Capacity Estimation of Centralized Reservation-Based Random Multiple-Access System

Andrey M. Turlikov and Alexey V. Vinel
Department of Information Systems
State University of Aerospace Instrumentation
Saint-Petersburg, Russia

Abstract

Modification of the basic random multiple-access model for the case of a centralized reservation-based data network is considered. Substance of the modification is the following. The channel time is composed of equal intervals, which are called frames. Each frame consists of some consequent mini-slots for the transmission of requests, used for the channel reservation, and consequent slots for the actual packets transmission. Upper and lower bounds of Tsybakov's capacity are estimated for the system. Problem of the optimal choice for the numbers of mini-slots and slots per frame is analyzed. It is shown that these values do not depend on the ratio between the duration of request and packet transmission.[1]

I. Introduction

Recent successes in wireless communications lead to constantly increasing interest in random multiple-access (RMA) theory nowadays. Since 70s RMA is widely known as an efficient method providing communication between large number of subscribers with bursty traffic sources in packet-switched data networks. Idea of reservation in multiple-access systems is the old one as well. In [1] Rubin, inspired mostly by the operation of satellite systems, considers centralized reservation-based multiple-access scheme. In the model from [1] the synchronized subscribers perform reservations, by transferring short requests to central repeater, and then transmit multiple-packets messages. Therefore, the shared broadcast channel is divided into so-called frames. Each frame consists of consequent mini-slots for reservation and slots for actual packet transmission. Access to the slots is regulated by time division technique, each mini-slot can be either assigned to a single subscriber or be potentially used by all subscribers in contention manner. It is very surprising, that some aspects of the operation of media-access control (MAC) sublayer of contemporary IEEE 802.16 broadband wireless networks in point-to-multipoint mode [2] can be modeled by Rubin's model.

In [1] time-probabilistic characteristics are computed for different scenarios, particularly considering large propagation delay values, and with the emphasis on reservations performed by means of time-division multiple-access (TDMA). The most commonly used model for the RMA systems analysis was described for instance in [3] by Tsybakov. Later its assumptions were expounded by Gallager in [4]. Throughout the rest of the paper we will refer this model as *basic*. In contrast to Rubin's model, where a finite number of subscribers is assumed, the basic model assumes an infinite number of subscribers. Under this assumption the TDMA system is principally incapable of providing finite mean packet delay, while many RMA algorithms are capable of doing it.

Tsybakov and Berkovskii [5] consider the reservation problem in the framework of the basic model. In contrast to [1], in [5] requests are not considered and the subscriber indicates how long it will require the channel in regular packets. Packets from various subscribers compete with each other according to some RMA algorithm. If a packet from some subscriber is received successfully, then all other subscribers in the system stop their transmissions during the specified time interval, thus enabling the subscriber sending the packet to transmit its information without conflict.

In this paper we introduce another reservation-based multiple-access model, which is a combination of the models from [1] and [3]. On the one hand, we analyze the centralized network, where frames are used for the packet transmission, under assumptions similar to those made by Rubin. On the other hand

[1]The authors would like to thank their colleagues from the Information Systems Department of Saint-Petersburg State University of Aerospace Instrumentation, particularly Dr. G. S. Evseev, for the discussions, which have led to the formulation of the paper main results. E-mail to vinel@ieee.org.

assuming infinite number of subscribers, we focus on the RMA algorithms and perform the estimation of Tsybakov's capacity for the considered system. Application of the results can be illustrated by means of the IEEE 802.16 MAC analysis. The usage of infinite subscribers number model is motivated by the fact, that the number of subscribers in IEEE 802.16 data network is expected to be fairly large. The rest of the paper is organized as follows. In Section II basic RMA model is explained and some auxiliary propositions are proved. Our *centralized reservation-based model* as well as the problem statement are described in Sections III and IV, respectively. Upper and lower bounds for the capacity are constructed in Section V. Some practical remarks are included in Section VI.

