Random Multiple Access in WiMAX: Problems and Solutions

Alexey Vinel, Dirk Staehle, Rastin Pries

Department of Distributed Systems, Institute of Computer Science,
University of Würzburg, Am Hubland, D-97074 Würzburg, Germany
vinel@ieee.org

Abstract. The development of media access control protocols for next generation metropolitan wireless networks is a challenging task nowadays. WiMAX is the most promising candidate for such a ubiquitous broadband wireless access system. This paper discusses both theoretical and practical issues of the random multiple access algorithm standardized in IEEE 802.16. A simple and efficient alternative for the binary exponential backoff is presented and analyzed. Tradeoffs, advantages and disadvantages of the introduced approach are investigated. The analysis is conducted by means of both analysis and simulation.

Keywords: WiMAX, Binary Exponential Backoff, Random Access, Collision Resolution.

1 Introduction

Random multiple access is a well-known method used in communication systems, where a very large number of stations occasionally transmit packets by means of a single channel. This is exactly the case on the uplink channel of next generation wireless networks, such as IEEE 802.16 WiMAX [1]. Consequently, efficient random multiple access algorithms should be developed to provide fast delivery of a request message from any subscriber station to the base station.

A classical model exists, which is used for the investigation of random multiple access algorithms (see the detailed description by Gallager [2]). In this model, all transmissions are synchronized by means of so-called slots. In case two or more stations simultaneously transmit in the same slot, a collision occurs. Retransmissions are performed using some algorithm, which is based on the knowledge of the situations occurred on the channel.

Two performance metrics are traditionally considered for the random access algorithm analyses: mean delay and tenacity. According to Capetanakis [3] the mean delay D is defined as a ratio between the total packet delay measured for a very long time period and the number of packets successfully transmitted during this time. Tenacity R is introduced by Tsybakov in [4] and defined as a maximal packet arrival rate λ, under which a finite mean delay is still provided. In case an infinite number of stations with a Poisson arrival process is assumed, tenacity becomes a simple measure of the algorithm efficiency. For instance tenacity of the simplest Aloha algorithm is known to be zero [2], the tree algorithm [5] has a tenacity value of 0.3662 and the fastest known part-and-try algorithm [6] provides 0.4878.

Though, random access algorithms have been developed since the beginning of 70s, only the so-called *binary exponential backoff* (BEB) is widely known and used in practice, for instance in local area networks. The BEB, which is implemented in real systems, is difficult to analyze due to a uniform distribution of a waiting time interval and its tenacity computation is an open question. Nevertheless, its modified "theoretical" version with geometrically distributed waiting interval is investigated by Aldous in [7]. He shows that the BEB is unstable in the sense that $N(t)/t$ converges to zero as t goes to infinity for any non-zero arrival rate, where $N(t)$ is the number of successful transmissions made during the time $[0,t]$.

The popularity and belief in the BEB efficiency comes from its simplicity and a good performance of classical IEEE 802.3 networks. Indeed, for the relatively small number of stations (which can be potentially connected into the collision segment of a local area network) the carrier-sense feature of IEEE 802.3 makes the probability of a collision very small. Thus, the collision resolution algorithm actually does not significantly influence the performance. In future wireless networks, the number of stations may vary from hundreds to thousands, while no carrier-sense mechanism can be implemented. This leads to a high collision probability and consequently the need of developing a simple and efficient algorithm for such a network. The latest and original survey on known random access algorithms in the framework of wireless communications can be found in Chlebus [8].

In contrast to the classical models from [2] and [8], where there is no central coordinator instance, in [9] a model with a base station is introduced. The presence of a base station provides an opportunity to implement special random access algorithms for such a system.

In this paper we introduce a simple algorithm for the centralized networks, which makes use of a base station ability to distinguish the situations in slots from the very beginning of the system operation. We numerically compute the tenacity of this algorithm, which is shown to be high enough, namely approximately 0.45.

Moreover, we continue our analysis of the BEB algorithm in the framework of IEEE 802.16, which has been started in [10]. We present a way to numerically compute the tenacity, which does not require Markov chain modeling.

