ISSN 0032-9460, Problems of Information Transmission, 2010, Vol. 46, No. 2, pp. 184–200. © Pleiades Publishing, Inc., 2010.
Original Russian Text © A.M. Turlikov, S.G. Foss, 2010, published in Problemy Peredachi Informatsii, 2010, Vol. 46, No. 2, pp. 91–109.

COMMUNICATION NETWORK THEORY

On Ergodic Algorithms in Random Multiple Access Systems with "Success-Failure" Feedback[1]

A. M. Turlikov[a] **and S. G. Foss**[b]
[a]*St. Petersburg State University of Aerospace Instrumentation*
turlikov@vu.spb.ru
[b]*Sobolev Institute of Mathematics, Siberian Branch of the RAS, Novosibirsk*
Heriot-Watt University, Edinburgh
s.foss@hw.ac.uk

Received August 27, 2009; in final form, February 15, 2010

Abstract—We consider a decentralized multiple access system with a binary "success-failure" feedback. We introduce a family of algorithms (protocols) called "algorithms with delayed intervals" and study stability conditions of one of them. Then we discuss some numerical results and a number of related and interesting problems and hypotheses.

DOI: 10.1134/S0032946010020067

1. INTRODUCTION

In the late 1970s, Tsybakov and Mikhailov [1] and Capetanakis [2] considered a model with infinite number of users and a single transmission channel which is available to all users and transmits messages between them. The authors proposed an algorithm that allows to transmit messages with a finite mean delay given that the input intensity is below a certain threshold. The algorithm is based on the use of the so-called ternary feedback. This means that the users can observe the channel output and distinguish three possible situations, either no transmissions ("Empty"), or transmission from a single user ("Success"), or a collision of messages from two or more users ("Conflict"). Soon afterwards, following [1, 2], algorithms with binary feedback, "Empty–Nonempty" and "Conflict–Nonconflict", were introduced and studied.

Performance of the model with "Success–Nonsuccess" (S–NS) feedback is less definite. Here a user cannot distinguish collisions and empty slots. Several algorithms that were proposed in [3–5] could guarantee a stable behaviour of the system only given that certain model extensions are made, like introducing a special testing file, etc. In [6] the authors proposed an idea of a new algorithm that may provide stable performance of the system with S–NS feedback and without a model extension, but they did not give a precise description. Then a description of one such algorithm was given in [7], where the authors presented balance equations for the stationary distribution of a corresponding Markov chain and numerically found the capacity of this algorithm, i.e., a number λ_0 such that the balance equations have a solution if and only if the input rate λ is smaller than λ_0. We are unaware of any further studies on algorithms for systems with the "S–NS" feedback.

Modern communication systems deal with a variety of multiple access algorithms including random multiple access; see, for example, systems based on standards IEEE 802.11 and IEEE 802.16. Moreover, one may say that these standards deal with the "S–NS" feedback. In particular, in standard IEEE 802.16, the base station does not distinguish collisions from empty slots. It is known that the algorithms used in practice do not provide stable dynamics if the number of users is infinite

[1] Supported in part by the Russian Foundation for Basic Research, project no. 07-01-00077.

(see, e.g., [8]). This problem is not of great importance for existing networks with a relatively small number of users, but an increase in this number may lead to essential delays. Therefore, designing algorithms that stabilize systems with "S–NS" feedback may have not only theoretical but also practical importance.

In this paper we introduce a rather general class of algorithms with binary "S–NS" feedback, which includes an algorithm from [7]. Then we analyze the simplest algorithm from this class, which is similar to that from [7] (see Theorems 1 and 2). Similar analysis is applicable to other algorithms, and the chosen one has the only advantage that it has the minimal possible number of free parameters. This allows us to supplement its study with a tractable figure (see Section 5.1).

The paper is organized as follows. In Section 2 we introduce a multiple access model and describe an algorithm chosen for the analysis. Then we provide a general description of the class of algorithms for "S–NS" feedback. The common feature of these algorithms is that they "browse" (messages that arrive within) some time intervals, then remove "successful" intervals and move "unsuccessful" ones to a queue, and then take them when a new "successful" interval appears. Following [7], this class is called a class of algorithms with delayed intervals. In Section 3 we study the chosen algorithm and show how one can determine the stability region and optimize parameters of the algorithm. The stability problem is here reduced to the study of conditions for positive recurrence and ergodicity of a certain two-dimensional Markov chain. In Section 3 we formulate several results for this Markov chain; their proofs are given in Section 4. In Section 5, we provide a comparative analysis of the chosen algorithm and of other algorithms that were introduced in Section 2; we also discuss a number of open problems. Finally, the Appendix contains a number of known auxiliary results.

2. MODEL AND MULTIPLE ACCESS ALGORITHM

2.1. Model

Here we describe and study a variant of the multiple access model introduced in [9]. There is an infinite number of users and a single transmission channel which is used by the users for message exchange. It is assumed that all messages have the same length and that it takes a unit of time to transmit a message. The message arrival process to the system forms a time-homogeneous Poisson flow with rate λ (thus, interarrival times form an i.i.d. sequence having a common exponential distribution with rate λ and mean $1/\lambda$).

We introduce a number of assumptions on the functioning and accessibility of the transmission channel, which almost coincide with assumptions from [9], and the only difference is in Assumption 3 (see below).

Assumption 1. Transmission time is slotted. All slots have the same length with the transmission time of a message. Time slots are numbered by positive integers, and slot t corresponds to time interval $[t, t+1)$ (for short, we say *slot* t instead of *slot number* t). Beginnings and ends of all time slots are known to all users. Message transmission may start only at the beginning of a time slot.

Assumption 2. Within any time slot, only one of the following three events may occur:

- Single transmission (event S, success);
- No transmission (event E, empty);
- Two or more transmissions (event C, conflict, collision).

Assumption 3. At the end of each time slot, each user can observe whether the slot was successful or not. If the latter, then "passive" users (those who did not transmit their messages within this time slot) only know that there was no success but cannot distinguish what has occurred, E or C. (This is the difference with the model from [9], where such a possibility was given.)

Denote by θ_i the indicator of the event {transmission in slot i is successful}, i.e., a random variable taking value 1 if the event occurs, and 0 otherwise. We say that a sequence $\theta(t) = \{\theta_1, \ldots, \theta_t\}$ is the *channel history* by time t. We assume that, at time t, this history is available to all users.

Assumption 4. A user can hold only one message. For short, we will say *message* x instead of *message arriving at time* x. Let t_x denote the integer part of a number x, i.e., an integer such that $t_x \le x < t_x + 1$.

A user that holds a message x may use the value x to make a decision in which slots he will transmit it in the future, starting from slot $t_x + 1$. Denote by $\nu_i^{(x)}$ the indicator of the event {message x was transmitted in slot i}. A sequence $\nu^{(x)}(t) = \{\nu_0^{(x)}, \ldots, \nu_t^{(x)}\}$ is the *history of message* x by time t. If message x is transmitted in slot t, $\nu_t^{(x)} = 1$, and if this is the only such message, $\theta_t = 1$, then x is transmitted successfully and is removed from the system after that.

2.2. Multiple Access Algorithm

Following [9], a *multiple access algorithm* is a rule under which, at the beginning of each time slot t, a decision on message x (either to transmit it or not) is made. It is based on the common *channel history* $\theta(t-1)$ and on the individual *message history* $\nu^{(x)}(t-1)$. Such a decision may be either deterministic or random.