II. Basic Random Multiple-Access System Model

Here we briefly explain basic RMA system model and review some necessary definitions from [3]. Packets transmission over the multiple-access channel is investigated. Infinite subscribers model, where each subscriber can have at most one packet requiring transmission is considered. The subscribers are assumed to transmit packets of a fixed length whose duration is taken as a time unit. The system is slotted, so that subscribers can begin packet transmissions only at times $t \in \{0, 1, 2, \ldots\}$. The time interval $(t, t+1)$ will be called a *slot*. The channel is noiseless and it is assumed that each subscriber knows by time $t+1$ which of the three possible events *idle slot*, *successful transmission*, or *conflict* (two or more simultaneous transmissions) did in fact occur in the slot $(t, t+1)$. The packet generation times of all subscribers form the overall input traffic, which is assumed to be discrete Poisson. The probability that j new packets are generated at some moment t equals to $e^{-\lambda}\lambda^j/j!$, where λ is the intensity of the overall input traffic.

An *RMA algorithm for the basic system* is defined as a rule that enables any subscriber with a ready-for-transmission packet at any time $t \in \{0, 1, 2, \ldots\}$, to determine whether or not it should transmit this packet in the next slot $(t, t+1)$. Thus we have a function of three arguments. The first argument is the time x of packet generation. The second argument is the sequence $\theta(t) = (\theta_1, \ldots, \theta_t)$ of channel events θ_i, here $\theta_i = 0$ if $(i-1, i)$ was an idle slot, $\theta_i = 1$ if only one subscriber transmitted in this slot, and $\theta_i = 2$ if two or more subscribers transmitted in this slot. The third argument is the sequence $\nu(x, t) = (\nu_1(x), \ldots, \nu_t(x))$ of events at the subscriber where a packet was generated at time x. Here $\nu_i(x) = 0$ if this subscriber has not transmitted a packet in the slot $(i-1, i)$, and $\nu_i(x) = 1$ if it has. Therefore, formally an RMA algorithm is defined as a function $f_0[x, \theta(t), \nu(x, t)]$ with values in the interval $[0, 1]$. Its value is the probability that a packet generated at time x will be transmitted in the slot $(t, t+1)$.

The delay of a packet is the time interval from the moment of its generation till the moment of its successful transmission. The delay $\delta^{(0)}(\lambda, f_0)$ is a random variable. Let a packet be generated at an arbitrary but fixed time t at some subscriber, and let $\delta_t^{(0)}(\lambda, f_0)$, be its delay. The *mean delay* (referred to as *virtual mean delay* in [3]) is defined as $D_0(\lambda, f_0) \triangleq \overline{\lim}_{t\to\infty} E\delta_t^{(0)}(\lambda, f_0)$.

The *transmission rate (tenacity)* of RMA algorithm f_0 is the maximum (more precisely, the supremum) intensity of the input traffic that can be transmitted by the algorithm with finite delay: $R_0(f_0) \triangleq \sup_\lambda\{\lambda\colon D_0(\lambda, f_0) < \infty\}$.

Finally, the *capacity* of the basic RMA system is defined as $C_0 \triangleq \sup_{f_0\in\mathcal{F}_0} R_0(f_0)$, where $\mathcal{F}_0$ is a set of all RMA algorithms (note, that capacities can be defined *over the class* in the sense, that any other class different from $\mathcal{F}_0$ can be used in the above definition). Exact value of the capacity C_0 is still unknown. As it was mentioned in [6] some researchers conjectured that the optimal value might be 0.5, but this claim was quickly abandoned as baseless. The best known upper bound for the capacity C_0 was found by Likhanov and Tsybakov in [7] and is shown to be $\overline{C}_0 = 0.587$. The fastest known algorithm is a part-and-try one, which rate is $R_{pt} = 0.487$, was found by Tsybakov and Mikhailov in [8]. Later it was slightly improved, but the core of the algorithm remained the same.

Before moving to the central problem of the paper, we will prove several auxiliary propositions for RMA systems having some form of *feedback delay*. Similar problem was addressed by Hajek in [9]. In that paper the feedback information θ_i is assumed to be announced to all the subscribers by time $i+N$, where N is the feedback delay. In the basic model the event in the slot i is known by the beginning of slot $i+1$, meaning that $N = 1$. In this section we assume, that all slots are grouped into equal consequent segments of length K. The values of function f_0 do not depend on the values of θ_i related to the current segment. For given K value any RMA algorithm and the set of all the RMA algorithms justifying this rule are denoted as $f_0^{(K)}$ and $\mathcal{F}_0^{(K)}$ respectively. Note, that $\mathcal{F}_0^{(1)} \triangleq \mathcal{F}_0$

Proposition 1: $C_0^{(K)} = \sup_{f_0^{(K)} \in \mathcal{F}_0^{(K)}} R_0(f_0) \leq C_0$.

