 1$), two terminals $\mathcal{T} = \{1, 3\}$ have collided in the first time-slot. The figure indicates the main variables of the system: the set of contending terminals $\mathcal{T}$, the set of terminals detected as active $\hat{\mathcal{T}}$, the set of terminals with correctly decoded signals $\mathcal{T}_d$, the collision multiplicity K, the estimated collision multiplicity $\hat{K}$ and the binary feedback flag ξ used to request retransmissions. Since two signals need to be recovered in the first epoch, then only one more retransmission is needed to potentially resolve the collision. Note that in this first epoch the set of detected terminals is identical to the set of contending terminals ($\mathcal{T}_d = \mathcal{T}$), which means no presence detection errors occurred. In this case, the number of collected signals is equal to two, which is enough to attempt the recovery of the two contending signals. Also note that the binary feedback is only set to $\xi = 1$ at the end of the first time-slot. Once the first retransmission has been received, the value is set to $\xi = 0$, which means that the current epoch has finalized so the contending terminals stop retransmitting while a new epoch-slot starts.

The second epoch ($e = 2$) experiences three contending terminals given by $\mathcal{T} = \{3, 8, 9\}$, which ideally requires two retransmissions plus the original transmission to be resolved. However, only 2 terminals given by $\hat{\mathcal{T}} = \{3, 8\}$ were detected as active (terminal $j = 9$ has not been correctly detected). Therefore, the system only requests one more retransmission instead of two. This leads to a *rank-deficient* MIMO system, which in turn can lead to excessive decoding errors. Only a subset of the contending signals have been actually correctly decoded: $\mathcal{T}_d = \{3\}$. The next epoch ($e = 3$) experiences $K = 2$ contending terminals given by $\mathcal{T} = \{1, 8\}$, but the BS has detected $\hat{K} = 3$ terminals given by $\hat{\mathcal{T}} = \{1, 3, 8\}$, thus falsely considering terminal $j = 3$ as active. Note that detecting one more terminal has caused the BS to request one more retransmission than actually needed. This is a waste of resources. However, it can be observed that in this case all the signals were correctly decoded by the BS, even with an error of collision multiplicity estimation. The last epoch ($e = 4$) shows the case where all terminals were correctly detected, but the detection process was still incorrect for one of the contending terminals. These examples of epoch realizations aim to illustrate the variety of cases of correct/incorrect detection and decoding that might arise in NDMA.

III. Signal models

A. Signal model for terminal detection

Each terminal is pre-assigned with a unique orthogonal code that is used for purposes of terminal activity detection and channel estimation [2]. The BS uses a matched-filter operation to extract the detection indicator of terminal j, denoted here by

z_j. This indicator is then compared to a detection threshold β to decide whether terminal j is active or not. If $z_j < \beta$ then the terminal is detected as inactive: $j \notin \widehat{\mathcal{T}}$. Otherwise, if $z_j \geq \beta$ then the terminal is detected as active ($j \in \widehat{\mathcal{T}}$). This means that $\hat{\mathcal{T}} = \{j | z_j \geq \beta\}$ is the set of all terminals whose detection indicator exceeds the threshold β. Since this detection process is prone to errors due to fading and noise, two cases of presence detection ($j \in \hat{\mathcal{T}}$) can be identified: 1) terminal j can be correctly detected as active with probability P_D (probability of correct detection) provided the terminal has transmitted a packet ($j \in \mathcal{T} \cup \hat{\mathcal{T}}$), and 2) terminal j is incorrectly detected as active with probability P_F (probability of false alarm) provided the terminal did not transmit a packet ($j \in \hat{\mathcal{T}}\ j \notin \mathcal{T}$). Analytical expressions for P_D and P_F have been obtained for Rayleigh fading channels in [2].