In this paper, we consider algorithms with the following two-phase decision rule:

1. For each t, at the beginning of slot t, all users observe the channel history $\theta(t-1)$ and choose somehow a common set $B(t)$ and a number $p_t \in [0,1]$;
2. Then the users take individual decisions (to transmit in slot t or not) as follows: if $x \in B(t)$, then message x is transmitted with probability p_t; otherwise, message x is not transmitted (with probability 1).

We will use the following terminology. The time when message x arrives is associated with point x on the time axis, and this point is removed after successful transmission of the message. If, at the beginning of time slot t, a set $B(t) = B$ and a number $p_t = p$ are chosen, then we say that B is *browsed with probability* p *in slot* t. If $p = 1$, then we say for short that the set is *browsed*, for short, and if $p < 1$, we say that it is *browsed with probability* p. If B is browsed, then it becomes *completely browsed* if $\theta(t) = 1$ and *partially browsed* otherwise (if $\theta(t) = 0$). Clearly, a union of completely browsed sets is completely browsed. This means that if a set B is browsed in slot t, then all messages from this set have been successfully transmitted and left the system by time $t+1$.

We describe operation of the algorithm following this terminology. The algorithm proceeds in sessions. The time axis is divided into time intervals of length $A + B$, where A and B are parameters of the algorithm. Session number 0 ends at time 0. Session k starts with browsing a new interval $[(k-1)(A+B), k(A+B))$, or, more precisely, browsing messages that arrive within this time interval. For $k \ge 1$, let s_k be the start, and e_k the end time of session k. After completion of session, say, $k-1$, the next session k starts immediately ($s_k = e_{k-1}$) if $e_{k-1} \ge k(A+B)$. Otherwise, the transmission channel stays empty during $\lceil (k(A+B) - e_{k-1}) \rceil$ time slots, and then session k starts; i.e., $s_k = e_{k-1} + \lceil (k(A+B) - e_{k-1}) \rceil$. Here $\lceil x \rceil$ is the smallest integer that is not smaller than x.

Each session k ($k \ge 1$) includes the following steps.

Step 1. In slot $s_k + 1$, interval $[(k-1)(A+B), k(A+B))$ is browsed. The interval contains two disjoint subintervals $[(k-1)A + B, kA + (k-1)B)$ and $[kA + (k-1)B, k(A+B))$, of lengths A and B. For short, we may speak about intervals $A + B$, A, and B.

If there is a success, $\theta(s_k + 1) = 1$, then interval $A + B$ becomes completely browsed (and is removed from consideration), and the session ends. Otherwise, the algorithm proceeds to Step 2.

Step 2. In slot s_k+2, interval A is browsed. If $\theta(s_k+2)=0$ (nonsuccess), then interval $A+B$ joins a queue of delayed intervals, and the session ends. Otherwise interval A becomes completely browsed. This means that interval B contains at least one message. Then there are three options.

Step 3.1. If the queue of delayed intervals is empty, then the algorithm applies a *procedure of browsing a nonempty set* (which is described below) to interval B. When the procedure ends, interval B becomes completely browsed (and is removed), and the session ends too.

Step 3.2. If the queue contains only one interval (say D), then the algorithm applies the procedure of browsing a nonempty set to the union of two intervals, B and D (which is also nonempty). When the procedure ends, both B and D become completely browsed (and then are removed), and the session ends too.

Step 3.3. If the queue contains two intervals or more, then the algorithm applies the procedure of browsing a nonempty set to the union of B and two first intervals from the queue. When the procedure ends, all three intervals become completely browsed and are then removed, and the session ends.

Now we describe the procedure of browsing a nonempty set.

Let V be a finite nonempty set. We start the procedure by letting $V=V_0$.

Action 1. The set V_0 is browsed with probability 1. If there is a success, then V_0 becomes completely browsed, and the procedure ends. Otherwise, we proceed to Action 2.

Action 2. The set V_0 is browsed with probability $\alpha\in(0,1)$. If there is no success, then the set is browsed again and again, until the first success. Then the "successful" element (say x) is removed. We let $V_0:=V_0\setminus\{x\}$ and return to Action 1.

Here α is a parameter of the procedure.

The algorithm that we have described depends on five parameters:

- Lengths of intervals A and B;
- Parameters α_0, α_1 and α_2 that are used in the procedure of browsing a nonempty set within Steps 3.1, 3.2, and 3.3, respectively (here an index of α shows how many intervals are taken from the queue).

Below we study this algorithm and, in particular, the following problems:

– Given values of the five parameters, do there exist values of λ that guarantee stable operation of the algorithm? If so, what are they?

– What is the maximum value of intensity λ (the capacity) which guarantees the stability?

2.3. Class of Algorithms with Delayed Intervals

The algorithm introduced in Section 2.2 may be called an *algorithm with delayed intervals.* Below we give a description of a class of such algorithms.

An algorithm proceeds in sessions. First, we choose an integer $N\geq 2$, and then, N positive numbers $D_1,D_2,\ldots,D_N$. A session contains a number of consecutive steps.

Step 1 of session k starts in slot s_k+1 with browsing a new interval of length $\sum\limits_{i=1}^{N} D_i$, which contains N disjoint intervals of lengths $D_1,D_2,\ldots,D_N$, respectively. If $\theta(s_k)=1$, then the session ends; otherwise, the algorithm proceeds to Step 2.

Step j (where $j<N$) proceeds in slot s_k+j with browsing an interval of length $\sum\limits_{i=1}^{N-j+1} D_i$, which is a union of the first $N-j+1$ intervals $D_1,\ldots,D_{N-j+1}$. If $\theta(s_k+j)=0$, then we proceed to Step $j+1$. Otherwise, the interval $\sum\limits_{i=1}^{N-j+1} D_i$ becomes completely browsed. This also means that the interval $\sum\limits_{i=n-j+2}^{N} D_i$ contains at least one message. Then one applies the procedure of browsing

a nonempty set to the union of the latter interval and some number of delayed intervals (if there are any). The number of delayed intervals that may be taken from the queue is another parameter of the algorithm.

Step N is analogous to Step $j < N$, with the only difference that if $\theta(s_k + N) = 0$, then the whole interval $\sum_{i=1}^{N} D_i$ joins the queue of delayed intervals.

For the algorithm from Section 2.2, we have

(a) $N = 2$, $D_1 = A$, and $D_2 = B$;
(b) If there are q delayed intervals in the queue, then the first $\min(2, q)$ of them are taken at Step 3.

This algorithm is the simplest stable algorithm in the introduced class: at the end of Section 3.3, we show that there is no stable algorithms with $N = 2$ if one cannot take more than one delayed interval from the queue.

For the algorithm introduced in [7], we have

(a) $N = 3$;
(b) The number of delayed intervals taken from the queue of length q equals $\min(q, 1)$.

3. ASYMPTOTIC ANALYSIS OF THE MAIN ALGORITHM: CONDITIONS FOR POSITIVE RECURRENCE AND ERGODICITY

3.1. Time Scaling

We scale the time axis by λ, so the length of a time slot becomes equal to λ. Then the input flow of messages becomes Poisson with parameter 1. An advantage of this scaling is that the new lengths of intervals $a = A\lambda$ and $b = B\lambda$ become free variables which are not related to λ.

Denote $L = a + b$. Introduce two new characteristics of the system: at each time t, they are

- The length $W(t)$ of a part of the input interval that has not been browsed yet;
- The number of delayed (and partly browsed) intervals $Q(t)$.