Proof: From the definition of class $\mathcal{F}_0^{(K)}$, it follows directly, that for any K: $\mathcal{F}_0^{(K)} \subset \mathcal{F}_0$ and thus proposition holds. ■

Proposition 2: For any algorithm $f_0 \in \mathcal{F}_0$, having transmission rate R_0, and any value of K an algorithm $f^{(K)} \in \mathcal{F}_0^{(K)}$ exists, which also has the transmission rate R_0.

Proof: Let us show how to construct the desired algorithm. Any algorithm $f_0 \in \mathcal{F}_0$ can be modified in the following way to be in the set $\mathcal{F}_0^{(K)}$. At the moment of a packet generation a subscriber chooses a number m uniformly from $\{1, 2, \ldots, K\}$ once and then "applies" algorithm f_0 only to slots having number m in any segment of K slots. This means, that each subscriber uses feedback from one fixed slot (which has number m in each segment) and can transmit only in such slots. Thus, we "split" our system into K independent basic systems, where each subscriber randomly chooses one system for his operation once and then works independently of those who have chosen different ones according to algorithm having transmission rate R_0/K. Thus, overall transmission rate achieved is R_0. ■

Note, that this approach does not necessarily guarantee, that the mean delay of the constructed algorithm will be "good". Moreover, it's easy to give examples when this "splitting" approach leads to unwarrantably high delay values [9].

Proposition 3: For any given K the capacity $C_0^{(K)}$ achieved over the class $\mathcal{F}_0^{(K)}$ equals to the capacity of the basic system C_0 (achieved over the class $\mathcal{F}_0$).

Proof: On the one hand from proposition 1 it follows, that $C_0^{(K)} \leq C_0$. On the other hand, from proposition 2 it follows, that any algorithm from $\mathcal{F}_0$ for any K can be modified in the way that it can be in $\mathcal{F}_0^{(K)}$, without reducing its transmission rate. Thus, $C_0^{(K)} = C_0$. ■

III. Centralized Reservation-Based System Model

Let us consider a data transmission system with one *central station* and infinite number of *subscribers*. Central station is connected to all the subscribers by means of two communication channels, namely uplink and downlink. *Uplink channel* is used for the data transmission from all subscribers to the central station and the *downlink channel* is used for the information transmission from the base station to the subscribers.

The traffic model used is exactly the same as in the basic model - the moments of *packets* arrivals represent a Poisson process, which provides an arrival rate equal to λ packets per unit of time. However, each subscriber, having a new packet, transmits a special *request* message to the central station in order to reserve uplink channel time. The duration of the request transmission is supposed to be $\alpha < 1$ units of time. In all the consequent consideration we assume, that *durations of request and packet transmissions are fixed* and uplink channel usage is organized in the following way. The time axis is slotted into equal intervals of time, which are called *frames*. All frames have fixed structure. Each frames comprises $K \geq 1$ intervals of time having duration α, which are called *mini-slots*, and $L \geq 1$ intervals of time having duration of the unit of time, which are called *s*lots. Slots are used by the subscribers for the transmission of packets, while mini-slots are used for request transmission.

The system is synchronized. Central station and all subscribers know the beginning of each i-th frame $i(\alpha K + L)$, each j-th slot $j + \alpha K \lfloor (j+1)/L \rfloor$ and each k-th mini-slot $k\alpha + L\lfloor k/K \rfloor$, where $i, j, k \in \{0, 1, 2, \ldots\}$ and transparent numeration of slots and mini-slots is assumed.