Finally, we perform a comparative analysis of the developed algorithm and BEB particularly by means of simulations and determine the conditions, when the usage of our approach is reasonable.

The rest of the paper is organized as follows. Section II provides a brief description of the IEEE 802.16 MAC protocol and the model for the random access used throughout this paper. The analysis of the binary exponential backoff is conducted in Section III, while in Section IV we present and investigate our alternative algorithm and compare it with BEB.

2 Random Access Model for WiMAX Networks

2.1 Basics of IEEE 802.16 MAC Operation

We focus on the IEEE 802.16 MAC operation as an example for possible applications of our analysis. Throughout this paper, we consider a network with a point-to-multipoint (PMP) architecture, which consists of one base station managing several subscriber stations. Transmissions between the BS and SSs are realized in fixed frames by means of time division multiple access (TDMA) / time division duplexing (TDD) mode of operation. The frame structure consists of a downlink sub-frame for transmissions from the BS to SSs and an uplink sub-frame for transmissions in the reverse direction. The Tx/Rx transition gap (TTG) and Rx/Tx transition gap (RTG) shall be inserted between the sub-frames to allow terminals to switch from reception to transmission and vice versa. In the downlink sub-frame the Downlink MAP (DL-MAP) and Uplink MAP (UL-MAP) messages are transmitted, which comprise the bandwidth allocations for data transmission in downlink and uplink direction, respectively.

Another important management message which is interconnected with the UL-MAP is the Uplink Channel Descriptor (UCD), which can be periodically transmitted in the downlink sub-frame. The values of the minimum backoff window, $W_{\min}$, and maximum backoff window, $W_{\max}$, are defined in this message, which are used for the BEB collision resolution algorithm.

The uplink sub-frame contains of transmission opportunities scheduled for bandwidth request purposes, in which a Bandwidth Request (BW-REQ) message can be transmitted, which serves for SSs to indicate to the BS that they need a UL bandwidth allocation. The BS manages the number of transmission opportunities through the UL-MAP message.

2.2 Model for the Random Access

Let us consider n subscriber stations, simply denoted by stations in the sequel, having a buffer sufficient to store exactly one request. A station, which has a request in the considered moment of time is referred to as “active”, otherwise it is called “non-active”. In this paper we consider two models of the system, namely, the *finite number of stations model* (denoted as $n < \infty$) and the *infinite number of stations model* ($n = \infty$).

The time axis is divided into frames. Each frame comprises K equal slots for the random access. The duration of a slot corresponds to the time needed for one bandwidth request transmission. The BS chooses K in order to make a trade-off between the duration of the contention period and the duration of payload transmissions within the whole frame duration, which is fixed. Therefore, in the following discussion, K is assumed to be a fixed value.

We define the arrival rate λ as the mean number of requests appearing in the system for the slot duration. For the $n < \infty$ model, each “non-active” station generates a request during the frame duration with probability $\pi=(K\lambda)/n$. For $n = \infty$ it is assumed that requests are arriving into the system according to a Poisson process with parameter λ. In both cases, a new request is put into the buffer and transmitted not earlier than in the next frame according to some RMA algorithm.

Three situations should be distinguished in any slot, namely: “empty” (E) – nobody transmitted in the slot, “success” (S) – exactly one station transmitted in the slot and “collision” (C) – at least two or more stations transmitted in the slot. We assume ideal channel conditions, i.e., if exactly one station transmits in a slot, the transmission is successful, otherwise a collision occurs. Furthermore, we assume that stations receive a feedback from the BS at the beginning of the next frame whether their transmission was successful or not.

3 Binary Exponential Backoff Analysis

3.1 Algorithm Description

In IEEE 802.16 a binary exponential backoff algorithm is introduced for collision resolution. Before each transmission attempt, a station uniformly chooses an integer number from the interval $[0, W_i–1]$, where W_i is the current value of its backoff window. A chosen value, referred to as the backoff counter, indicates the number of slots the station has to wait before the transmission of a request. For the first transmission attempt, the backoff window W_0 is set to $W_{\min}$. In case of a collision, a station doubles its backoff window value, and so the backoff window after i collisions, W_i, becomes $2^i W_{\min}$. The window is not doubled if it reaches the maximum value $W_{\max} = 2^m W_{\min}$, where m is referred to as the maximum backoff stage. In the case of a successful transmission, the backoff window is set to the minimum value $W_{\min}$.

