ON A FUNCTIONAL EQUATION ARISING IN THE ANALYSIS OF A PROTOCOL FOR A MULTI-ACCESS BROADCAST CHANNEL

GUY FAYOLLE,* *INRIA*
PHILIPPE FLAJOLET,* *INRIA* AND
MICHA HOFRI,** *Technion–Israel Institute of Technology*

Abstract

We analyse a stack protocol of the Capetanakis–Tsybakov–Mikhailov type for resolving collisions in a random multiple-access channel. We obtain a functional equation for the generating function of the expected collision resolution interval (CRI) durations, which is non-local with a non-commutative iteration semigroup. Using Mellin transform techniques and geometric properties of the iteration semigroup we show that for arrival rates smaller than a fixed threshold, the mean CRI duration for n initial colliders is asymptotically proportional to n. Ergodicity conditions are also demonstrated.

ASYMPTOTIC ANALYSIS; FUNCTIONAL EQUATION; MELLIN TRANSFORM; PROTOCOL; RANDOM ACCESS; TREE ALGORITHM

1. Introduction

Since the mid 1970s many papers have been published on the analysis of the performance of single-channel message-switching communications networks, springing from the ALOHA network concept. The application to these networks of standard techniques used for the evaluation of performance of computing systems has proved to be either difficult or inadequate, because the processes governing the behaviour of these networks do not behave like those familiar from queueing theory. As will be observed below, multitype branching processes are more often what one finds, and the duration of such a process is the underlying quantity we wish to compute.

We briefly recall the salient principles of the original ALOHA network, as they apply to the analysis presented in this paper.

(a) A single error-free channel is shared among many users which transmit messages of constant length (packets). Time is slotted, and may be considered discrete. Users are synchronized with respect to the slots.

Each slot is equal to the time required to transmit a packet, and packets are transmitted at the beginning of slots only.

(b) Each transmission is receivable by every user. Thus, when two or more users transmit simultaneously, packets are said to 'collide' (interfere) and none is received correctly: these collisions are treated as transmission errors and each user must strive to retransmit its colliding packet until it is correctly received. The users all employ the same algorithm for this purpose, and have to resolve the contention without the benefit of any other source of information on other users' activity save the common channel.

The collision resolution algorithm is clearly the Gordian knot of the behavior of the transmission process, affecting, among other things, the delay experienced by messages until they are sucessfully transmitted, the buffering requirements at the nodes that maintain the broadcast activity, and perhaps most significantly, the 'efficiency' of this multiple-access scheme, in terms of the maximum traffic rate it will allow before destabilizing (unless more involved measures, such as slot reservation policies, are used to reduce its criticality).

Many protocols for collision resolution have been suggested and investigated. Most are based on the original ALOHA concept: each user with a 'colliding' packet will repeatedly retransmit, each time with certain probability until it hits a free slot, and thus succeeds. The main drawback of these protocols is that, left to their own devices, the nodes congest the channel, and in the absence of additional controls they are unstable (Fayolle (1975)).

Our analysis concerns only the maximum traffic rate that the channel can carry under a particular channel access scheme and collision resolution algorithm (CRA). These are defined in the next section. It turns out (this is shown in Section 4, and see also Fayolle and Hofri (1982)) that the only quantity one needs in order to evaluate the limiting rate is the expected duration of the collision resolution interval (CRI) which is defined in terms of the specific CRA. These expected durations are shown to satisfy a difference equation, which we convert to a non-local functional equation for their exponential generating function (Section 2.1). The analysis of this equation is the main thrust of this paper:

– we show a formal series solution of the equation (Section 2);
– the solution uses two non-commutative transformations in the complex plane.

The investigation of the group generated by these transformations allows us to obtain an asymptotic estimate of the expected CRI durations (Section 3).

The desired channel capacity is obtained in Section 5.