In the new time scale, we will use the same notation, s_k and e_k, as above for the start and end times of sessions.

Recall that, for $k = 1, 2, \ldots$, session k starts immediately after the end of session $(k-1)$ (i.e., $s_k = e_{k-1}$) if $W(e_{k-1}) \geq L$. Otherwise, there is a delay of i time units, where i is the smallest integer such that $W(e_{k-1}) + i\lambda \geq L$. In other words, $i = \lceil (L - W(e_{k-1}))/\lambda \rceil$, and then $s_k = e_{k-1} + \ i\lambda$ and $W(s_k) = W(e_{k-1}) + i\lambda$.

Let T_k be the duration of session k, $k = 1, 2, \ldots$. During the session, the length of a nonbrowsed interval increases by λT_k, i.e.,

$$W(e_k) = W(s_k) - L + \lambda T_k,$$

and at the end of the session there can be three scenarios:

- either the browsed interval is removed and the number of delayed intervals is unchanged,

$$Q(e_k) = Q(e_{k-1});$$

- or the browsed interval is removed together with the only interval in the queue,

$$Q(e_k) = Q(e_{k-1}) - 1,$$

or with two intervals from the queue,

$$Q(e_k) = Q(e_{k-1}) - 2;$$

- or the interval is browsed only partially and joins the queue,

$$Q(e_k) = Q(e_{k-1}) + 1.$$

We first consider a subsequence of two-dimensional vectors at embedded epochs of session ends

$$(W_k, Q_k) := (W(e_k), Q(e_k)). \tag{1}$$

By the model construction and since the input is a Poisson process, sequence (1) forms a time-homogeneous Markov chain. We study conditions for its recurrence/transience in terms of parameters λ, a, and b and additional parameters α_0, α_1, and α_2. Note that the sequence $(W(\lambda n), Q(\lambda n))$ is not Markov in general. We will show that transience of the embedded chain implies a similar property of the sequence $(W(\lambda n), Q(\lambda n))$, and that recurrence of the embedded chain implies that, under an extra technical condition, the sequence $(W(\lambda n), Q(\lambda n))$ becomes regenerative and aperiodic. This, in turn, will imply existence of a stationary version of this sequence and convergence to that in the total variation norm.

3.2. Probabilities of Events in a Session

Recall that, for each session, there are three possibilities:

- Session ends at Step 1, and the number of delayed intervals does not change;
- Session ends at Step 2, and the queue increases by one;
- Session ends at Step 3, and if the queue was $q \geq 1$, it decreases by $\min(2, q)$.

Let p_0, p_1, and p_- be the respective probabilities of these events, and X_a and X_b be the numbers of points of Poisson process in disjoint time intervals of lengths a and b.

In Step 1, an interval of length $a + b$ is browsed, and it becomes completely browsed if

$$X_a + X_b = 1.$$

The probability of this event is

$$p_0 = \mathbf{P}(X_a + X_b = 1) = (a + b)e^{-a-b}.$$

Further, we have

$$\begin{aligned} p_- &= \mathbf{P}(X_a + X_b \neq 1,\ X_a = 1) \\ &= \mathbf{P}(X_a = 1,\ X_b \geq 1) \\ &= \mathbf{P}(X_a = 1)\,\mathbf{P}(X_b \geq 1) \\ &= ae^{-a}(1 - e^{-b}) \end{aligned}$$

and

$$p_1 = 1 - p_0 - p_- = 1 - be^{-a-b} - ae^{-a}.$$

3.3. Embedded Markov Chain

Recall again that if at the beginning of a session there were some delayed intervals (say q), then at the end of a session their number either decreases by $\min(q, 2)$ (with probability p_-), or increases by one (with probability p_1), or stays unchanged (with probability p_0). If there were no delayed intervals ($q = 0$), then a new delayed interval either appears at the end (with probability p_1) or not (with probability $p_- + p_0$). Therefore, the one-dimensional sequence Q_n is also Markov,

$$Q_{n+1} = \max(Q_n + \xi_n, 0), \tag{2}$$

where $\{\xi_n\}$ is a sequence of i.i.d. random variables,

$$\mathbf{P}(\xi_n = 1) = p_1, \qquad \mathbf{P}(\xi_n = 0) = p_0, \qquad \mathbf{P}(\xi_n = -2) = p_-.$$

This Markov chain admits a (unique) stationary distribution if and only if $\mathbf{E}\,\xi_n < 0$, i.e., $h := p_1/2p_- < 1$. Denote this stationary distribution by $\{\pi_i\}_{i\ge 0}$.

Sequence (2) is known as an *integer-valued random walk stopped at zero*. In our case this random walk is *right-continuous*; i.e., $\mathbf{P}(\xi_n \ge 2) = 0$. Note that all other algorithms described in Section 2.3 are also right-continuous. It is known (see, e.g., [10, ch. 11]) that for a right-continuous random walk its stationary distribution is geometric,

$$\pi_i = \pi_0(1-\pi_0)^i, \quad i \ge 0,$$

and that π_0 is a unique solution z to the equation $\sum \mathbf{P}(\xi_n = j)(1-z)^{-j} = 1$ in the domain $z \in (0,1)$. In our case,

$$\pi_0 = \frac{3-\sqrt{1+8h}}{2}.$$

Recall also the following well-known facts (which can also be found, say, in [10]). Assume that $h < 1$. Let the initial value be $Q_0 = m \ge 0$, and let

$$\tau^{(m)} = \min\{n \ge 1 :\ Q_n = 0 \mid Q_0 = m\}.$$

Then $\tau^{(m)}$ has all power moments finite, $\mathbf{E}\big(\tau^{(m)}\big)^k < \infty$ for all $k > 0$; moreover, there exists a finite exponential moment, $\mathbf{E}\,e^{c\tau^{(m)}} < \infty$, for some $c = c_m > 0$. In particular, $\mathbf{E}\,\tau^{(0)} = 1/\pi_0$ and $\mathbf{E}\,\tau^{(m)} \le C + m/(2p_- - p_1)$ for some C and all $m \ge 1$. Further, this Markov chain is geometrically ergodic, i.e., there exists a constant C and, for any $m \ge 0$, a constant c_m such that, for all $n \ge 0$,

$$\sup_k \big|\mathbf{P}(Q_n = k \mid Q_0 = m) - \pi_k\big| \le c_m e^{-Cn}.$$

Note that the inequality $h < 1$ may hold for some positive parameters a and b. Indeed, it is equivalent to the inequality

$$2p_- - p_1 = 3ae^{-a} + (b-2a)e^{-a-b} - 1 > 0,$$

which holds if, say, $a = 1$ and $b = 2$.

Note also that, for a simpler algorithm where at most one delayed interval can be taken from the queue, the embedded Markov chain might be positive recurrent only if $p_- > p_1$. But the latter is equivalent to the inequality

$$2ae^{-a} - ae^{-a-b} + be^{-a-b} > 1,$$

which has no solutions in the set of positive real numbers.

3.4. Procedure of Browsing a Nonempty Set

Assume that we know that a set D is nonempty, but its cardinality, say X, is unknown. Assume further that X is random and has a known distribution. Introduce the following "identification" algorithm for elements of this set.

At the first step, the whole set is browsed with probability one. If it contains only one element, the procedure stops.