Since simultaneous transmissions of subscribers are possible in the mini-slots, three different situations can be distinguished in an arbitrary mini-slot $l \in \{1, 2, \ldots, K\}$ of frame number $(i-1)$ (we denote them by $\theta_i^{(l)}$): *successful transmission* of some subscriber ($\theta_i^{(l)} = 1$), *empty* mini-slot meaning that there is not any transmission ($\theta_i^{(l)} = 0$), and *collision*, when two or more subscribers transmit in the mini-slot ($\theta_i^{(l)} = 2$). By the beginning of each i-th frame the central station in the downlink transmits information about the situation in all the mini-slots of the frame $i-1$ to all the subscribers. This information is represented by the *feedback vector* $\bar{\theta}_i = (\theta_i^{(1)}, \theta_i^{(2)}, \ldots, \theta_i^{(K)})$.

Subscribers transmit requests by means of some *centralized reservation-based system RMA algorithm* $f^{(K)}$, which is a rule, that bases on the situations in mini-slots of previous frames, and is used by the subscribers at the beginning of each frame to determine whether to transmit a request in a mini-slot of this frame or not. Analogously to the basic model $f^{(K)}$ is defined as a function of three arguments $f^{(K)}[x, \theta(n), \nu(x, n)]$, $n \in \{0, 1, 2, \ldots\}$. Here, x is the moment of time, when the packet is generated and $\theta(n) = (\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_n)$ is a sequence of feedback vectors till the beginning of frame n. Finally, $\nu(x, n) = (\bar{\nu}_1(x), \bar{\nu}_2(x), \ldots, \bar{\nu}_n(x))$ is a sequence of vectors for the subscriber x, $\bar{\nu}_i(x) = (\nu_i^{(1)}(x), \nu_i^{(2)}(x), \ldots, \nu_i^{(K)}(x))$. We denote $\nu_i^{(l)}(x) = 0$ if the subscriber whose packet has been generated at time x did not transmit a request in the l-th mini-slot of the i-th frame and $\nu_i^{(l)}(x) = 1$ otherwise. The possible values of the function f are set of vectors: $\bar{p} = (p^{(1)}, p^{(2)}, \ldots, p^{(K)})$, where each element $p^{(l)}$ represents the probability of the subscriber's transmission in the l-th mini-slot of the n-th frame.

We suppose, there is an infinite queue buffer for the requests at the central station. When some request is transmitted successfully, it arrives to the tail of this queue. The central station serves the requests from the conducted queue according to some rule, which is referred to as *service discipline* $g^{(L)}$. The serving is done by means of a downlink transfer information at the beginning of each frame, describing which subscribers are allowed to transmit their packets in each of the L slots of the next frame. We suppose, that the central station uses the number of mini-slots, where some request has been successfully transmitted to inform the corresponding subscriber about the assignment of slots to him. That is why, throughout this paper we assume that a subscriber can not make more than one request transmission attempt per frame. This leads to the following restriction for considered algorithms. For any $f^{(K)}$: the weight of vector $\bar{\nu}_i(x)$ is either one or zero for any subscriber x and frame i.

Noiseless uplink and downlink channels are assumed. Neither packets nor requests can be distorted by noise. Situations in mini-slots are always correctly distinguished by the central station. Feedback vectors and slots allocation information is always successfully transmitted in the downlink to all subscribers.

IV. Definitions and Problem Statement

We call the pair $(f^{(K)}, g^{(L)})$ the *multiple access protocol* for centralized reservation-based system with parameters (K, L). Here, we introduce definitions analogous to those given previously for the basic RMA model, with extensions corresponding to our system. The time interval from the moment when a packet was generated to the moment it has been successfully transmitted is referred to as packet delay transmission. Then in some arbitrary but fixed frame (having number n) let an additional packet arrive in the system, whose transmission delay is denoted by $\delta_n(\lambda, K, L, f^{(K)}, g^{(L)})$. According to the algorithm of the system operation the transmission delay consists of two components. The first one is the request delay for random access $\delta_n^{(1)}(\lambda, K, L, f^{(K)})$. It is the time from the moment of packet generation, to the moment of the corresponding request successful transmission. The second one is the time from the moment of request successful transmission, to the corresponding packet will be successfully transmitted $\delta_n^{(2)}(\lambda, K, L, g^{(L)})$. We will refer to this value as queuing delay. The value $D(\lambda, K, L, f^{(K)}, g^{(L)}) \triangleq \overline{\lim}_{n\to\infty} E\delta_n = \overline{\lim}_{n\to\infty} E(\delta_n^{(1)} + \delta_n^{(2)})$ for a given arrival rate λ, K mini-slots, L slots and multiple access protocol $(f^{(K)}, g^{(L)})$ will be referred to as the *mean delay of packet transmission*. Further we will need notation for mean request delay for the random access $D_1 \triangleq \overline{\lim}_{n\to\infty} E\delta_n^{(1)}$.