1.1. *The CTM algorithm.* In this paper we consider a variety of the Capetanakis–Tsybakov–Mikhailov (CTM) CRA. We shall show that this CRA induces stable behavior of the channel as long as the rate of arrival of new packets is below a certain bound. The particular variety is described in Fayolle and Hofri (1982) as channel protocol #3 there. It does not require external controls to stabilize, and allows users to transmit a packet as soon as it is generated, regardless of the activity of the other transmitters. Thus users do not have to monitor the channel continuously, but only when they wish to transmit. Such users are called 'active' users.

Broadly speaking the CTM CRA is a 'divide and conquer' algorithm.

Specification of the CTM CRA with continuous input. Points (a) and (b) of the introduction hold, and we also make the following assumptions:

(c) Each user monitoring the channel knows, by the end of the slot, if that slot produced a collision or not.

(d) Each active user maintains a conceptual stack. When an inactive user becomes active, it enters level 0 in the stack. It will transmit at the nearest slot, and will always do so when at stack level 0.

At each slot end, a user determines its position in the stack according to the following procedure (identical to all users, who are unable, however to communicate their stack state).

Case 1: *following a non-collision slot.* A user in stack level 0 (there can be *at most* one user, system-wide) becomes inactive, and users at any other stack level decrease their stack level by 1.

Case 2: *following a collision slot.* All users at stack level i, $i \geqq 1$ change to level $i + 1$. The users at level 0 are split into two groups; one group remains at level 0, while the members of the other push themselves into level 1. Note that no such user is aware then of how many users there are at each level. This partition can be made on the basis of a random variable, such as the flipping of a coin, on the basis of the time when the user became active, etc.

The CRI duration, that is the time it takes under this algorithm, to dispose of a group of n colliding users, is denoted by L_n. This includes the slot of the initial collision and subsequent slots, until all active users who were at level i in their respective stacks, return there. Or, if it was empty initially (and known to be such to any listener), it returns to this marked state. We formulate the process in this manner, rather than saying that L_n is the time until the nth successful transmission after the collision because:

(a) there will generally be newly arriving packets that have to be cleared before the CRI is over, and

(b) even when all the colliding packets have actually been successfully transmitted (as well as subsequent arrivals) there may be empty levels in the other active users' stacks, which are not known to be empty and must be disposed of by silent slots. Brief reflection will show that those users cannot know whether some users occupy the corresponding levels in their stacks or not.

1.2. *The mathematical model.* To the above description we add the following assumptions:

(i) The numbers of new packets generated in each slot (or the numbers of new active users) form a sequence of i.i.d. random variables, denoted by $\{X_i, i \geq 1\}$, which follow the Poisson distribution with parameter λ.

(ii) The decisions described in step (d) above, taken by each 'colliding' user at the end of a collision slot, form a sequence of independent Bernoulli trials, with p being the probability of staying at level 0. This probability is assumed to be uniform for all users and time homogeneous. For definiteness and with no loss of generality we assume $1 > p \geq q$ $(= 1 - p)$.

(iii) It will be useful to define the 'degenerate' CRIs L_0 and L_1 to be of size 1.

2. The basic equations

2.1. *The functional equation for the generating function of mean CRI durations.* The definition of the channel protocol provides the following recursive relation for the random variables L_n:

(2.1) $$L_0 = L_1 = 1$$

(2.2) $$L_n = 1 + L_{I+X} + L_{n-I+Y}, \quad n \geq 2$$

where I, the number of messages immediately retransmitted, follows the binomial distribution $B(n, p)$, X is the number of new arrivals in the collision slot and Y is the number of new arrivals in the slot following L_{I+X}.

The L_n's are defined on the canonical space $B^{N^+} \times P$, where B and P are the underlying probability spaces associated respectively with each Bernoulli trial and with a Poisson process of intensity λ. (N^+ is the set of natural integers.)