The standard IEEE 802.16 does not define any relationship between the parameters $W_{\min}$, $W_{\max}$, and K. If $W_{\min}<K$, then some time slots will never be used during the first transmission attempt. We set $W_{\min} = lK$, where l is a natural number, in order to uniformly distribute the transmission attempts over the available random access slots.

3.2 Tenacity Derivation

Let us consider the $n < \infty$ model and assume that each station has always a request to transmit. Following the approach of [10] and [11], we assume that the behavior of an arbitrary station does not depend on the behavior of the other $i – 1$ stations, and the collision probability p that a station transmits and falls into collision or the request is distorted by noise, is constant. Under such an assumption, a two-step procedure is proposed: in the first step, a station uniformly chooses one of the L_w frames to transmit, where $L_w = 2^w l$, $w = 0,\ldots,m$, and w describes the current backoff stage. In the second step one out of K slots is uniformly chosen in the given frame.

A discrete and integer time scale is adopted, where t and t+1 correspond to the beginning of two consequent frames. Let $c(t)$ be the stochastic process representing the integer number of frames a station has to wait before transmission at time t. So, the station transmits in a frame, which starts at the moment t, if $c(t)$ equals to zero. Let b(t) be the backoff stage of a station at the moment t. Note that the two-dimensional process $\{c(t), b(t)\}$ is a renewal one, because according to protocol rules after a successful transmission its evolution does not depend on its previous history.

Consider the process of transmitting a request. Let $\overline{N}$ be the average number of transmission attempts, and $\overline{K}$ – the mean number of frames where subscriber does not transmit during this process. Consequently, the equation for the probability of a station to transmit in a frame x can be obtained [12]:

$$x = \frac{\overline{N}}{\overline{N} + \overline{K}}. \tag{1}$$

It is clear that number of transmission attempts is a geometrically distributed random variable, thus $\overline{N} = 1/(1-p)$. Let $\overline{K}_i$ be the mean number of frames a subscriber deferred its transmission, in case exactly i attempts have been required for a successful transmission, then

$$\overline{K} = (1-p)\sum_{i=1}^{\infty} \overline{K}_i p^{i-1}. \tag{2}$$

It can be shown (we omit mathematical simplifications for short) that the following equations hold

$$\overline{K}_i = \begin{cases} 2^{i-1} l - (l+i)/2 & 1 \le i \le m+1 \\ 2^m l \dfrac{i-m+1}{2} - \dfrac{l+i}{2} & i > m+1 \end{cases}. \tag{3}$$

Substituting (2) and (3) into (1) after the algebraic simplifications we obtain

$$x = \frac{2(1-2p)}{(1-2p)(l+1) + pl(1-(2p)^m)}. \tag{4}$$

Let us consider that one particular station transmits in a frame. Under this condition, the probability y_u that u stations from the remaining i-1 that transmit in the same frame, is equal to

$$y_u = \binom{i-1}{u} x^u (1-x)^{i-1-u} \,. \quad \textbf{(5)}$$

and the probability that all of them transmit in the slots different from the one chosen by the considered station is $(1 - 1/K)^u$. Thus, the conditional collision probability p is

$$p = 1 - \sum_{u=0}^{i-1} \binom{i-1}{u} x^u (1-x)^{i-1-u} \left(1 - \frac{1}{K}\right)^u \,. \quad \textbf{(6)}$$

So, the non-linear system is represented by Equations (4) and (6) with two unknowns p and x and probability x could be calculated numerically.