The maximal arrival rate value (more precisely the supremum of the arrival rate), which can be transmitted by means of some multiple access protocol $(f^{(K)}, g^{(L)})$ for some frame structure (K, L), with finite mean delay $R(K, L, f^{(K)}, g^{(L)}) \triangleq \sup_\lambda\{\lambda\colon D(\lambda, K, L, f^{(K)}, g^{(L)}) < \infty\}$ will be referred to as *transmission rate (tenacity)* of the multiple access protocol.

Then if the multiple access protocol is not fixed, the *system capacity* is the following value

$$C(K, L, \mathcal{F}^{(K)}, \mathcal{G}^{(L)}) \triangleq \sup_{\substack{f^{(K)} \in \mathcal{F}^{(K)} \\ g^{(L)} \in \mathcal{G}^{(L)}}} R(K, L, f^{(K)}, g^{(L)}),$$

where $\mathcal{F}^{(K)}$ is set of all RMA algorithms defined for the system with K mini-slots and $\mathcal{G}^{(L)}$ is set of all service disciplines, which can be defined for the system with L slots.

Our aim is to compute the upper and lower bounds for the capacity $C(K, L, \mathcal{F}^{(K)}, \mathcal{G}^{(L)})$.

V. Capacity Estimation

Let us first consider only the part of the whole system operation, namely request transmission during reservation. Packet transmission is not considered. This system is referred to as a *reduced* one. Then, transmission rate R_1 and capacity C_1 definitions analogous to those previously mentioned can be introduced for the reduced system, namely $R_1(K, L, f) \triangleq \sup_\lambda\{\lambda\colon D_1(\lambda) < \infty\}$ and $C_1(K, L, \mathcal{F}^{(K)}) \triangleq \sup_{f\in\mathcal{F}^{(K)}} R_1(K, L, f)$. Then following propositions are proved.

Proposition 4: If there is exactly one mini-slot in each frame then the capacity of the reduced system equals to $C_0/(\alpha + L)$, where C_0 is the capacity of the basic RMA system ($C_1(1, L, \mathcal{F}^{(1)}) = C_0/(\alpha + L)$).

Proof: It is easy to notice, that for $K = 1$, when each frame consists of only one mini-slot we have exactly the basic RMA system, for which vectors $\bar{\theta}_i$, $\bar{\nu}_i(x)$ and the output of function f turn to scalars. Thus, $\mathcal{F}^{(1)} = \mathcal{F}_0$. The only difference is that one "slot", which is used in basic system corresponds to one frame of length $(\alpha + L)$ in our reduced system, what is taken into account by means of corresponding normalization. ■

Proposition 5: If there are more than one mini-slot in each frame then the capacity of the reduced system equals to $(C_0K)/(\alpha K + L)$, where C_0 is the capacity of the basic RMA system ($C_1(K, L, \mathcal{F}^{(K)}) = C_0K/(\alpha K + L)$, $K \geq 2$).

Proof: It is easy to notice, that $\mathcal{F}_0^{(K)} = \mathcal{F}^{(K)}$. Thus, for $K \geq 2$ we have exactly the basic RMA system with slots grouped into segments of length K (as it is explained in section II), whose capacity is proved to be C_0 (proposition 3). The only difference is that one "slot", which is used in the basic system corresponds to one frame of length $(\alpha K + L)$ in our reduced system, what is taken into account by means of corresponding normalization. ■

Now we are finishing with the analysis of the reduced system and consider the overall reservation model. Below are two necessary conditions for the system stability.

Proposition 6: The mean request delay for the random access D_1 and the mean delay of packet transmission D may be finite if the inequality

$$\lambda(\alpha K + L) < C_0K \tag{1}$$

holds.

