A Lower Bound on Mean Delay for Free Access Class of RMA Algorithms

Andrey M. Turlikov,
Department of Information Systems
State University of Aerospace Instrumentation
St.-Petersburg, 190000, Russia
turlikov@vu.spb.ru

Sergey D. Andreev,
Department of Information Systems
State University of Aerospace Instrumentation
St.-Petersburg, 190000, Russia
Serge.Andreev@gmail.com

***Abstract*— The performance of a packet-switched network with the Poisson input source of rate λ and the infinite user population is studied. The users transmit their data packets over the communications channel with the *success-empty-collision* feedback information. We restrict our exploration to the class of the random multiple access algorithms for which a newly arrived packet is transmitted by a user as soon as possible. A function of λ is established such as the mean packet delay for any algorithm that belongs to the considered class is not below the value given by this function.**

I. INTRODUCTION

In many modern data networks the Random Multiple Access (RMA) is used to arbitrate access of the large user population to a shared broadcast channel. IEEE 802.16-based systems allow of RMA when transmitting the bandwidth requests prior to the scheduled data transmission. In GSM networks RMA is used similarly in the GPRS mode. In the networks which are built over IEEE 802.3 and IEEE 802.11 telecommunication standards RMA directly arbitrates the transmission of the data packets.

For the moderate values of arrival rate (of data packets into the system) RMA allows of reaching an unrivaled efficiency in comparison to the other known channel access techniques. One of the most essential performance characteristics is the expected value for the time period from the instant a data packet (request) arrived into the system to the instant it is successfully transmitted. This value represents the mean packet delay in the RMA system.

The implementers in the field of telecommunications face the challenging task of making the mean delay value as small as possible. In particular, IEEE 802.16 standard leaves some parameters out of its scope. By changing these parameters one can influence the value of the mean request delay. Clearly, the lower bound on the mean delay is of both theoretical and practical importance. Thus, the actual mean delay value cannot be lower than this fundamental threshold. Initially, the mean delay value for the set of all the RMA algorithms was found by Tsybakov and Likhanov in [1]. Later these authors together with Hajek analyzed the mean delay for the communications channel with a large propagation delay [2].

This paper considers a certain class of the RMA algorithms. For any algorithm that belongs to this class a newly arrived packet is transmitted by a user as soon as possible. The term *free access* is used in this paper to indicate the considered class. The first known upper bound on the capacity of the free access class of the RMA algorithms was given by Kelly in 1985 [3] and since then it has not been improved. In this paper we follow Kelly's approach to the capacity estimation to obtain the lower bound on the mean delay.

Note that the vast majority of the modern telecommunication systems that use RMA implement (as the means of collision resolution) the binary exponential backoff algorithm [4], [5], [6] which belongs to the considered free access class. The tree algorithms with free access [7], [8] also belong to this class.

The rest of the paper is structured as follows. Section II briefly summarizes the RMA model from [9], [10], [11], [12] and introduces the fundamental definitions. Section III describes the Hypothetic RMA algorithm. Section IV studies the operation of the Hypothetic algorithm and establishes its mean delay. Section V shows that for any algorithm that belongs to the free access class the mean delay is not less than that for the Hypothetic algorithm. *Conclusion* summarizes the paper.

II. SYSTEM MODEL

This paper considers the RMA model from [9], [10], [11], [12]. The user population in the system is infinite and each user has access to both input and output of the communications channel. Users generate data packets and exchange them via the channel. It is assumed that the packets have equal lengths. The transmission time of one packet is taken as a unit of the system time. Packet inter-arrival times are assumed to be statistically independent random variables which are distributed exponentially with the mean value of $\frac{1}{\lambda}$. Thus, λ is the arrival rate of the new packets into the system. It equals the mean number of packets that arrive into the system during a unit of time.