Otherwise, at each of the subsequent steps, each element of the set is browsed with a fixed probability $\alpha \in (0,1)$, until the only element is identified. Clearly, given $X = n$, the number of attempts is random and has a geometric distribution with parameter $r_{n,\alpha} = n\alpha(1-\alpha)^{n-1}$ and mean $1/r_{n,\alpha}$.

Then the identified element is removed, and the procedure is repeated: first, all the elements are browsed with probability one and then, if their number exceeds one, the elements are browsed repeatedly with probability α, and a next one is identified and removed. The procedure continues until the set becomes empty.

Denote by $R_\alpha(X)$ the duration (i.e., the number of steps) of this procedure. Then

$$\begin{aligned}\mathbf{E}\,R_\alpha(X) &= \sum_{n=1}^{\infty} \mathbf{E}(R_\alpha(X) \mid X = n)\,\mathbf{P}(X = n)\\ &= \sum_{n=1}^{\infty} \mathbf{P}(X = n)\Big(1 + \sum_{m=2}^{n} (1 + 1/r_{m,\alpha})\Big)\\ &= \mathbf{E}\,X + \sum_{n=2}^{\infty} \mathbf{P}(X = n) \sum_{m=2}^{n} \frac{1}{r_{m,\alpha}}.\end{aligned}$$

Note that

$$\max_\alpha r_{m,\alpha} = r_{m,1/m} = (1 - 1/m)^{m-1}.$$

Therefore, for any $\alpha \in (0,1)$,

$$\mathbf{E}\,R_\alpha(X) \ge \mathbf{E}\,X + \sum_{n=2}^{\infty} \mathbf{P}(X = n) \sum_{m=2}^{n} (1 - 1/m)^{-m+1}. \tag{3}$$

Since $(1-1/m)^{-m+1} > 2$ for all $m \ge 2$, inequality (3) implies, in particular, the following simple lower bound:

$$\mathbf{E}\,R_\alpha(X) \ge 3\,\mathbf{E}\,X - 2. \tag{4}$$

3.5. Mean Duration of a Session

We continue to assume that $h < 1$. Let T be the duration of a typical session in the stationary regime. Recall that T depends on five parameters, $T = T(a, b, \alpha_0, \alpha_1, \alpha_2)$. Recall also that the delayed intervals form a queue, and denote by $Y^{(i)}$ the number of messages in the ith interval in the queue. By the construction, the random variables $Y^{(i)}$, $i = 1, 2, \ldots$, are i.i.d.

Recall that there are three options:

1. With probability p_0, there is only one message in the interval $a + b$; then $T = 1$;
2. With probability p_1, consecutive browsing of intervals $a + b$ and a shows that the number of messages in each of them differs from one; then $T = 2$;
3. If $X_{a+b} \ne 1$ and $X_a = 1$ (this occurs with probability p_-), then the session duration is $T = 2 + T_+$, where T_+ equals either $R_{\alpha_0}(\widetilde{X}_b)$ if the queue of delayed intervals is empty (i.e., with probability π_0), or $R_{\alpha_1}(Y^{(1)} + \widetilde{X}_b)$ if there is only one interval in the queue (i.e., with probability π_1), or $R_{\alpha_2}(Y^{(1)} + Y^{(2)} + \widetilde{X}_b)$ if there are two or more intervals in the queue (i.e., with probability $(1 - \pi_0 - \pi_1)$).

Here $\widetilde{X}_b$ is a random variable with distribution $\mathbf{P}(\widetilde{X}_b \in \cdot) = \mathbf{P}(X_b \in \cdot \mid X_b \ge 1)$. We have the equalities

$$\begin{aligned}\mathbf{E}\,T &= p_0 + 2p_1 + (2 + \mathbf{E}\,T_+)p_- = 2 - p_0 + p_-\,\mathbf{E}\,T_+\\ &= 2 - p_0 + (E_0\pi_0 + E_1\pi_1 + E_2(1 - \pi_0 - \pi_1))\,p_-,\end{aligned}$$

where

$$E_0 = \mathbf{E}\, R_{\alpha_0}(\widetilde{X}_b) = \mathbf{E}\, \widetilde{X}_b + \sum_{m\ge 2} \frac{1}{r_{m,\alpha_0}} \mathbf{P}(\widetilde{X}_b \ge m),$$

$$E_1 = \mathbf{E}\, R_{\alpha_1}(Y^{(1)} + \widetilde{X}_b) = \mathbf{E}\, Y^{(1)} + \mathbf{E}\, \widetilde{X}_b + \sum_{m\ge 2} \frac{1}{r_{m,\alpha_1}} \mathbf{P}(Y^{(1)} + \widetilde{X}_b \ge m),$$

$$E_2 = \mathbf{E}\, R_{\alpha_2}(Y^{(1)} + Y^{(2)} + \widetilde{X}_b) = \mathbf{E}\, Y^{(1)} + \mathbf{E}\, Y^{(2)} + \mathbf{E}\, \widetilde{X}_b + \sum_{m\ge 2} \frac{1}{r_{m,\alpha_2}} \mathbf{P}(Y^{(1)} + Y^{(2)} + \widetilde{X}_b \ge m).$$

Below are formulas for means of the random variables $\widetilde{X}_b$ and Y (where Y has the same distribution with the random variables $Y^{(i)}$, $i = 1, 2$), while the above sums can be explicitly found only by using numerical methods. We have

$$\mathbf{E}\, \widetilde{X}_b = \mathbf{E}(X_b \mid X_b \ge 1) = \frac{\mathbf{E}\, X_b}{\mathbf{P}(X_b \ge 1)} = \frac{b}{1 - e^{-b}},$$

$$\begin{aligned}\mathbf{E}\, Y &= \mathbf{E}(X_a + X_b \mid X_a \ne 1,\ X_a + X_b \ne 1)\\ &= \frac{1}{p_1} \mathbf{E}\big((X_a + X_b)\big(\boldsymbol{I}(X_a \ge 2) + \boldsymbol{I}(X_a = 0,\ X_b \ge 2)\big)\big)\\ &= \frac{1}{p_1}\big(a + b - ae^{-a} - be^{-a-b} - abe^{-a}\big).\end{aligned}$$

Here $\boldsymbol{I}$ is the indicator function: $\boldsymbol{I}(B) = 1$ if event B occurs, and $\boldsymbol{I}(B) = 0$ otherwise.

3.6. Conditions for Positive Recurrence and Stability

Recall that, in the stationary regime, from $N = 0$ to $N = 3$ time intervals of length $a + b$ can be browsed in a single session, with probabilities

$$\begin{aligned}&\mathbf{P}(N = 0) = p_1, &&\mathbf{P}(N = 1) = p_0 + p_-\pi_0,\\ &\mathbf{P}(N = 2) = p_-\pi_1, &&\mathbf{P}(N = 3) = p_-(1 - \pi_0 - \pi_1).\end{aligned} \tag{5}$$

Here $\mathbf{E}\, N = 1$, and the mean cumulative length of all removed (completely browsed) slots is

$$L = (a + b)\, \mathbf{E}\, N = a + b$$

(the latter makes sense since, within a typical session, the mean number of removed slots can be neither greater nor smaller than the number of new slots, which is 1). Further, assume that the Markov chain starts from the state $(W_0, 0)$ with $Q_0 = 0$, and let $\tau = \min\{n :\ Q_n = 0\}$. Let N_i be the number of slots browsed in slot i, and T_i the length of session i. Then

$$\mathbf{E}\left(\sum_{i=1}^{\tau} N_i\right) = \mathbf{E}\, \tau\, \mathbf{E}\, N = \mathbf{E}\, \tau \qquad \text{and} \qquad \mathbf{E}\left(\sum_{i=1}^{\tau} T_i\right) = \mathbf{E}\, \tau\, \mathbf{E}\, T, \tag{6}$$

where T is the length of a typical session in the stationary regime (which was studied in Section 3.5).