Compute the probability $R_{r,k}$ that r stations out of total k active ones successfully transmit in a frame. Denote $N(r,k,K)$ – the total number of ways to put k undistinguishable balls into K boxes under the condition that exactly r boxes contain one ball. This function is computed recursively using the following rule

$$\begin{gathered} N(0,0,K) = 1 \,,\ N(0,k,0) = 0 \,, \\ N(0,k,K) = K^k - \sum_{s=1}^{\min(k,K)} N(s,k,K) \,, k > 0 \,; \\ N(r,k,K) = \binom{k}{r}\binom{K}{r} \cdot r! \cdot N(0, k-r, K-r) \,,\ 0 < r \le \min(k,K) \,. \end{gathered} \quad \textbf{(7)}$$

and the conditional collision probability $R_{r,k}$ equals to

$$R_{r,k} = \frac{N(r,k,K)}{K^k} \,. \quad \textbf{(8)}$$

Finally, tenacity of BEB R_1 for a given number of stations n may be computed as a mean value of requests successfully transmitted per slot:

$$R_1(n) = \Big(\sum_{k=0}^{n} \sum_{r=0}^{k} r R_{r,k} \binom{n}{k} x^k (1-x)^{n-k}\Big) / K \,. \quad \textbf{(9)}$$

4 Alternative Algorithm Analysis

4.1 Algorithm Description

We present the following algorithm as an alternative for the BEB. First, we introduce its informal description. Let all available K slots be divided into two groups, namely first S slots are used for the transmission of new requests and $N = K - S$ slots for collision resolution. A subscriber, having a request to transmit, waits till the beginning of the nearest frame and then uniformly chooses one of K slots to transmit. In case a collision occurs, all involved subscribers form a so-called *collision set*. A collision set is put into a specially organized virtual distributed queue and is considered to be served, when all subscribers in this set transmit successfully.

The first N collision sets in the head of the queue are served in parallel, corresponding to the number of slots for collision resolution. The decision to transmit is made by a subscriber from the collision set with probability 0.5. In case "collision" or "empty" situations take place in a slot for collision resolution, subscribers perform another transmission attempt with probability 0.5. Otherwise, (in case of a "success") subscribers left in the collision set transmit with probability 1. Thus, in order to determine the current length of the distributed queue and current position of each subscriber, simple rules may be used. "Collision" occurred in any out of K slots means that a new collision set will be formed. Two situations "success", happened in two consequent frames with subscribers from some collision set, mean that this set has been served successfully.

Now let us present the formal rules which should be implemented on the base station and subscribers' side to enable the algorithm operation. As mentioned in Section II, a BS should determine the situations in the slots and at the beginning of frame t provide the subscribers a ternary feedback vector $F_t = (f_1(t), f_2(t), \ldots, f_K(t))$, where f_i has three possible values, depending on the situation seen in each slot of frame $t - 1$: 0 – "empty", 1 – "success", 2 – "collision". Moreover, other values should be transmitted to all stations, namely: Q_t – number of sets in the distributed queue by the beginning of frame number t and a binary vector $D_t = (d_1(t), d_2(t), \ldots, d_N(t))$, where $d_i(t) = 1$ in case some collision set has been served in slot number $S + i$ of frame number $t - 1$, and $d_i(t) = 0$ otherwise.

Assume that a BS "sees" the channel state in contention slots from the very beginning of the system operation. By the beginning of frame t, the BS itself stores Q_{t-1}, F_{t-1} (processed according to algorithm below), and F_t and uses this information to compute Q_t and D_t in the following way:

- for each i from 1 to S, if $f_i(t) = 2$ then $Q_t \leftarrow Q_{t-1} + 1$;

- for each i from S+1 to N, if $f_i(t - 1) = f_i(t) = 1$ then $Q_t \leftarrow Q_{t-1} - 1$, $f_i(t) \leftarrow 0$, $d_i(t) \leftarrow 1$.

For all i: $d_i(t) = 1$ perform the shift – for any j: $i < j < N$ do $f_j(t) \leftarrow f_{j+1}(t)$. Then make assignment $F_{t-1} \leftarrow F_t$. Finally, $d_i(t) \leftarrow 0$ for all i.

In contrast to the base station, subscriber stations may enter the system at an arbitrary moment of time. Each subscriber station stores the following data by the beginning of frame t: q_t – the current position of the subscriber in the distributed queue, $c(t)$ – the number of a slot to transmit a request. For new subscribers q_t equals to zero and $c(t)$ equals to –1. After receiving Q_t, F_t, and D_t, each subscriber station performs the following actions:

1. If the subscriber has a new request ready to transmit, then $c(t)$ is uniformly chosen from [1, S]. The Subscriber performs a transmission attempt in the chosen slot.