Below we formulate additional assumptions [9], [10], [11], [12] about the communications channel and the way users access it:

Assumption 1. The system time is slotted into equal slots. The duration of each slot is a unit of the system time which is exactly the transmission time of one data packet. Each

978-1-4244-2036-0/08/$25.00 ©2008 IEEE.

slot is assigned an integer nonnegative number and number t slot corresponds to the time interval of $[t, t+1)$. Hereinafter we refer to number t slot simply as slot t for the sake of brevity. Slot borders are known to all the users and each user is restricted to start its packet transmission only in the beginning of the current slot.

Assumption 2. In each slot any and only one of the following events may occur:

- only one user transmits (**S** event - success);
- none of the users transmit (**E** event - empty);
- two or more users transmit (**C** event - collision).

Assumption 3. When monitoring the channel activity a user is notified of the channel event by the end of the current slot.

Assumption 4. When monitoring the channel activity a user differentiates between the three possible channel events infallibly.

Assumption 5. Each user is supplied with a buffer sufficient to store only one packet. The packet is stored from the instant of time it arrived into the system to the instant of time it is successfully transmitted. When a packet arrives during interval $[t, t+1)$ it is transmitted not earlier than in slot $t' \geq t+1$.

Following the approach of [13] we introduce the definitions of the RMA algorithm, its delay and rate together with the capacity of the RMA system. Define the RMA algorithm as a rule which a user follows to decide whether to transmit its pending data packet in each time slot. Consider a packet arrival during interval $[t, t+1)$, that is, in slot t and its successful transmission in slot t_{Tr}. Then the delay of this packet is a random variable given by

$$D = t_{Tr} - t.$$

Consider the operation of algorithm f. Enumerate (with integer and nonnegative numbers) all the packets that arrive into the system from the very start of its operation. Denote the delay of packet number n by D_n. The mean delay for algorithm f is thus

$$d_f = \limsup_{n\to\infty} E[D_n].$$

The rate of algorithm f is the upper bound on the arrival rate which still provides finite mean delay

$$R_f = \sup_{\lambda}\{\lambda : d_f < \infty\}.$$

Consider set F of algorithms. The capacity of the RMA system for this set is given by

$$C_F = \sup_{f\in F} R_f.$$

Let us now specify a subset of all the RMA algorithms and denote it by F_{free}. A packet that arrives in randomly chosen slot t is transmitted in slot $t+1$ independently of all the other events in the system for any algorithm that belongs to F_{free}. Specified subset F_{free} will be referred to as the free access class of the RMA algorithms. In [3] it was shown that

$$C_{F_{free}} < \gamma,$$

where γ is the solution of equation $x = e^{-x}$ and $\gamma \approx 0.567$.

III. Hypothetic Algorithm

Denote the number of *backlogged* packets by q_t. Backlogged packets are the ones that participated in the collision and have not yet been transmitted successfully by the end of slot t. Further denote the number of the packets that arrived into the system during slot t by v_t. Assume the presence of the coordinating center which monitors the channel activity together with all the users. Additionally, the coordinating center is notified of the exact number of the backlogged packets in the system as well as of the IDs of q_t backlogged users by the end of each slot t. Consider the following algorithm. In the end of slot t the coordinating center randomly selects one of q_t backlogged users and the selected user transmits its packet in slot $t+1$. Clearly, in the model of Section II this algorithm is impossible to implement as there is no coordinating center. Therefore, hereinafter we regard this algorithm as Hypothetic.

The dynamics of the number of the backlogged packets in the system can be described by the following recurrent formula

$$q_{t+1} = q_t + v_t - delta(q_t, v_t), \tag{1}$$

where $delta(q_t, v_t)$ equals 1 if $(q_t = 0, v_t = 1)$ or $(q_t > 0, v_t = 0)$ and equals 0 otherwise.

From (1) it follows that the sequence of q_t is given by the homogeneous irreducible and aperiodic Markov chain with the countable state space.

Lemma 1. Denote the root of equation $x = e^{-x}$ by γ.

- For $\lambda < \gamma$ the considered Markov chain is ergodic.
- For $\lambda \leq \gamma$ the considered Markov chain is recurrent.
- For $\lambda > \gamma$ the considered Markov chain is null-recurrent.