Definition 1. A Markov chain (W_n, Q_n) is *recurrent* if there is a bounded set $A = \{W \le c_1,\ Q \le c_2\}$ such that

(1) $\tau_{(W,Q)} = \tau_{(W,Q)}(A) = \min\{n \ge 1 :\ (W_n, Q_n) \in A \mid W_0 = W,\ Q_0 = Q\} < \infty$ a.s., for any initial value (W, Q).

A recurrent Markov chain is *positive recurrent* if

$$(2) \quad \sup_{(W,Q)\in A} \mathbf{E}\,\tau_{(W,Q)} < \infty,$$

and *null recurrent* otherwise.

We say that a Markov chain (W_n, Q_n) is *transient* if $W_n + Q_n \to \infty$ a.s. as $n \to \infty$, for any initial value $W_0 = W$, $Q_0 = Q$.

Remark 1. Definitions of positive and null recurrence are standard. In fact, there are several definitions of transience, and our definition corresponds to the most restrictive one.

Definition 2. An algorithm is *positive/null recurrent* or *transient* if the corresponding Markov chain (W_n, Q_n) is such.

Theorem 1. *For a Poisson input flow with intensity λ and for any positive numbers a and b and any collection of probabilities α, the algorithm described above is*

(a) *positive recurrent if $2p_- > p_1$ and $\lambda < L/\mathbf{E}\,T$, and*
(b) *transient if either $\lambda > L/\mathbf{E}\,T$ or $2p_- < p_1$.*

Remark 2. One can show that, under conditions (a) of Theorem 1, the set A is positive recurrent for *any* choice of positive numbers c_1 and c_2.

Remark 3. One can also show that if $2p_- = p_1$ and $\lambda < L/\mathbf{E}\,T$, then the Markov chain is *null recurrent*. Most likely, the same holds if $2p_- > p_1$ and $\lambda = L/\mathbf{E}\,T$. If this is true, then Theorem 1 can be formulated as a criterion: the algorithm under consideration is positive recurrent if and only if conditions (a) hold.

Remark 4. The ratio $L/\mathbf{E}\,T$ is the *rate* of the algorithm. Recall that this ratio depends on five parameters.

Definition 3. A Markov chain $\{(W_n, Q_n)\}$ (and the corresponding algorithm) is *ergodic* if it has a unique stationary distribution Π and, moreover, for any initial condition (W_0, Q_0), distributions of random vectors $\{(W_n, Q_n)\}$ converge, as n grows, to the stationary one; and *strong ergodic* if in addition the convergence is in the *total variation norm*, i.e.,

$$\sup |\mathbf{P}((W_n, Q_n) \in B) - \Pi(B)| \to 0, \quad n \to \infty,$$

where the supremum is taken over all two-dimensional measurable sets B.

Note that, in general, positive recurrence of a Markov chain does not guarantee existence (and uniqueness) of its stationary distribution.

Theorem 2. *Let $C = \sup L/\mathbf{E}\,T$, where the supremum is taken over all values of the five parameters for which $2p_- > p_1$ (one may call C the capacity of the family of algorithms under consideration).*

(a) *If $\lambda < C$, then one can choose parameters a and b and a collection of probabilities α which make this algorithm strong ergodic. Then the basic process $(W(t), Q(t))$ is also ergodic and aperiodic, and therefore there exists a proper stationary distribution, which is the limiting distribution of vectors $(W(t), Q(t))$ in the total variation norm as t grows.*
(b) *If $\lambda > C$, then all algorithms under consideration are transient.*

Remark 5. One can quite easily find an upper bound for C. Namely, one can apply (three times) the lower bound (3) and then consider a simpler 2-parameter optimization problem.

4. PROOFS

4.1. Proof of Theorem 1 (a)

Consider embedded moments of starts of the sessions k_n for which $Q_{k_n} = 0$. These moments divide time into cycles of lengths $\{k_n - k_{n-1}\}$, which are i.i.d. random variables having a common distribution with random variable τ introduced in Section 3.6. Let $\widetilde{W}_n = W_{k_n}$.

First we show that the Markov chain $\widetilde{W}_n$ is positive recurrent. For that, we apply the first part of Foster's criterion (see Theorem 3 in the Appendix). In this part of the proof, we may assume $k_0 = 0$. Then, clearly, $\tau = k_1 - k_0$.

Denote

$$\Delta_x = \mathbf{E}(\widetilde{W}_1 \mid \widetilde{W}_0 = x) - x$$

and show that

- the number Δ_x is bounded from above by the same constant for all x, and
- $\limsup\limits_{x\to\infty} \Delta_x < 0$.

Then Foster's criterion can be applied.

Indeed, the cumulative length of all browsed intervals in this cycle is $L\sum\limits_{i=1}^{\tau} N_i$, where $L = a + b$ and N_i is the number of browsed intervals during the ith session. Note that $\mathbf{E}\sum\limits_{i=1}^{\tau} N_i = \mathbf{E}\,\tau$, due to (6).

Since the total increase of W during time t is λt, its increase during the first cycle is not smaller than $\lambda\sum\limits_{i=1}^{\tau} T_i$, and not greater than

$$\lambda\sum_{i=1}^{\tau} T_i + (L+1)\sum_{i=1}^{\tau} \boldsymbol{I}(W_i^x < L).$$

Here T_i is the duration of the ith session, and the upper index x means that the first session in the cycle starts with $W_0 = \widetilde{W}_0 = x$. Then $W_i^x \ge x - 2iL$ for all x and i, and therefore

$$0 \le \sum_{i=1}^{\tau} \boldsymbol{I}(W_i^x < L) \le \sum_{i=1}^{\tau} \boldsymbol{I}(x - iL < L) \le \tau \boldsymbol{I}(x < 2L\tau + L),$$

where the upper bound $\tau\boldsymbol{I}(x < 2L\tau + L)$ tends monotonically to zero as x grows, both a.s. and in mean (this follows from Lebesgue's theorem). Therefore, as $x \to \infty$, we have

$$\Delta_x \to \lambda\,\mathbf{E}\,\tau\,\mathbf{E}\,T - L\,\mathbf{E}\,\tau\,\mathbf{E}\,N = \mathbf{E}\,\tau\,\mathbf{E}\,T\,(\lambda - L/\mathbf{E}\,T) < 0.$$

Here N is the total number of browsed intervals within a cycle, and it has distribution (5). Since

$$\Delta_x \le \mathbf{E}\Bigg(\lambda\sum_{i=1}^{\tau} T_i + L\sum_{i=1}^{\tau} \boldsymbol{I}(W_I^x < L)\Bigg) \le \lambda\,\mathbf{E}\,\tau\,\mathbf{E}\,T + L\,\mathbf{E}\,\tau < \infty$$

for all x, the Markov chain $\{\widetilde{W}_n\}$ is positive recurrent.

Now we show that (W_n, Q_n) is positive recurrent too. For that, we apply the first part of the generalized Foster criterion (see Theorem 4 from the Appendix).