Defining $\alpha_n = E(L_n)$ and taking expectation in (2.1) and (2.2) yields

(2.3) $$\begin{cases} \alpha_0 = \alpha_1 = 1 \\ \alpha_n = 1 + \sum_{k \geq 0} \sum_{l \geq 0} \sum_{i=0}^{n} \binom{n}{i} p^i q^{n-i} \exp(-2\lambda) \frac{\lambda^k \lambda^l}{k! \, l!} (\alpha_{i+k} + \alpha_{n-i+l}). \end{cases}$$

Relations of type (2.3) have been studied in Tsybakov and Vvedenskaya

(1980), for $p = q = \frac{1}{2}$ only, although it was for the case of more general input. The arguments were different and, as will appear later, the mathematical problem is *completely* altered when $p \neq q$. Introducing

(2.4)
$$
\begin{cases}
\alpha(z) \overset{\text{def}}{=} \sum_{n=0}^{\infty} \alpha_n \frac{z^n}{n!} \\
\Psi(z) \overset{\text{def}}{=} \exp(-z)\alpha(z),
\end{cases}
$$

we get from (2.3)

(2.5) $\quad \Psi(z) - \Psi(\lambda + pz) - \Psi(\lambda + qz) = 1 - \exp(-z)[2\Psi(\lambda)(1 + z) + z\Psi'(\lambda)]$

where

$$
\Psi'(\lambda) = \frac{d\Psi(z)}{dz}\bigg|_{z=\lambda}.
$$

Evaluating (2.5) for $z = \lambda/p$ and $z = \lambda/q$, and eliminating $\Psi(2\lambda)$, one obtains $\Psi'(\lambda)$ in terms of $\Psi(\lambda)$:

(2.6)
$$
\Psi'(\lambda) = 2(K - 1)\Psi(\lambda),
$$

where

(2.7)
$$
K = \frac{\exp(-\lambda/p) - \exp(-\lambda/q)}{\frac{\lambda}{q}\exp(-\lambda/q) - \frac{\lambda}{p}\exp(-\lambda/p)}.
$$

Note. While this reduction is convenient and simplifies some of the computations below, it is inessential from a mathematical point of view (see Fayolle and Hofri (1982)).

Finally, (2.5) can be written as follows.

Lemma 1. The generating function $\Psi(z)$ of the α_n defined in (2.4) satisfies the functional equation

(2.8) $\quad \Psi(z) - \Psi(\lambda + pz) - \Psi(\lambda + qz) = 1 - 2\Psi(\lambda)\exp(-z)(1 + Kz).$

Equations (2.3) and (2.4) provide $\Psi(0) = 1$, $\Psi'(0) = 0$.

2.2. *An iteration scheme for the functional equation.* In this section we develop an iterative scheme for solving the basic functional equation (2.8) with its boundary conditions, and for expressing the Taylor coefficients of

(2.9)
$$
\alpha(z) = \exp(z)\Psi(z).
$$

This requires the introduction of a *non-commutative iteration semigroup* with

properties that are also of use in the later asymptotic analysis. The solution appears as sums indexed on this iteration semigroup.

Define $\sigma_1(z) = \lambda + pz$ and $\sigma_2(z) = \lambda + qz$, and introduce the following notation:

(i) Let H be the semigroup of linear substitutions generated by σ_1 and σ_2, where the semigroup operation is the composition of functions. The identity of H, denoted by ε, is thus the function $\varepsilon(z) = z$ for all $z \in C$ (the complex plane).

(ii) Any member of H can be written in the form

$$\sigma = \sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_n} \quad \text{where} \quad n \geq 0 \quad \text{and} \quad i_j \in \{1, 2\}.$$

Define

$$|\sigma|_1 = \text{card } \{j \,|\, i_j = 1\}; \qquad |\sigma|_2 = \text{card } \{j \,|\, i_j = 2\};$$
$$|\sigma| = |\sigma|_1 + |\sigma|_2;$$

this last quantity is called the *length* of the substitution σ.