The above lemma may be proved taking into account the results in [14]. It immediately follows from it that for any $\lambda < \gamma$ the mean delay for the Hypothetic algorithm is finite, otherwise it increases without bound. Further we formulate and prove an important proposition for the Hypothetic algorithm which is used in the subsequent derivations.

Proposition 1. For any value of arrival rate λ the conditional expected value of the number of the backlogged packets in the system $E[q_{t+1}|q_t = n]$ is given by

$$E[q_{t+1}|q_t = 0] = \lambda - \lambda e^{-\lambda},$$

for $n = 0$ and by

$$E[q_{t+1}|q_t = n] = n + \lambda - e^{-\lambda},$$

for $n > 0$.

Proof. Firstly we calculate the conditional expected value for the number of the backlogged packets in slot $t+1$ conditioning on th fact that there were n backlogged packets in slot t

$$\begin{aligned} &E[q_{t+1} \mid q_t = n] = E[n + v_t - delta(q_t, v_t) \mid q_t = n] = \\ &= n + +E[v_t] - E[delta(q_t, v_t)|q_t = n] = \\ &= n + \lambda - E[delta(q_t, v_t)|q_t = n]. \end{aligned}$$

From the equation above it immediately follows that for $n = 0$

$$\begin{aligned} &E[q_{t+1} \mid q_t = 0] = \lambda - E[delta(q_t, v_t)|q_t = 0] = \\ &= n + \lambda - Pr\{v_t = 1\} = \lambda - \lambda e^{-\lambda}, \end{aligned}$$

and for $n > 0$

$$E[q_{t+1} \mid q_t = n] = n + \lambda - E[delta(q_t, v_t)|q_t = n] = \\ = n + \lambda - Pr\{v_t = 0\} = n + \lambda - e^{-\lambda},$$

which completes the proof. ■

IV. Mean Delay for Hypothetic Algorithm

Consider the operation of the Hypothetic algorithm described in the previous section. After we find the mean number of the backlogged packets in the system the mean delay is calculated using the Little's formula. In order to derive the expected value for the number of the backlogged packets we follow the approach from [15]. Assume that $\lambda < \gamma$, that is, the underlying Markov chain is ergodic. Hence, the stationary distribution for the number of the backlogged packets exists. Denote random variable q_t for $t \to \infty$ by q. Note that as the chain is ergodic both the first and the second moments of q_t exist for $t \to \infty$ and they are

$$E[\lim_{t\to\infty} q_t] = \lim_{t\to\infty} E[q_t] = E[q], \\ E[\lim_{t\to\infty} q_t^2] = \lim_{t\to\infty} E[q_t^2] = E[q^2].$$

Coming to the mathematical expectation in both parts of (1) we establish that

$$E[q_{t+1}] = E[q_t] + E[v_t] - E[delta(q_t, v_t)].$$

Note that random variables v_t are independent and identically distributed (i.i.d.) for all the values of t. Therefore, we can further omit index t and denote the considered random variable simply by v. Taking the above equation to the limit for $t \to \infty$ and accounting for $\lim_{t\to\infty} E[q_{t+1}] = \lim_{t\to\infty} E[q_t]$, we readily obtain the following

$$E[q] = E[q] + E[v] - E[delta(q, v)], \\ E[v] = E[delta(q, v)]. \tag{2}$$

Now, accounting for the statistical independence of random variables q and v, we derive the expression for the expected value of random variable $delta(q, v)$ as follows

$$E[delta(q, v)] = Pr\{delta(q, v) = 1\} = Pr\{v = 1 \wedge \\ \wedge q = 0\} + Pr\{v = 0 \wedge q > 0\} = Pr\{v = 1\}Pr\{q = 0\} + \\ + Pr\{v = 0\}Pr\{q > 0\} = e^{-\lambda}(\lambda Pr\{q = 0\} + Pr\{q > 0\}).$$