2. If the subscriber has transmitted a request in frame $t-1$ in slot number $1 \le i \le K$ and $f_i(t)$=1, then this request is considered to be successfully transmitted.

3. If the subscriber has transmitted a new request in frame $t-1$ in slot number $1 \le i \le S$ and $f_i(t)$=2, then this request becomes backlogged, $q_t \leftarrow Q_t$ and for each j from $i+1$ to S, if $f_j(t) = 2$ then $q_t \leftarrow q_t - 1$. This step is done to determine the position of new collided requests in a distributed queue.

4. For any subscriber having $q_t \ge 1$ for each j from 1 to $\min(q_t, N)$, if $D_j(t) = 1$ then $q_t \leftarrow q_t - 1$. This step is done to determine the position of backlogged requests in a distributed queue due to shift.

5. If for some subscriber $1 \le q_t \le N$ and $c(t) = -1$ then this subscriber transmits with probability 0.5 in the slot number $q_t + S$ of the t-th frame and assigns $c(t) \leftarrow q_t + S$.

6. If for some subscriber $1 \le q_t \le N$ and $c(t) \ne -1$ and $f_{c(t)}(t) = 1$ then this subscriber transmits with probability 1 in slot number $q_t + S$ of the t-th frame ($c(t) \leftarrow q_t + S$).

7. If for some subscriber $1 \le q_t \le N$ and $c(t) \ne -1$ and $f_{c(t)}(t) \ne 1$ then this subscriber transmits with probability 0.5 in slot number $q_t + S$ of t-th frame ($c(t) \leftarrow q_t + S$).

4.2 Tenacity Derivation

Let us consider the $n = \infty$ model and, as mentioned in Section II, assume a Poisson request arrival process with arrival rate λ. Compute R_2 – tenacity of the developed algorithm. In each slot for new requests, in average $s\lambda = \lambda(S + N)$/S requests arrive. The mean number of the collision sets arriving into the distributed queue for the frame duration equals to

$$\Lambda = \sum_{i=0}^{S} i \binom{S}{i} (1 - e^{-\lambda s} - \lambda s e^{-\lambda s})^i (e^{-\lambda s} + \lambda s e^{-\lambda s})^{S-i} = S(1 - e^{-\lambda s} - \lambda s e^{-\lambda s}),$$

while the mean number of collision sets served in N slots may be computed in the

$$\Psi = \frac{1 - e^{-\lambda s} - \lambda s e^{-\lambda s}}{\sum_{i=2}^{\infty} T_i \frac{(\lambda s)^i}{i!} e^{-\lambda s}} N,$$

following way

We denote T_k as the mean time needed for serving the collision set of k subscribers. It is calculated recursively

$$T_2 = 3, T_k = \frac{1}{k\frac{1}{2}(1-\frac{1}{2})^{k-1}} + 1 + T_{k-1} = 2^k / k + 1 + T_{k-1}\cdot$$

Finally, using the stability equation for the considered queuing system $\Lambda = \Psi$ and applying (10) we may numerically obtain tenacity R_2 from:

$$\sum_{i=2}^{\infty} T_i \frac{(R_2 s)^i}{i!} e^{-\lambda s} = \frac{N}{S}\,. \quad \textbf{(10)}$$

Now let us compute S and N which maximize the tenacity (see Table). Some optimal parameters values for different number of contention slots $K = S + N$ are shown.

Table. Optimal Tenacity Values

S / *N*	1	2	3	4	5	6	7	8
1	**0,4493**	0,4363	0,4075	0,38	0,3557	0,3345	0,3159	0,2994
2	0,413	**0,4493**	0,4478	0,4363	0,4221	0,4075	0,3934	0,38
3	0,3748	0,4334	**0,4493**	**0,4501**	0,4446	0,4363	0,427	0,4172
4	0,3437	0,413	0,4397	0,4493	**0,4507**	0,4478	0,4427	0,4363
5	0,3185	0,3931	0,4267	0,4426	0,4493	**0,4508**	0,4493	0,4459
6	0,2979	0,3748	0,413	0,4334	0,4442	0,4493	**0,4508**	0,4501
7	0,2807	0,3584	0,3996	0,4233	0,4373	0,4452	0,4493	**0,4508**
8	0,2661	0,3437	0,3868	0,413	0,4295	0,4397	0,4459	0,4493
9	0,2535	0,3304	0,3748	0,4029	0,4213	0,4334	0,4414	0,4464
10	0,2425	0,3185	0,3637	0,3931	0,413	0,4267	0,4362	0,4426