(iii) The subset of H formed with substitutions of length n is denoted by H_n, so that

$$H_n = \{\sigma \in H: |\sigma| = n\}.$$

The semigroup H satisfies the obvious decompositions:

(2.10)
$$\begin{cases} H = \{\varepsilon\} \cup \sigma_1 H \cup \sigma_2 H \\ H = \{\varepsilon\} \cup H\sigma_1 \cup H\sigma_2 \end{cases}$$

and correspondingly for H_n, $n \geq 1$:

(2.11)
$$\begin{cases} H_n = \sigma_1 H_{n-1} \cup \sigma_2 H_{n-1} \\ H_n = H_{n-1}\sigma_1 \cup H_{n-1}\sigma_2. \end{cases}$$

(iv) For α, β complex numbers, define

$$(\alpha; \beta)^\sigma = \alpha^{|\sigma|_1}\beta^{|\sigma|_2}.$$

We can now state the following result.

Lemma 2. If α, β are complex numbers satisfying the *contraction condition*

(2.12) $$|\alpha| + |\beta| < 1$$

and if $t(z)$ is an entire function, then the functional equation

(2.13) $$f(z) - \alpha f(\sigma_1(z)) - \beta f(\sigma_2(z)) = t(z)$$

has a unique entire solution given by

(2.14) $$f(z) = \sum_{\sigma \in H} (\alpha; \beta)^\sigma t(\sigma(z)).$$

Proof.

(i) *Existence*: For z in C and any $\sigma \in H$, we have

$$|\sigma(z)| \leqq \max\left(|z|, \frac{\lambda}{q}\right).$$

The proof is straightforward by induction on $|\sigma|$. Let

$$\mu(z) = \sup\left\{|t(x)| : x \in C, |x| \leqq \max\left(|z|, \frac{\lambda}{q}\right)\right\}.$$

Then the sum in (2.14) is absolutely convergent and its modulus is bounded by

$$\mu(z) \sum_{\sigma \in H} (|\alpha|; |\beta|)^\sigma = \frac{\mu(z)}{1 - |\alpha| - |\beta|}.$$

Thus for z in any bounded domain, $f(z)$ as given by (2.14) is a uniformly convergent sum of analytic functions and therefore is analytic. Using decomposition (2.10), we have

$$f(z) = t(z) + \sum_{\sigma \in H\sigma_1} (\alpha; \beta)^\sigma t(\sigma(z)) + \sum_{\sigma \in H\sigma_2} (\alpha; \beta)^\sigma t(\sigma(z))$$

$$= t(z) + \alpha \sum_{\tau \in H} (\alpha; \beta)^\tau t(\tau\sigma_1(z)) + \beta \sum_{\tau \in H} (\alpha; \beta)^\tau t(\tau\sigma_2(z))$$

$$= t(z) + \alpha f(\sigma_1(z)) + \beta f(\sigma_2(z)),$$

so that $f(z)$ indeed satisfies Equation (2.13).

(ii) *Uniqueness*: To discuss the solutions of (2.13), we have only to consider domains D such that

$$\sigma_1 D \subset D \qquad \sigma_2 D \subset D.$$

Such domains necessarily contain the real interval $[\lambda/p; \lambda/q]$, since for any z, the set $H[z] = \{\sigma(z) \mid \sigma \in H\}$ admits this interval as the set of its accumulation points (see Figure 1). Let f_1 and f_2 be two analytic solutions of
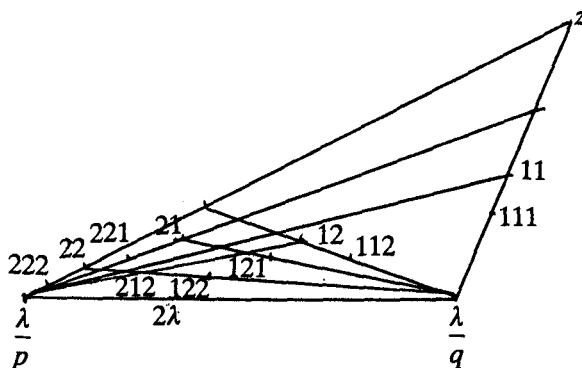