Combining the above result together with (2) and $E[v] = \lambda$ we calculate the probability of at least one backlogged packet being present in the system $Pr\{q > 0\}$, which is

$$\lambda = \lambda e^{-\lambda} Pr\{q = 0\} + e^{-\lambda} Pr\{q > 0\}, \\ Pr\{q > 0\} = \frac{\lambda - \lambda e^{-\lambda}}{e^{-\lambda} - \lambda}. \tag{3}$$

Putting both parts of (1) in the power of two and coming to the expected values for $t \to \infty$ we obtain that

$$E[q^2] = E[q^2] + E[v^2] + E[delta(q, v)^2] - \\ -2E[q \cdot delta(q, v)] + 2E[q \cdot v] - 2E[v \cdot delta(q, v)].$$

Contracting both parts of the above equation with $E[q^2]$ we establish that

$$E[v^2] + E[delta(q, v)^2] - 2E[q \cdot delta(q, v)] + \\ +2E[q \cdot v] - 2E[v \cdot delta(q, v)] = 0. \tag{4}$$

Further we calculate the terms of (4) individually.

Note that random variable v follows a Poisson distribution with the mean value λ, which implies $E[v^2] = \lambda + \lambda^2$.

Using (2) and taking into account that $delta(q, v) = delta(q, v)^2$ we get that

$$E[delta(q, v)^2] = E[delta(q, v)] = E[v] = \lambda.$$

Remember that $delta(q, v)$ equals 1 if $(q_t = 0, v_t = 1)$ or $(q_t > 0, v_t = 0)$ and equals 0 otherwise. Again, accounting for the statistical independence of random variables q and v, it follows that

$$E[q \cdot delta(q, v)] = E[q]Pr\{v = 0 \wedge q > 0\} = \\ = E[q]Pr\{v = 0\} = E[q]e^{-\lambda}, \\ E[q \cdot v] = E[q]E[v] = E[q]\lambda, \\ E[v \cdot delta(q, v)] = Pr\{v = 1 \wedge q = 0\} = \\ = Pr\{v = 1\}Pr\{q = 0\} = Pr\{q = 0\}\lambda e^{-\lambda}.$$

Putting the above equations in (4) gives us

$$\lambda + \lambda^2 + \lambda - 2E[q]e^{-\lambda} + \\ +2E[q]\lambda - 2Pr\{q = 0\}\lambda e^{-\lambda} = 0, \\ E[q] = \frac{\lambda^2 + 2\lambda - 2Pr\{q = 0\}\lambda e^{-\lambda}}{2(e^{-\lambda} - \lambda)}.$$

Now substituting for (3) in the above equation and after elementary transformations we establish that

$$E[q] = \frac{\lambda(2 + \lambda + \lambda^2 - 2e^{-\lambda})}{2(1 - \lambda)(e^{-\lambda} - \lambda)}.$$

Combining the above equation together with the Little's formula we finally derive the mean delay $d_h(\lambda)$ for the Hypothetic algorithm as follows

$$d_h(\lambda) = \frac{E[q]}{\lambda} = \frac{2 + \lambda + \lambda^2 - 2e^{-\lambda}}{2(1 - \lambda)(e^{-\lambda} - \lambda)}. \tag{5}$$

V. Lower Bound on Mean Delay for Free Access Class of Algorithms

Consider algorithm $f \in F_{free}$. Denote the total number of the backlogged packets in slot t by y_t and the number of the transmitted backlogged packets in the same slot by s_t. The dynamics of the number of the backlogged packets in the system is presented by the following recurrent formula

$$y_{t+1} = y_t + v_t - \varphi(s_t, v_t), \tag{6}$$

where $\varphi(s_t, v_t)$ equals 1 if $(s_t = 0, v_t = 1)$ or $(s_t = 1, v_t = 0)$ and equals 0 otherwise. Note that by contrast to the Hypothetic algorithm the considered sequence of the random variables is not given by the Markov chain. Since the number of the backlogged packets transmitted in slot t generally depends not only on y_t but also on all the events that have occurred in the system from the start of its operation to time slot t.