Let $W_0 = W \ge 0$ and $Q_0 = m \in \{0, 1, 2, \ldots\}$. Then, in the notation of Section 3.3, the random variable $\tau^{(m)}$ is a.s. finite and, moreover, has a finite mean. In this part of the proof, we have to let $k_0 = \tau^{(m)}$. Further, $\widetilde{W}_0 = W_{\tau^{(m)}}$ and

$$\mathbf{E}\,\widetilde{W}_0 \le W + \mathbf{E}\,\tau^{(m)} C = W + m\widetilde{C},$$

where $C = L + K_0$,

$$K_0 = 2 + \max\Big(\mathbf{E}\, R_{\alpha_0}(\widetilde{X}_b),\, \mathbf{E}\, R_{\alpha_1}(Y^{(1)} + \widetilde{X}_b),\, \mathbf{E}\, R_{\alpha_2}(Y^{(1)} + Y^{(2)} + \widetilde{X}_b)\Big) < \infty,$$

and $\widetilde{C} = C/(2p_- - p_1)$. We take a test function g of the form $g(w,m) = w + m$.

Let $\widetilde{\mu} = \min\{n \ge 0 :\ \widetilde{W}_n \le g_0\}$. Then, as follows from the above,

$$\mathbf{E}(\widetilde{\mu} \mid \widetilde{W}_0) \le K(\widetilde{W}_0 + 1)$$

for some constant K. If we denote

$$\gamma = \min\{n :\ W_n + Q_n \le g_0\},$$

then

$$\mathbf{E}(\gamma \mid W_0 = W, Q_0 = m) \le \mathbf{E}\,\tau^{(m)} + \mathbf{E}\Bigg(\sum_{i=1}^{\widetilde{\mu}} z_i\Bigg),$$

where z_i is the length of cycle i (these lengths are i.i.d. with a finite mean). Thus, one can find another absolute constant $\widehat{K}$ such that

$$\mathbf{E}(\gamma \mid W_0 = W,\, Q_0 = m) \le \widehat{K}(W + m + 1).$$

Therefore, Theorem 4 can be applied, and the Markov chain $\{W_n, Q_n\}$ is positive recurrent.

4.2. Proof of Theorem 1 (b)

If $\mathbf{E}\,\xi_n = p_1 - 2p_- > 0$, then $\sum\limits_{i=1}^{n} \xi_i \to \infty$ a.s., eventually all Q_n are strictly positive, and

$$\frac{Q_n}{n} \to \mathbf{E}\,\xi_1 > 0 \quad \text{a.s.,} \quad \text{as } n \to \infty,$$

by the strong law of large numbers.

Let now $p_1 < 2p_-$ and $\lambda > L/\mathbf{E}\,T$. Then again the cycles have finite mean $\mathbf{E}\,\tau$, and the mean number of sessions per cycle is $\mathbf{E}\,\tau\,\mathbf{E}\,T$. Since the total increase in the first coordinate of the Markov chain per typical cycle is not smaller than $\lambda \sum\limits_{i=1}^{\tau} T_i$, then again, by the strong law of large numbers,

$$\liminf \frac{\widetilde{W}_n}{n} \ge \mathbf{E}\,\tau\,\mathbf{E}\,T(\lambda - L/\mathbf{E}\,T) > 0 \quad \text{a.s.}$$

Now we show that $W_n/n \to \infty$ a.s. too. For that, we refer to basic facts from renewal theory.

Denote again by τ_i the length of cycle i. The random variables $\{\tau_i\}$ are mutually independent for $i \ge 1$ and identically distributed (with finite mean) for $i \ge 2$. Let $S_m = \sum\limits_{i=1}^{m} \tau_i$ and, for $n \ge 1$,

$$\eta_n = \min\{m :\ S_m \ge n\} \qquad \text{and} \qquad \chi_n = S_{\eta_n} - n.$$

It is known that $\chi_n/n \to 0$ a.s. Since $W_{S_m} = \widetilde{W}_m$, we have

$$\frac{W_n}{n} \ge \frac{\widetilde{W}_{\eta_n} - \lambda\chi_n}{n} = \frac{\widetilde{W}_{\eta_n}}{\eta_n}\frac{\eta_n}{n} - \lambda\frac{\chi_n}{n} \to \infty \quad \text{a.s.}$$

Indeed, since $\eta_n \to \infty$ and $\eta_n/n \to 1/\mathbf{E}\,\tau > 0$ as $n \to \infty$, the divergence $\widetilde{W}_n/n \to \infty$ implies also that $\widetilde{W}_{\eta_n}/\eta_n \to \infty$ a.s. and then that $W_n/n \to \infty$ a.s.

4.3. Proof of Theorem 2 (a)

We can always find binary rational values of λ, a, b, and c and a collection of probabilities α for which $2p_- > p_1$ and $\lambda < L/\mathbf{E}\,T$. With these parameters, the Markov chain (W_n, Q_n) lives on a lattice. Since $\mathbf{P}(T = 1) > 0$ and $\mathbf{P}(T = 1,\, \tau = 1) > 0$, the Markov chain is aperiodic, and all its states commute. Therefore, the second part of the generalized Foster criterion is applicable, and distributions of (W_n, Q_n) converge to the stationary one in the total variation. Further, the random sequence $(W(t), Q(t))$ is regenerative, and the regenerative cycle length may take value 1 with positive probability. Therefore, we can apply Theorem 5, and statement (a) follows.

4.4. Proof of Theorem 2 (b)

Convergence $W_n + Q_n \to \infty$ a.s. follows from statement (b) of Theorem 1.

5. DISCUSSION AND OPEN PROBLEMS

5.1. Computing the Capacity of the Algorithm

Recall that, for our algorithm, $\mathbf{E}\,T \equiv \boldsymbol{T}(a, b, \alpha_0, \alpha_1, \alpha_2)$ is the mean number of slots in a session for given parameters a, b, α_0, α_1, and α_2. Also, the parameters $p_0 = p_0(a, b)$, $p_- = p_-(a, b)$, and $p_1 = p_1(a, b)$ depend on a and b.

By Theorem 2, the capacity is a solution to the following optimization problem: find the value of

$$C = \sup\{(a + b)/\boldsymbol{T}(a, b, \alpha_0, \alpha_1, \alpha_2)\},$$

where the supremum in taken over all values of parameters α_0, α_1, and α_2 from the interval $(0, 1)$ and over all nonnegative values of a and b such that $p_1(a, b)/(2p_-(a, b)) < 1$.

One can reduce this problem to a simpler one. Introduce a function of two parameters a and b

$$\varphi(a, b) = \max(a + b)/\boldsymbol{T}(a, b, \alpha_0, \alpha_1, \alpha_2),$$

where the supremum is taken over all α_0, α_1, and α_2 from $(0, 1)$.