Figure 1. The successive transforms of a point z, in the case $p = \frac{2}{3}$, $q = \frac{1}{3}$. The point $\sigma_{i_1}\sigma_{i_2} \cdots \sigma_{i_m}(z)$ is labelled $i_1 i_2 \cdots i_m$

Equation (2.13) in such a domain D. By iteration of the functional relation, f_1 and f_2 are necessarily entire. Setting $\delta(z) = f_1(z) - f_2(z)$, we see that $\delta(z)$ satisfies the relation

$$(2.15) \qquad \delta(z) - \alpha\delta(\sigma_1(z)) - \beta\delta(\sigma_2(z)) = 0,$$

whence by iteration of (2.15) for any $n > 0$,

$$(2.16) \qquad \delta(z) = \sum_{\sigma \in H_n} (\alpha; \beta)^\sigma \delta(\sigma(z)).$$

We prove that $\delta(z)$ is of bounded modulus in the whole of the complex plane, hence constant by Liouville's theorem: let

$$M = \max\left\{ |\delta(z)| : |z| \leq \frac{2\lambda}{q} \right\}.$$

For any $z \in C$, there exists some $n \in N$ such that

$$|\sigma(z)| \leq \frac{2\lambda}{q} \quad \text{for all} \quad \sigma \in H_n.$$

Thus with (2.16) for this value of n

$$|\delta(z)| \leq M \sum_{\sigma \in H_n} (|\alpha|; |\beta|)^\sigma = M(|\alpha| + |\beta|)^n < M,$$

which therefore establishes the uniform boundedness of δ. Obviously we can only have $\delta(z) = 0$ which proves the two solutions f_1 and f_2 coincide.

In what follows we encounter equations of form similar to (2.13) where the coefficients α and β do not satisfy the 'contraction condition'. However, we can state the following particular case.

Corollary. The equation

$$(2.17) \qquad f(z) - f(\sigma_1(z)) - f(\sigma_2(z)) = t(z)$$

has the solution

$$(2.18) \quad f(z) = S(t(z); z) = f(0) + zf'(0) + \sum_{\sigma \in H} [t(\sigma(z)) - t(\sigma(0))$$
$$- z(p; q)^\sigma t'(\sigma(0))].$$

Proof. Equation (2.17) reduces to (2.13) by double differentiation. Integrating the solution (2.14), which now involves $t''(z)$, term by term twice produces (2.18).

The equation (2.17) has to satisfy certain consistency relations, such as $t(\lambda/q) = t(\lambda/p)$. In all the applications we have encountered, $t(z)$ contains unknown constants and this relation is instrumental in determining them.

Theorem 1. Define

$$(2.19) \quad D(\lambda) = \sum_{\sigma \in H} \exp\left(-\sigma(0)\right)\{(1 + K\sigma(\lambda)) \exp\left(-(p;q)^{\sigma}\lambda\right)$$

$$- 1 - K\sigma(0) - \lambda(p;q)^{\sigma}[K - 1 - K\sigma(0)]\}.$$

A necessary and sufficient condition for Equation (2.8) to have an entire solution is that

$$D(\lambda) \neq -\tfrac{1}{2}, \quad \text{and} \quad K \neq \infty.$$

If this condition is satisfied, one has

$$(2.20) \qquad \alpha_n = 1 - \frac{1}{1 + 2D(\lambda)}[T_n + KU_n + KnV_n]$$

where

$$T_n = \sum_{\sigma \in H} \exp\left(-\sigma(0)\right)[(1 - (p;q)^{\sigma})^n - 1 + n(p;q)^{\sigma}],$$

$$(2.21) \qquad U_n = \sum_{\sigma \in H} \exp\left(-\sigma(0)\right)\sigma(0)[(1 - (p;q)^{\sigma})^n - 1 + n(p;q)^{\sigma}],$$

$$(2.22) \qquad V_n = \sum_{\sigma \in H} \exp\left(-\sigma(0)\right)(p;q)^{\sigma}[(1 - (p;q)^{\sigma})^{n-1} - 1].$$