B. Signal model for multi-packet reception

Each terminal j transmits packets with Q QAM symbols denoted by $\mathbf{x}_j = [x_j(0), x_j(1) \ldots x_j(Q-1)]^T$, where $(\cdot)^T$ is the vector transpose operator. Considering unitary power transmission $E[\mathbf{x}_j^H \mathbf{x}_j] = 1$, where $(\cdot)^H$ is the Hermitian transpose operator, the signal vector received at the beginning of an epoch is given by $\mathbf{y} = \sum_{j \in \mathcal{T}} h_j \mathbf{x}_j + \mathbf{v}$, where $\mathbf{v} = [v(0), v(1) \ldots, v(Q-1)]^T$ is the zero-mean and white complex Gaussian noise vector with variance σ_v^2: $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}_Q, \sigma_v^2 \mathbf{I}_Q)$ where $\mathbf{0}_Q$ and $\mathbf{I}_Q$ denote, respectively, the vector of Q zeroes and the identity matrix of order Q. The BS proceeds to estimate the collision multiplicity by means of terminal activity detection (explained in the previous subsection) and requests the number of necessary retransmissions (given by $\widehat{K} - 1$) to resolve the collision. All the collected (re)transmissions are stored in memory to create a virtual MIMO system that can be expressed as follows [2] [5]: $\mathbf{Y}_{\widehat{K} \times Q} = \mathbf{H}_{\widehat{K} K} \mathbf{S}_{K \times Q} + \mathbf{V}_{\widehat{K} \times Q}$, where $\mathbf{Y}$ is the array formed by the collection of all received signals from all the $\widehat{K}$ time-slots of the epoch, $\mathbf{H}$ is the mixing matrix or MIMO (multiple-input multiple-output) channel, $\mathbf{S}$ is the array of stacked packets from all the contending terminals, each one with Q symbols, and finally $\mathbf{V}$ represents the collected Gaussian noise components. The mixing matrix $\mathbf{H}$ can be estimated by using the outcome of the matched filter operation from each retransmission [2]. The estimate $\hat{\mathbf{H}}$ can be used to recover the contending packets. The contending signals can be estimated at the BS by means of a linear decoding matrix $\mathbf{G}$: $\hat{\mathbf{S}} = \mathbf{GY} = \mathbf{GHS} + \mathbf{GV}$. This expression can be rewritten as follows: $\hat{\mathbf{S}} = \mathbf{W}_1 \mathbf{S}_1 + \mathbf{W}_2 \mathbf{S}_2 + \mathbf{GV}$, where $\mathbf{W}_1 = \mathbf{GH}_1$, $\mathbf{W}_2 = \mathbf{GH}_2$, $\mathbf{H}_1$ is the mixing matrix of the contending terminals that have been detected as active ($j \in \hat{\mathcal{T}}, j \in \mathcal{T}$), and $\mathbf{H}_2$ is the mixing matrix of the contending terminals that have not been detected as active ($j \notin \hat{\mathcal{T}}, j \in \mathcal{T}$). The decoding matrix can be calculated using zero-forcing (ZF) or minimum mean square error (MMSE) equalization. The decoding signal for terminal j will experience a signal-to-interference-plus-noise ratio (SINR) given by:

$$\Gamma_j = \frac{|W_1(j,j)|^2}{\sum_{k \neq j} |W_1(j,k)|^2 + \sum_k |W_2(j,k)|^2 + |\mathbf{g}_j|^2 \sigma_v^2}, \quad (1)$$

where $W_1(j,k)$ and $W_2(j,k)$ denote the entries of matrix $\mathbf{W}_1$ and $\mathbf{W}_2$, respectively, that correspond to row and column of terminal j and terminal k, respectively, and $\mathbf{g}_j$ is the row of matrix $\mathbf{G}$ corresponding to terminal j. It is assumed that a packet is correctly received when the SINR in (1) exceeds a decoding threshold denoted here by β_d. The probability of a terminal transmission to be correctly decoded is thus denoted by $\Pr\{\Gamma_j > \beta_d\}$.

IV. Reception model and performance metrics

The correct packet reception probability of q out of K contending signals provided K_d contending signals were correctly detected as active and K_f inactive terminals were incorrectly detected as active (false alarm) can be expressed as follows:

$$C_{K_d,K_f}^{(q,K)} = \left(\begin{array}{c} K \\ K_d \end{array} \right) \left(\begin{array}{c} J-K \\ K_f \end{array} \right) \times$$
$$\Pr\{\cup_{j \in \mathcal{T}_d} \Gamma_j > \beta_d | K, K_d, K_f\}, \quad q = |\mathcal{T}_d|, \quad \mathcal{T}_d \subset \mathcal{T} \cap \hat{\mathcal{T}}, \quad (2)$$

where $\left(\begin{array}{c} n_1 \\ n_2 \end{array} \right)$ is the combinatorial number of n_1 elements in n_2 positions, and $\Pr\{\cup_{j \in \mathcal{T}_d} \Gamma_j > \beta_d | K, K_d, K_f\}$ is the probability that q contending signals are correctly decoded as their experienced SINR exceeds the decoding threshold β_d conditional on the number of contending terminals K, the number of contending terminals correctly detected as active K_d and the number of inactive terminals incorrectly detected as active K_f (false alarm).