Proof: From proposition 5 it directly follows that the request delay for the random access D_1 is infinite if the arrival rate does not satisfy $\lambda < C_0K/(\alpha K + L)$. Obviously, the same is valid for the mean delay D. ■

Proposition 7: Let, the arrival rate value λ is chosen such as the request delay for the random access D_1 is finite, then mean delay of packet transmission in the system D may be finite if inequality

$$\lambda(\alpha K + L) < L \tag{2}$$

holds.

Proof: Generation and transmission of packets can be described in terms of queueing theory ([10]). We have Poisson packet arrivals with rate $\lambda(\alpha K+L)$ per frame. On the other hand not more than L packets can be transmitted per frame using any service discipline $g^{(L)}$. Thus this queuing system is unstable if (2) does not hold. ■

Now we will construct the upper bound for the system capacity C.

Proposition 8: For given α value, the following inequality

$$\max_{K,L} C(K, L, \mathcal{F}^{(K)}, \mathcal{G}^{(L)}) \leq \frac{1}{1 + \alpha/C_0}, \tag{3}$$

holds for the capacity of centralized reservation-based RMA systems.

Proof: Since from proposition 6 mean delay of packet transmission may be finite if $\lambda(\alpha K + L) < C_0K$, we easily obtain that it may be finite if arrival rate λ satisfies

$$\lambda < \frac{C_0\frac{K}{L}}{\alpha\frac{K}{L} + 1}. \tag{4}$$

On the other hand from proposition 7 mean delay of packet transmission may be finite if $\lambda(\alpha K + L) < L$, hence it may be finite if λ satisfies

$$\lambda < \frac{1}{\alpha \frac{K}{L} + 1}. \tag{5}$$

From (4) and (5) we obtain, that

$$\lambda < \min\left(\frac{C_0 \frac{K}{L}}{\alpha \frac{K}{L} + 1}, \frac{1}{\alpha \frac{K}{L} + 1}\right),$$

what leads to $\max_{K/L} C(K, L, \mathcal{F}^{(K)}, \mathcal{G}^{(L)}) = \frac{1}{\alpha/C_0 + 1}$ for $K/L = 1/C_0$ and proves (3). ■

Finally, let us construct a lower bound for the system capacity C. For this purpose we consider the part-and-try RMA algorithm, which, as previously mentioned, is the fastest known one for the basic model. From proposition 2 follows, that algorithm exists in class $\mathcal{F}^{(K)}$, which has exactly the same transmission rate. Moreover, an explicit way to construct it is provided in the proof of proposition 2. Let us denote this RMA algorithm as $\phi^{(K)}$. Then the following proposition can be proven.

Proposition 9: Let in the centralized reservation-based RMA system $\phi^{(K)}$ RMA algorithm and first-input-first-output (FIFO) service discipline (denoted as $\varphi^{(L)}$) be used. Then maximal transmission rate of multiple-access protocol $(\phi^{(K)}, \varphi^{(L)})$ for all K and L can be made arbitrary close to $\frac{R_{pt}}{\alpha + R_{pt}}$, where R_{pt} is the transmission rate of the part-and-try-algorithm.

Proof: One can show, that necessary and sufficient condition, that request mean delay for random access is finite, is

$$\lambda(\alpha K + L) < R_{pt} K. \tag{6}$$

Let λ justifies condition (6). Then central station queue becomes $G/D/L$ FIFO queuing system, which input traffic represents the outcome of K basic RMA systems, where subscribers independently operate according to part-and-try algorithm. One can show, that for this queuing system Baccelli-Foss conditions [10] are satisfied. Therefore, condition

$$\lambda(\alpha K + L) < L. \tag{7}$$

is the necessary and sufficient condition, that mean packet delay in the queue is finite.

From conditions (6) and (7) and using approach analogous to one used in the proof of proposition 8 we obtain, that mean packet delay is finite if and only if both $\lambda < \frac{R_{pt} \frac{K}{L}}{\alpha \frac{K}{L} + 1}$ and $\lambda < \frac{1}{\alpha \frac{K}{L} + 1}$ and taking into account the fact, that for any $\epsilon > 0$ a pair (K, L) exists for which $|K/L - 1/R_{pt}| < \epsilon$, the proposition is proven. ■

From the proof of this proposition the *corollary* directly follows: the maximal transmission rate of multiple-access protocol $(\phi^{(K)}, \varphi^{(L)})$ is achieved, when $\frac{K}{L} \approx \frac{1}{R_{pt}}$.