Proof. Equation (2.19) follows from Lemma 1 and the corollary to Lemma 2 upon substitution of $z = \lambda$:

$$(2.23) \quad \psi(z) = 1 - 2\psi(\lambda)\sum_{\sigma} \{(1 + K\upsilon(z)) \exp\left(-\sigma(z)\right)$$

$$- (1 + K\sigma(0)) \exp\left(-\sigma(0)\right) - z(p;q)^{\sigma} \exp\left(-\sigma(0)\right)[K - 1 - K\sigma(0)]\}$$

and observing $\sigma(z) = \sigma(0) + (p;q)^{\sigma}z$.

This provides the linear equation for $\psi(\lambda)$

$$(2.23a) \qquad\qquad \psi(\lambda) = 1 - 2\psi(\lambda)D(\lambda),$$

which has a solution iff $D(\lambda) \neq -\tfrac{1}{2}$.

When the condition is satisfied, $\psi(\lambda) = 1/1 + 2D(\lambda)$ and substituting this in (2.23) and equating the Taylor coefficients on both sides of (2.23) yield the last part of the theorem.

Note. For computational purposes it is preferable to compute the coefficients of the differentiated series. Then we define for $n \geq 2$

$$(2.24) \qquad \begin{cases} g_n = (-1)^n \sum_{\sigma} \exp(-\sigma(0))(p^n;q^n)^{\sigma} \\[2ex] k_n = (-1)^n \sum_{\sigma} \exp\left(-\sigma(0)\right)\sigma(0)(p^n;q^n)^{\sigma} \end{cases}$$

and obtain

$$(2.25) \qquad D(\lambda) = \sum_{n \geq 2} [(1 - Kn)g_n + Kk_n] \frac{\lambda^n}{n!}.$$

3. Asymptotic analysis

Having obtained an explicit result for $\psi(z)$, and through it for the α_n, we note that these do not provide an analytically usable characterization of the behaviour of α_n as n increases. To obtain the latter we have to resort to different means.

3.1. *The exponential approximation.* The first step is to use exponential approximations for the coefficients T_n, U_n, V_n (cf. Knuth (1973), p. 131, for a similar situation). Replacing $(1 - a)^n$ by $\exp(-an)$ we introduce the quantities

$$(3.1) \qquad \begin{cases} t(x) = \sum_{\sigma \in H} \exp(-\sigma(0))[\exp(-a(\sigma)x) - 1 + a(\sigma)x], \\[2mm] u(x) = \sum_{\sigma \in H} \exp(-\sigma(0))\sigma(0)[\exp(-a(\sigma)x) - 1 + a(\sigma)x], \\[2mm] v(x) = \sum_{\sigma \in H} \exp(-\sigma(0))a(\sigma)[\exp(-a(\sigma)x) - 1] \end{cases}$$

where we have set $a(\sigma) = (p; q)^\sigma$. We prove the following result.

Lemma 3. The collision resolution times satisfy the relation

$$(3.2) \qquad \alpha_n = \frac{2}{1 + 2D(\lambda)}[t(n) + Ku(n) + Knv(n)] + O(n^{1-\eta})$$

for any positive η, such that

$$(3.3) \qquad \eta < \frac{\log(p^2 + q^2)}{\log q} \qquad (p \geq q).$$

Proof. We use the expression of Lemma 1 and show that T_n, U_n, nV_n are approximated for large n by $t(n)$, $u(n)$, $nv(n)$, respectively, to the stated tolerance. We prove the result only for T_n, the other cases being identical. From the definition

$$T_n - t(n) = \sum_{\sigma \in H} \exp(-\sigma(0))[(1 - a(\sigma))^n - \exp(-a(\sigma)n)] \overset{\text{def}}{=} \sum_{\sigma \in H} \delta(\sigma).$$

To evaluate this sum, we split it and define

$$S_1(v) = \sum_{\{\sigma: |\sigma| \leq v\}} \delta(\sigma); \qquad S_2(v) = \sum_{\{\sigma: |\sigma| > v\}} \delta(\sigma).$$