A. Throughput

Packet throughput can be defined here as the ratio of the average number of packets correctly received (denoted by S) to the average length of an epoch-slot ($E[l]$):

$$T = \frac{S}{E[l]}, \quad (3)$$

With the help of the reception model in (2), the numerator of (3) can be mathematically written as:

$$S = \sum_{K=1}^{J} \sum_{q=1}^{K} \sum_{K_d=q}^{K} \sum_{K_f=0}^{J-K} \left(\begin{array}{c} J \\ K \end{array} \right) q (p P_D)^{K_d} (\bar{p} P_F)^{K_f} C_{K_d,K_f}^{(q,K)}, \quad (4)$$

where $\bar{(\cdot)} = 1 - (\cdot)$, which means that $\bar{p} = 1 - p$. The expression in (4) represents the average of correct packet reception over all possible cases of transmission and terminal activity detection (correct and incorrect). The average length of an epoch in the denominator of (3) can be obtained by averaging over all possible cases of terminal activity detection, i.e., when an active terminal is correctly detected as active, or when an inactive terminal is incorrectly detected as active (false alarm). We recall the reader that the number of time-slots of each epoch is determined by the number of retransmissions necessary to make the MIMO system full-rank, which in our setting is given by $\widehat{K}$ [6]. The probability mass function (PMF) of length of an epoch l is thus given by:

$$\Pr\{l = m\} = \left\{ \begin{array}{cc} \Pr\{\hat{K} = m\}, & m > 1 \\ \Pr\{\hat{K} = 0\} + \Pr\{\hat{K} = 1\}, & m = 1 \end{array} \right. \quad (5)$$

It can be also proved that $\hat{K}$ has a binomial distribution with parameter $P_A = pP_D + \bar{p}P_F$, which can be written as $\Pr\{\hat{K} = k\} = \left(\begin{array}{c} J \\ k \end{array} \right) P_A^k \bar{P}_A^{J-k}$, $k = 0, \dots J$. Therefore, $E[l]$ can be obtained by averaging over the PMF of l in 5, which yields $E[l] = JP_A + \bar{P}_A^J$, where the second term $\bar{P}_A^J$ stands for the contribution of one time slot in case any terminal is detected as active: $\Pr\{\hat{K} = 0\} = \bar{P}_A^J$. The parameter P_A is thus regarded as the total probability of terminal activity detection, and is given by the probability of correct detection in case of transmission plus the probability of false alarm in case of no transmission: $P_A = \Pr\{j \in \hat{\mathcal{T}}\} = \Pr\{j \in \hat{\mathcal{T}} | j \in \mathcal{T}\}\Pr\{j \in \mathcal{T}\} + \Pr\{j \in \hat{\mathcal{T}} | j \notin \mathcal{T}\}\Pr\{j \notin \mathcal{T}\} = pP_D + \bar{p}P_F$.