The function $\boldsymbol{T}(a, b, \alpha_0, \alpha_1, \alpha_2)$ can be represented as

$$\boldsymbol{T}(a, b, \alpha_0, \alpha_1, \alpha_2) = \gamma(a + b) + E_0(a, b, \alpha_0) + E_1(a, b, \alpha_1) + E_2(a, b, \alpha_2), \tag{7}$$

where

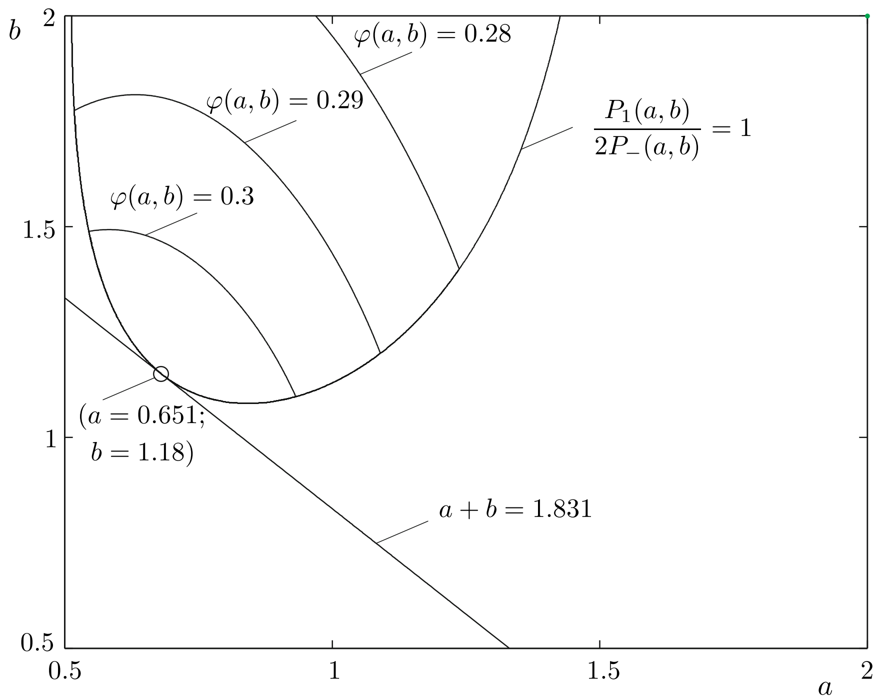
$$\gamma(a + b) = 2 - p_0,$$

$$E_0(a, b, \alpha_0) = \Big(\mathbf{E}\,\widetilde{X}_b + \sum_{m\ge 2} \frac{1}{r_{m,\alpha_0}} \mathbf{P}(\widetilde{X}_b \ge m)\Big)\pi_0 p_-,$$

$$E_1(a, b, \alpha_1) = \Big(\mathbf{E}\,Y^{(1)} + \mathbf{E}\,\widetilde{X}_b + \sum_{m\ge 2} \frac{1}{r_{m,\alpha_1}} \mathbf{P}(Y^{(1)} + \widetilde{X}_b \ge m)\Big)\pi_1 p_-,$$

$$E_2(a, b, \alpha_2) = \Big(\mathbf{E}\,Y^{(1)} + \mathbf{E}\,Y^{(2)} + \mathbf{E}\,\widetilde{X}_b + \sum_{m\ge 2} \frac{1}{r_{m,\alpha_2}} \mathbf{P}(Y^{(1)} + Y^{(2)} + \widetilde{X}_b \ge m)\Big)(1 - \pi_0 - \pi_1)p_-.$$

It follows from equation (7) that, in order to find values of $\varphi(a, b)$ for fixed a and b, it suffices to independently minimize the functions $E_0(a, b, \alpha_0)$, $E_1(a, b, \alpha_1)$, and $E_2(a, b, \alpha_2)$ by varying α_0, α_1, and α_2, respectively. Since all these functions are unimodal, the minimization can be done efficiently.



Stability region of the queue.

Given the function $\varphi(a,b)$, the problem of finding the capacity is reduced to solving a simpler problem: find the value

$$C = \sup \varphi(a,b),$$

where the supremum is over all nonnegative values of a and b such that $p_1(a,b)/(2p_-(a,b)) < 1$.

The latter problem can be solved numerically. The capacity value is 0.3098 (with an absolute error of 10^{-4}), and the optimal parameters are $a \approx 0.651$ and $b \approx 1.18$.

Computation of the function φ is very laborious. We guess that one can get a relatively good approximation for $C = \sup \varphi(a,b)$ as follows: consider pairs (a,b) such that $p_1(a,b)/(2p_-(a,b)) = 1$ and find a pair that minimizes the value $p_0 = (a+b)e^{-a-b}$, which is the probability of success in browsing the initial interval.

We cannot substantiate this conjecture as yet but would like to support it by the following illustration.

In the figure we consider the plane (a,b) and mark out the area where $p_1(a,b)/(2p_-(a,b)) < 1$. We call this area the *stability region of the queue*. In the interior of the region, we draw three level lines $\varphi(a,b) = 0.3$, $\varphi(a,b) = 0.29$, and $\varphi(a,b) = 0.28$. Note that the function $p_0(a,b) = (a+b)e^{-a-b}$ has linear level lines $b = D-a$, where D is a constant and the value of $p_0(a,b)$ at this level is De^{-D}. Any point in the interior satisfies $a+b > 1$, and the value of De^{-D} increases as the level lines move closer to the origin, within the stability region. The supremum of $p_0(a,b) = (a+b)e^{-a-b}$ in the stability region seems to be equal to the value of the function $p_0(a,b) = (a+b)e^{-a-b}$ at the contact point of the line $b = D-a$ and the boundary of the stability region. We have a numerical confidence for this.

5.2. Open Problems

The methods and computations that we used can be applied to all algorithms with delayed intervals introduced earlier. However, the greater the number of segmentations of the basic interval, the greater the number of parameters involved and, therefore, the more complex the optimization problem. For example, in [7] the basic interval is segmented into three intervals, say a, b, and c, and the function $\varphi(a, b, c)$ depends on three parameters. Again, the problem of finding the capacity can be reduced to finding the supremum of this function in the stability region. Numerical computations show that the capacity of this algorithm is 0.318; this refines a lower bound obtained in [7]. Again, the stability region is a convex set, and the capacity is attained at the point of contact of the level line $a + b + c = D$ and the stability region. As in the previous case, this point is the closest to the origin among all boundary points. Therefore, we may conjecture that for this algorithm it suffices to find a point a, b, and c on the boundary of the stability area which minimizes the probability of success of browsing the basic interval—this gives us the capacity.

For the class of algorithms with delayed intervals, numerical analysis shows that the capacity does not increase (as compared to the algorithm of [7]) with the increase of the number of segments and/or with complication of the way of taking delayed intervals from the queue. Therefore, we conjecture that the capacity of the whole class is approximately equal to 0.318.

We have to point out that the algorithm under consideration has the following drawback. Assume that the input intensity λ is smaller than the capacity C. By Theorem 2, with this intensity one can always find parameters A and B which make the system stable. Here $a = \lambda A$, $b = \lambda B$, and the point (a, b) belongs to the stability region (see the figure). However, for fixed A and B, one can always choose an intensity $\lambda_0 < \lambda < C$ so that the point $(\lambda_0 A, \lambda_0 B)$ is outside of the stability region. This means that if the algorithm is stable for a given value of input rate λ, it may become unstable not only with increase, but also with decrease of λ. The closer λ to the capacity, the narrower the interval of its values that make the algorithm stable. All algorithms with delayed intervals have this drawback. Development and analysis of algorithms that do not have such a shortcoming is an interesting and open problem.

The main part of this paper was completed during a visit of the first author to the Heriot-Watt University under the financial support of Spotnet project, EURONF Network of Excellence, Framework 7. The authors acknowledge the support of the program and the hospitality of the Heriot-Watt University.

APPENDIX

We recall a number of well-known results (the first two can be found, e.g., in survey [11]). The first statement is known as the *Moustafa–Foster–Tweedie criterion*, or, for short, *Foster's criterion*.