Since for a positive a, $(1 - a) < \exp(-a)$, and since for $|\sigma| \leq v$, $a(\sigma) \geq q^v$, we find:

$$(3.4) \qquad S_1(v) = O(2^v \exp(-nq^v)).$$

Choosing v such that $p^v < n^{-\frac{1}{2}}$ ensures that for all $\sigma : |\sigma| > v$,

$$(1 - a(\sigma))^n - \exp(-a(\sigma)n) = \exp(-a(\sigma)n)O(na^2(\sigma)),$$

uniformly in σ. (Easy to see by writing $(1 - a(\sigma))^n = \exp(n \log(1 - a(\sigma)))$ and developing.) Noting that $\sum_{\sigma \in H_i} a^2(\sigma) = (p^2 + q^2)^i$, we obtain

$$(3.5) \qquad S_2(v) = O\left(n \sum_{|\sigma| > v} a^2(\sigma)\right) = O(n(p^2 + q^2)^v).$$

We can now select

$$(3.6) \qquad v = \log_q \left(\frac{\log^2 n}{n}\right),$$

to ensure that $S_1(v)$ is exponentially small by (3.4):

$$S_1(v) = O(n^{c_1} \exp(-\log^2 n)) \quad \text{for some finite } c_1.$$

The above condition on p^v is also satisfied, so that by (3.5)

$$(3.7) \qquad S_2(v) = O(n(p^2 + q^2)^v) = O((p^2 + q^2)^{2\log_q \log n} n^{1 - \log(p^2 + q^2)/\log q})$$

which satisfies the lemma requirement.

3.2. *Mellin transform techniques.* We now propose to study the asymptotic behaviour of α_n as n gets large. An outline of the rest of this section follows: Lemma 3 has reduced the problem to that of estimating the asymptotic equivalents $t(x)$, $u(x)$ and $v(x)$ as $x \to \infty$. We first compute the Mellin transforms of t, u, v, which have factored forms in which both the gamma function and certain Dirichlet series related to the iteration group appear. We then use the classical correspondence between the singularities of Mellin transforms in the right half-plane and terms in the asymptotic expansion of the original functions for large values of the arguments (a fact which devolves from the inversion theorem for Mellin transforms).

Locating singularities of the Dirichlet series, and in particular estimating the dominant terms in their asymptotic expansions around their poles, requires some deeper properties of the iteration group H. Once this is done, we can conclude with the asymptotic estimates of $t(x)$, $u(x)$, $v(x)$. The discussion distinguishes two cases based on certain arithmetical properties of the probabilities p, q and we can conclude finally that α_n has a linear growth in n.

We start by introducing the two Dirichlet series:

(3.8)
$$\theta(s) = \sum_{\sigma \in H} \exp(-\sigma(0))(p^s; q^s)^\sigma$$

(3.9)
$$\xi(s) = \sum_{\sigma \in H} \exp(-\sigma(0))\sigma(0)(p^s; q^s)^\sigma,$$

in which the sums are absolutely convergent for $\mathrm{Re}(s) > 1$. We also consider the function

(3.10)
$$\beta(x) = t(x) + Ku(x) + Kxv(x),$$

which appears in the approximation of α_n, and which is defined for all $x \geqq 0$.

The Mellin transform of a function $f(x)$ defined on R^+ is denoted by $f^*(s)$ or $M[f(x); s]$ and is given by

(3.11)
$$f^*(s) = \int_0^\infty f(x)x^{s-1}\, dx$$

and satisfies the important functional property:

(3.12)
$$M[f(ax); s] = a^{-s}f^*(s)$$

for any positive a. (See Doetsch (1955), Davies (1978) for basic properties and definitions of the Mellin transform.) We have the following result.

Lemma 4. The Mellin transform of the function $\beta(x)$ of (3.10) is

$$\beta^*(s) = \theta(-s)[\Gamma(s) + K\Gamma(s+1)] + K\xi(-s)\Gamma(s)$$

and the integral (3.11) defining β^* is absolutely convergent for $s : -2 < \mathrm{Re}(s) < -1$.