V. Results

Let us now present some results that show the concepts explored in the previous sections. Consider a scenario with $J = 16$ terminals with an average signal-to-noise ratio (SNR) of $\frac{\gamma}{\sigma_v^2} = 10$ dB. All simulation results assume a packet decoding threshold for the signal-to-interference-plus-noise (SINR) ratio of $\beta = 2.5$, above which a packet is considered to be correctly received by the BS. Fig. 2 shows the results for packet throughput versus traffic load ($J\lambda$) for different values of temporal correlation coefficient ρ_r. Two types of decoder were used: MMSE and ZF. Fig. 2 shows the results of the detection throughput (labelled "Detect. T."), which is the assumption made in the conventional analysis of NDMA where packets are only correctly received by the BS when there is no detection errors. It does not include the potential errors due to multi-user decoding. This curve is useful as reference for all other results. The results that include the errors due to multi-user decoding are labelled in Fig. 1 with the subscript 0". For example, the results using ZF decoder with correlation coefficient $\rho_r = 0.2$ are simply labelled $ZF_0, \rho_r = 0.2$. The results show that further decoding errors of the multi-user detection stage reduce the throughput with respect to the predicted value given by the detection throughput. Also note that the effect of correlation tends to reduce throughput performance at high values of traffic load, even affecting the stability bound (the maximum value of traffic load before the throughput curve rapidly decreases). However, at low traffic load, the results show that the highly correlated case can even slightly surpass the case with low correlation. The results without subscript have been obtained by using the concepts developed in this paper. Even in the case of incorrect detection of the collision multiplicity, the system attempts the decoding of the contending signals. It can be observed that for both types of decoder ZF and particularly for MMSE, the throughput of the protocol surpasses that one of the conventional assumption (with subscript 0). In one of the cases the obtained throughput can even surpass the detector throughput, which means that more potential gains in NDMA can be obtained by using signal processing post-collision multiplicity estimation. The effects of correlated retransmissions are similar to the previous case analysed here. These results suggest that there can be some cases where NDMA can obtain benefit from correlated retransmissions.

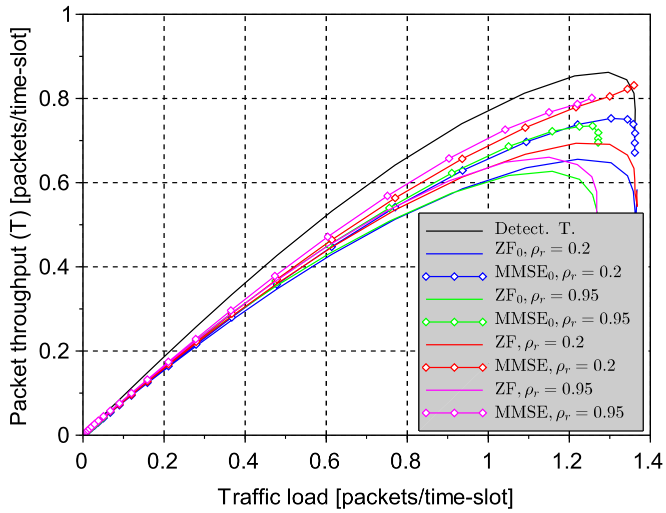

Fig. 2. Packet throughput versus traffic load for different decoding schemes and using different values of time-correlation ρ_r.

VI. Conclusions

This paper has presented a more detailed analysis of the operation of a class of random access protocols assisted by retransmission diversity and signal processing tools for multi-user detection. The conventional analysis of the protocol ignores several details of correct detection and packet decoding that have been addressed in this work. It was found that temporal correlation and imperfect channel and collision multiplicity estimation can affect the stability and throughput performance at high traffic loads. By comparison, low temporal correlation and MMSE decoding can even surpass the predicted detection throughput of the protocol, which opens further possibilities for improvement of the performance of this type of algorithm.

References

[1] L. Tong, V. Naware and P. Venkitasubramaniam, "Signal processing in random access," *IEEE Sig. Proc. Mag.,* , vol. 21, no. 5, Sep. 2004, pp. 29-39.

[2] M.K. Tsatsanis, R. Zhang, and S. Banerjee, "Network-assisted diversity for random access wireless networks," *IEEE Transactions on Signal Processing*, Vol. 48, No. 3, March 2000, pp. 702-711.

[3] L. Dong and A.P. Petropulu, "Multichannel ALLIANCES: A cooperative cross-layer scheme for wireless networks," *IEEE Transactions on Signal Processing*, Vol.56, No. 2, pp. 771-784, Feb. 2008.

[4] R. Samano-Robles, M. Ghogho, and D.C. McLernon, "A multi-access protocol assisted by retransmission diversity and multipacket reception," *IEEE ICASSP*, Las Vegas, Nevada, 2008, pp. 3005-3008.

[5] 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, Aug. 2003.

[6] 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 Comm.,* Vol. 2013, No. 984956, pp. 1 - 10, Dec. 2013.