VI. Some Practical Remarks

We have introduced upper and lower bounds for Tsybakov's capacity of centralized reservation-based RMA system. In contemporary IEEE 802.16 broadband wireless network a version of so-called binary exponential back-off RMA algorithm is used for lower priority traffic requests [2]. This algorithm is shown to have zero transmission rate for infinite-users basic RMA model in [11]. For finite, but fairly large number of users, value $\ln(2)/2$ can represent some analog of transmission rate [12].

Computation of mean packet delay in the centralized reservation-based RMA system for the general case is an open question and is out of the scope of this paper. Nevertheless, an interesting application of the results is observed, if we apply our analysis for the case when some practical system is investigated, where "rational" algorithm $f^{(K)}$ having transmission rate R_0, which is independent of K, and some "simple" service discipline $g^{(L)}$ (like FIFO), are implemented. Thus, transmission rate of this multiple-access protocol is $R = \min(\frac{R_0 K}{\alpha K + L}, \frac{L}{\alpha K + L})$. For such system we would like to set up the following *hypothesis*: the ratio K/L, which minimizes the mean packet delay value $D(\lambda, K, L, f^{(K)}, g^{(L)})$, is a non-decreasing function of arrival rate λ and for any α, values of this function lie in a narrow interval not wider than $[1, 1/R_0]$. Moreover, mean delay itself is minimized, when K and L are minimal among those satisfying optimal ratio K/L. Thus, taking into account our hypothesis, frame structure can be optimally designed and is *almost independent of the ratio between the duration of request and packet transmission*.

REFERENCES

[1] I. Rubin, "Access-Control Disciplines for Multi-Access Communication Channels: Reservation and TDMA Schemes," *IEEE Transactions on Information Theory*, Vol. IT-25, No. 25, pp. 516–538, September 1979.

[2] IEEE Std 802.16-2004 - IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems.

[3] B. S. Tsybakov, "Survey of USSR Contributions to Random Multiple-Access Communications," *IEEE Transactions on Information Theory*, Vol. IT-31, No. 2, pp. 143–165, March 1985.

[4] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1st ed., 1987; 2nd ed., 1992.

[5] B. S. Tsybakov and M. A. Berkovskii, "Multiple Access with Reservation," *Problems of Information Transmission*, Vol. 16, No. 1, pp. 35-54, January–March 1980.

[6] A. Ephremides and B. Hajek, "Information Theory and Communication Networks: An Unconsummated Union," *IEEE Transactions on Information Theory*, Vol. 44, No. 6, pp. 2416–2434, October 1998.

[7] B. S. Tsybakov and N. B. Likhanov, "Upper Bound on the Capacity of a Random Multiple-Access System," *Problems of Information Transmission*, Vol. 23, No. 3, pp. 224–236, July–September 1987.

[8] B. S. Tsybakov and V. A. Mikhailov, "Random Multiple Packet Access: Part-and-Try Algorithm," *Problems of Information Transmission*, Vol. 16, No. 4, pp. 305–317, October–December 1980.

[9] B. Hajek, N. B. Likhanov and S. Tsybakov, "On the Delay in a Multiple-Access System with Large Propagation Delay," *IEEE Transactions on Information Theory*, Vol. 40, No. 4, pp. 1158–1166, July 1994.

[10] F. Baccelli and S. Foss, "On the Saturation Rule for the Stability of Queues," *Journal of Applied Probability*, 32, 2, pp. 494–507, 1995.

[11] D. Aldous,"Ultimate Instability of Exponential Back-off Protocol for Acknowledgment-based Transmission Control of Random Access Communication Channels," *IEEE Transactions on Information Theory*, Vol. 33, No. 2, pp. 219–233, March 1987.

[12] N.-O. Song, B.-J. Kwak and L. E. Miller, "On the Stability of Exponential Backoff," *Journal of Research of the National Institute of Standards and Technology*, Vol. 108, No. 4, pp. 289–297, July-August 2003.