Proposition 2. For any value of arrival rate λ the conditional expected value of the number of the backlogged packets in the system $E[q_{t+1}|q_t = n]$ is estimated by

$$E[y_{t+1}|y_t = 0] = \lambda - \lambda e^{-\lambda},$$

for $n = 0$ and by

$$E[y_{t+1}|y_t = n] \geq n + \lambda - e^{-\lambda},$$

for $n > 0$.

Proof. Firstly we consider the case of $n = 0$. For $y_t = 0$ it always holds that $s_t = 0$, that is, the number of backlogged packets in slot $t+1$ is only determined by the packet arrivals in slot t. Thus, from (6) we have

$$E[y_{t+1}|y_t = 0] = E[v_t] - E[\varphi(s_t, v_t)|y_t = 0] = \\ = E[v_t] - Pr\{v_t = 1\} = \lambda - \lambda e^{-\lambda}.$$

For $y_t = n > 0$ the number of backlogged packets s_t, which are transmitted in slot t is generally a random variable. The distribution of this random variable s_t depends on all the events that occurred in the system from the start of its operation to slot t. Note that this distribution does not depend on the number of arrivals in slot t. Summarizing the above and accounting for (6), we establish that

$$E[y_{t+1}|y_t = n] = n + E[v_t] - Pr\{v_t = 1 \wedge s_t = 0|y_t = n\} - \\ -Pr\{v_t = 0 \wedge s_t = 1|y_t = n\} = n + \lambda - \\ -\lambda e^{-\lambda} Pr\{s_t = 0|y_t = n\} - e^{-\lambda} Pr\{s_t = 1|y_t = n\}.$$

Accounting for the fact that $Pr\{s_t = 0|y_t = n\} + Pr\{s_t = 1|y_t = n\} \leq 1$ and using the above we obtain the lower bound on $E[y_{t+1}|y_t = n]$ as

$$E[y_{t+1}|y_t = n] = n + \lambda - \lambda e^{-\lambda} Pr\{s_t = 0|y_t = n\} - \\ -e^{-\lambda} Pr\{s_t = 1|y_t = n\} \geq n + \lambda - \\ - \max_{\alpha \geq 0, \beta \geq 0 : \alpha + \beta \leq 1} (\lambda e^{-\lambda} \alpha + e^{-\lambda} \beta) = n + \lambda - e^{-\lambda},$$

which completes the proof. ■

Proposition 3. Suppose in initial time instant $t = 0$ the system is empty. Then for any value of arrival rate λ and for any slot t for the conditional expected values of the number of the backlogged packets (for the Hypothetic $E[q_t|q_0 = n]$ and any algorithm $f \in F_{free}$ $E[y_t|y_0 = n]$, respectively) it holds the following

$$E[q_t|q_0 = n] \leq E[y_t|y_0 = 0].$$

The proof of Proposition 3 is based on the arguments similar to that in Proposition 1 and Proposition 2, however, it is excluded due to the space limitations. We again consider the case when the system is empty in some initial instant of time. Combining the above inequality together with the Little's formula we formulate the following theorem.

Theorem 1. Denote the mean delay (for any value of arrival rate λ) for algorithm f and for the Hypothetic algorithm by $d_f(\lambda)$ and $d_h(\lambda)$, respectively. Let $\lambda < \gamma$ then it holds that $d_h(\lambda) \leq d_f(\lambda)$.

From the above theorem it follows that the mean delay for the Hypothetic algorithm (5) represents the lower bound on the mean delay for any algorithm that belongs to the free access class of the RMA algorithms.