Compare the BEB and our algorithm. From (9) and applying optimal algorithm parameters obtained in [10] we see that for BEB $R_1(100) \approx 0.36$ and there is a hypothesis, mentioned in the introduction that $R_1 = 0$ for an infinite number of stations. Our algorithm always provides high tenacity $R_2 \approx 0.45$ (from (10)) no matter how many stations are active in the system. This guarantees the stable system operation in case there are thousands of subscribers in a cell, which is supposed to be the case in new generation wireless networks.

Finally, we would like to note that all analytical results have been validated by means of simulations in MatLab and the introduced approaches have shown high accuracy. We do not include simulation results here because of the space limit constraints.

5 Conclusion

In this paper a new simple and efficient random multiple access algorithm for next generation, centralized wireless networks, where number of stations may exceed several thousands, has been presented. In comparison to the binary exponential backoff, which is potentially unstable for a very large number of subscribers, our algorithm provides high tenacity, which is close to the fastest known model for an infinite number of stations. The cost for this advantage is a more complicated implementation in comparison to the BEB. For further investigations, we suppose to extend our analysis considering an error-prone wireless channel.

Acknowledgment

This work was partly funded for Alexey Vinel by the German Academic Exchange Service (DAAD) jointly with the Russian Federal Agency for Education in the framework of “Mikhail Lomonosov 2006/2007” program.

References

1. IEEE Std. 802.16-2004, IEEE standard for local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems.
2. D. Bertsekas, and R.G. Gallager, *Data Networks*, Prentice Hall, 1991.
3. J.I. Capetanakis “Generalized TDMA. The Multi-Accessing Tree Protocol Channels,” *IEEE Trans. Commun.* 1979. V. 27. №10. P.1476- 1483.
4. B.S. Tsybakov *Random Multiple Access*, USSR Academy of Science (preprint), Moscow, 1984 (in Russian).
5. B.S. Tsybakov, and V.A. Mikhailov “Free synchronous packet access in a broadcast channel with feedback,” *Probl. Information Transmission* 14 (1978) 259-280.
6. B.S. Tsybakov, and V.A. Mikhailov “Random multiple access of packets: Part and try algorithm,” *Probl. Information Transmission* 16 (1980) 65-79.
7. D. J. Aldous, "Ultimate Instability of Exponential Back-Off Protocol for Acknowledgement-Based Transmission Control of Random Access Communication Channels," *IEEE Transactions on Information Theory* 33(2), pp. 219-223 (March 1987).
8. B.S. Chlebus, *Randomized Communication in Radio Networks,* Handbook on Randomized Computing, Kluwer Academic Publishers, P.M. Pardalos, S. Rajasekaran, J. Reif and J.D.P. Rolim, Eds., 2000.

9. D.V. Cortizo, J. Garcia J., C. Blondia, B. VanHoudt, "FIFO by Sets ALOHA (FS-ALOHA): A Collision Resolution Algorithm for the Contention Channel in Wireless ATM Systems," IFIP WG 7.3 Performance, Istanbul, 1999.
10. A. Vinel, Y. Zhang, M. Lott and A. Tiurlikov, "Performance Analysis of the Random Access in IEEE 802.16," in *Proceedings of the 16th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'05)*, Berlin, Germany, 2005.
11. G. Bianchi "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, V18, N3, 2000. Pp 535-548.
12. V.M. Vishnevsky, A.I. Lyakhov "802.11 LANs: Saturation Throughput in the Presence of Noise," in *Proc. of 2nd Int. IFIP TC6 Networking Conf.* Pisa, Italy, May 19-24, 2002/ Lecture Notes in Computer Science, vol.2345, Springer-Verlag, 2002. Pp. 1008-1019.