Theorem 3. *Let $\{Z_n\}$ be a time-homogeneous Markov chain taking values in a measurable state space $(\mathcal{Z}, \mathcal{B}_{\mathcal{Z}})$, and let $g\colon \mathcal{Z} \to [0, \infty)$ be a measurable function. If, for some positive C, g_0, and ε,*
(1) $\mathbf{E}(g(Z_1) \mid Z_0 = z) \le C$ *a.s. for all z such that $g(z) \le g_0$,*
(2) $\mathbf{E}(g(Z_1) \mid Z_0 = z) \le -\varepsilon$ *a.s. for all z such that $g(z) \ge g_0$,*
then the set $\{z : g(z) \le g_0\}$ is positive recurrent and, moreover, for any z, the random variable

$$\mu_z = \min\{n \ge 1 : g(Z_n) \le g_0 \mid Z_0 = z\}$$

has a finite mean and

$$\mathbf{E}\, \mu_z \le g(z)/\varepsilon.$$

If, in addition, the set $\{z : g(z) \le g_0\}$ is finite and the Markov chain is irreducible and aperiodic, then it is ergodic, i.e., it has a unique stationary distribution and, for any initial value, the distribution of Z_n converges to the stationary one in the total variation norm.

The following result can be regarded as the "generalized Foster criterion" (see also [11]). It is applicable in more general situations and, in particular, to increments of a Markov chain on random time intervals.

Theorem 4. *Let* $\{Z_n\}$ *be a time-homogeneous Markov chain with values in a measurable space* $(\mathcal{Z}, \mathcal{B}_{\mathcal{Z}})$, *and let* $g\colon \mathcal{Z} \to [0,\infty)$ *be a measurable function. Let* ν_z *be a sequence of random stopping times* (*this means that, for any* $z \in \mathcal{Z}$ *and for the Markov chain* $Z_0, Z_1, \ldots$ *with initial value* $Z_0 = x$, *there exists a positive and integer-valued random variable* ν_z *such that, for any* $n = 0, 1, \ldots,$ *the event* $\{\nu_z \le n\}$ *belongs to the* σ*-algebra generated by the random variables* $Z_0 = z, Z_1, \ldots, Z_n$).

If, for some positive C, c_1, c_2, g_0, *and* ε,

(1) $\mathbf{E}(g(Z_{\nu_z}) \mid Z_0 = z) \le C$ *a.s. for all* z *such that* $g(z) \le g_0$,
(2) $\mathbf{E}(g(Z_{\nu_z}) \mid Z_0 = z) \le -\varepsilon$ *a.s. for all* z *such that* $g(z) \ge g_0$,
(3) $\mathbf{E}\,\nu_z \le c_1 + c_2 g(z)$ *for all* $z \in \mathcal{Z}$,

then the set $\{z : g(z) \le g_0\}$ *is positive recurrent and, moreover, for any* z, *the random variable*

$$\mu_z = \min\{n \ge 1 : g(Z_n) \le g_0 \mid Z_0 = z\}$$

has a finite mean and

$$\mathbf{E}\,\mu_z \le g(z)/\varepsilon.$$

If, in addition, the set $\{z : g(z) \le g_0\}$ *is finite and the Markov chain is irreducible and aperiodic, then it is ergodic, i.e., it has a unique stationary distribution and, for any initial value, the distribution of* Z_n *converges to the stationary one in the total variation norm.*

Now we formulate a statement on convergence of regenerative processes in discrete time. A sequence $\{Z_n\}$ is *regenerative* if there are integer-valued random times $S_0 = 0 \le S_1 < S_2 < \ldots$ such that random elements $V_k = (S_k - S_{k-1}, Z_{S_{k-1}}, Z_{S_{k-1}+1}, \ldots, Z_{S_k - 1})$ are mutually independent for $k \ge 1$ and identically distributed for $k \ge 2$.

The following result may be found in many texts on renewal theory (see, e.g., [12]).

Theorem 5. *If a sequence* $\{Z_n\}$ *is regenerative and if, in addition,*

(a) $\mathbf{E}(S_2 - S_1) < \infty$, *and*
(b) *the greatest common divisor of all* j *such that* $\mathbf{P}(S_2 - S_1 = j) > 0$ *is equal to* 1,

then the distribution of Z_n *converges in the total variation norm, as* $n \to \infty$, *to a limiting distribution* Π *of the form*

$$\Pi(B) = \frac{\mathbf{E}\left(\sum\limits_{n=S_1}^{S_2-1} \boldsymbol{I}(Z_n \in B)\right)}{\mathbf{E}(S_2 - S_1)}.$$

REFERENCES

1. Tsybakov, B.S. and Mikhailov, V.A., Free Synchronous Packet Access in a Broadcast Channel with Feedback, *Probl. Peredachi Inf.*, 1978, vol. 14, no. 4, pp. 32–59 [*Probl. Inf. Trans.* (Engl. Transl.), 1978, vol. 14, no. 4, pp. 259–280].
2. Capetanakis, J.I., Tree Algorithms for Packet Broadcast Channels, *IEEE Trans. Inform. Theory*, 1979, vol. 25, no. 5, pp. 505–515.
3. Mehravari, N. and Berger, T., Poisson Multiple-Access Contention with Binary Feedback, *IEEE Trans. Inform. Theory*, 1984, vol. 30, no. 5, pp. 745–751.
4. Tsybakov, B.S. and Beloyarov, A.N., Random Multiple Access in a Channel with Binary "Success–No Success" Feedback, *Probl. Peredachi Inf.*, 1990, vol. 26, no. 3, pp. 67–82 [*Probl. Inf. Trans.* (Engl. Transl.), 1990, vol. 26, no. 3, pp. 245–260].
5. Tsybakov, B.S. and Beloyarov, A.N., Random Multiple Access in Binary Feedback Channel, *Probl. Peredachi Inf.*, 1990, vol. 26, no. 4, pp. 83–98 [*Probl. Inf. Trans.* (Engl. Transl.), 1990, vol. 26, no. 4, pp. 354–366].

6. Paris, B.-P. and Aazhang, B., Near-Optimum Control of Multiple-Access Collision Channels, *IEEE Trans. Commun.*, 1992, vol. 40, no. 8, pp. 1298–1309.

7. Malkov, A. and Turlikov, A., Random Multiple Access Protocols for Communication Systems with "Success-Failure" Feedback, in *Proc. 1995 IEEE Workshop on Information Theory, Multiple Access, and Queueing, St. Louis, USA*, 1995, p. 39.

8. Aldous, D.J., Ultimate Instability of Exponential Back-off Protocol for Acknowledgment Based Transmission Control of Random Access Communication Channels, *IEEE Trans. Inform. Theory*, 1987, vol. 33, no. 2, pp. 219–223.

9. Tsybakov, B.S. and Mikhailov, V.A., Random Multiple Packet Access: Part-and-Try Algorithm, *Probl. Peredachi Inf.*, 1980, vol. 16, no. 4, pp. 65–79 [*Probl. Inf. Trans.* (Engl. Transl.), 1980, vol. 16, no. 4, pp. 305–317].

10. Borovkov, A.A., *Teoriya veroyatnostei* (Probability Theory), Moscow: Editorial URSS, 1999, 3rd ed.

11. Foss, S. and Konstantopoulos, T., An Overview of Some Stochastic Stability Methods, *J. Oper. Res. Soc. Japan*, 2004, vol. 47, no. 4, pp. 275–303.

12. Gut, A., *Stopped Random Walks. Limit Theorems and Applications*, New York: Springer, 2009.