VI. Conclusion

We introduced the lower bound on the mean delay for the free access class of the RMA algorithms. By contrast to the previously know bound for all the RMA algorithms [1] our bound is given by a simple analytical expression

$$d_h(\lambda) = \frac{2 + \lambda + \lambda^2 - 2e^{-\lambda}}{2(1-\lambda)(e^{-\lambda} - \lambda)}.$$

The introduced bound allows a quantitative performance estimation of the practical algorithms. For instance, in [16] the mean delay of the non-blocking stack algorithm for low arrival rates is established as

$$d_{stack}(\lambda) = 3\lambda + 11.33...\lambda^2 + o(\lambda^2).$$

By representing $d_h(\lambda)$ analogously we obtain the following expression

$$d_h(\lambda) = \frac{3}{2}\lambda + \frac{9}{2}\lambda^2 + o(\lambda^2).$$

By comparing the above expressions ($d_{stack}(\lambda) \approx 2d_h(\lambda)$) we conclude that for low arrival rates the non-blocking stack algorithm is 'almost optimal' in the free access class of the RMA algorithms.

References

[1] B. S. Tsybakov and N. B. Likhanov, "Lower bound on delay in a random multiple access system," *Probl. of Inf. Trans.*, vol. 27, no. 3, pp. 247–260, 1991.

[2] B. Hajek, N. B. Likhanov, and B. S. Tsybakov, "On the delay in a multiple-access system with large propagation delay," *IEEE Trans. on Inf. Theory*, vol. 40, no. 4, pp. 1158–1166, 1994.

[3] F. P. Kelly, "Stochastic models of computer communication systems," *J. of the Royal Stat. Society (Series B)*, vol. 47, pp. 379–395, 1985.

[4] R. M. Metcalfe and D. R. Boggs, "Ethernet: Distributed packet switching for local computer networks," *ACM Communications*, vol. 19, no. 5, pp. 395–404, 1976.

[5] N. O. Song, B. J. Kwak, and L. E. Miller, "On the stability of exponential backoff," *J. Research of NIST*, vol. 108, pp. 289–297, 2003.

[6] N. D. Vvedenskaya and Y. M. Suhov, "Multi-access system with many users: Stability and metastability," *Probl. of Inf. Trans.*, vol. 43, no. 3, pp. 263–269, 2007.

[7] G. Fayolle, P. Flajolet, M. Hofri, and P. Jacquet, "Analysis of a stack algorithm for random multiple-access communication," *IEEE Trans. on Inf. Theory*, vol. 31, no. 2, pp. 244–254, 1985.

[8] G. T. Peeters, B. V. Houdt, and C. Blondia, "A multiaccess tree algorithm with free access, interference cancellation and single signal memory requirements," *Perform. Eval.*, vol. 64, pp. 1041–1052, 2007.

[9] B. S. Tsybakov and V. A. Mikhailov, "Free synchronous packet access in a broadcast channel with feedback," *Probl. of Inf. Trans.*, vol. 14, no. 4, pp. 259–280, 1978.

[10] D. Bertsekas and R. Gallager, *Data Networks*, S. Edition, Ed. Prentice Hall, 1992.

[11] L. Gyorfi, S. Gyori, and J. L. Massey, "Principles of stability analysis for random accessing with feedback," *Multiple Access Channels: Theory and Practice*, vol. IOS Press, pp. 214–249, 2007.

[12] G. S. Evseev and A. M. Turlikov, "A connection between characteristics of blocked stack algorithms for random multiple access system," *Probl. of Inf. Trans.*, vol. 43, no. 4, pp. 83–92, 2007.

[13] B. S. Tsybakov, "Survey of ussr contributions to random multiple-access communications," *IEEE Trans. on Inf. Theory*, vol. 31, no. 2, pp. 143–165, 1985.

[14] F. G. Foster, "On the stochastic matrices associated with certain queueing processes," *Annals of Math. Stat.*, vol. 24, pp. 355–360, 1953.

[15] D. G. Kendall, "Some problems in the theory of queues," *Journal of the Roal Statistical Society (Series B)*, vol. 13, pp. 151–185, 1951.

[16] N. D. Vvedenskaya and B. S. Tsybakov, "Computing packet delay for some random multiple access stack algorithms," *Probl. of Inf. Trans.*, vol. 24, no. 3, pp. 246–253, 1988.