Proof. Applying (3.12) repeatedly we see that a function of the form

(3.13)
$$F(x) = \sum_{\sigma \in H} c_\sigma f(d_\sigma x)$$

has a transform of the form

(3.14)
$$F^*(s) = f^*(s) \sum_{\sigma \in H} c_\sigma (d_\sigma)^{-s}$$

which is valid provided s is in the intersection of the domain of absolute convergence of $f^*(s)$ and of the domain of absolute convergence of the sum that appears in (3.14).

We use the classical transforms

(3.15)
$$\begin{cases} \text{(a)} \quad \int_0^\infty (\exp(-x) - 1 + x)x^{s-1}\, dx = \Gamma(s) \quad \text{for} \quad s : -2 < \mathrm{Re}(s) < -1 \\[2mm] \text{(b)} \quad \int_0^\infty (\exp(-x) - 1)x^{s-1}\, dx = \Gamma(s) \quad \text{for} \quad s : -1 < \mathrm{Re}(s) < 0. \end{cases}$$

Applying (3.14) to the sum defining $t(x)$, we find that

$$t^*(s) = \Gamma(s) \sum_{\sigma \in H} \exp(-\sigma(0))\{(p;q)^\sigma\}^{-s}$$

where the condition on s from (3.15a) is $-2 < \text{Re}(s) < -1$ and the condition on s from the convergence requirement of the sum is $\text{Re}(s) < -1$. The transforms of u and v are dealt with in a similar way, whence the result by linearity of the transform.

By the inversion theorem, $\beta(x)$ is expressible in terms of $\beta^*(s)$ as the integral

(3.16) $$\beta(x) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \beta^*(s)x^{-s}\, ds, \quad \text{for any } c \text{ in } (-2, -1).$$

3.3. *Analytic properties of some Dirichlet series.* We propose to evaluate (3.16) by shifting the line of integration to the right, taking residues of the integrand into account. The first residues will give the dominant terms in the asymptotic expansion of $\beta(x)$ as $x \to \infty$. Evaluation of these terms requires however some more detailed analytic information on $\theta(s)$ and $\xi(s)$.

To treat $\theta(s)$ and $\xi(s)$ simultaneously, we consider Dirichlet series given by a sum of the form

(3.17) $$\omega(s) = \sum_{\sigma \in H} r(\sigma(0))(p^s; q^s)^\sigma$$

where $r(u)$ is any continuously differentiable function on $[0, \lambda/q]$. Strictly speaking, we should also require r not to vanish identically on $(\lambda/p, \lambda/q)$. We have the following result.

Theorem 2. For any function $r(.)$ continuously differentiable on $[0, \lambda/q]$ the function $\omega(s)$ in (3.17) is meromorphic for $\text{Re}(s) > 0$. It has a simple pole at $s = 1$, and around that point admits the expansion

(3.18)
$$\omega(s) = \frac{1}{(s-1)} \frac{1}{h(p, q)\left(\dfrac{\lambda}{q} - \dfrac{\lambda}{p}\right)} \int_{\lambda/p}^{\lambda/q} r(u)\, du + \int_0^{\lambda/q} r'(u)w(u)\, du$$
$$+ \frac{p\log^2 p + q\log^2 q}{2\left(\dfrac{\lambda}{q} - \dfrac{\lambda}{p}\right)h^2(p, q)} \int_{\lambda/p}^{\lambda/q} r(u)\, du + o(s-1)$$

where $h(p, q)$ is the entropy function

$$h(p, q) = p\log p^{-1} + q\log q^{-1}$$

and $w(u)$ is a weight function independent of $r(u)$, to be computed later: $w(u) = w_1(u) + w_2(u)$, with w_1, w_2 defined in Equations (3.42), (3.43).

Briefly, the idea is first to find a function that has the same behaviour at $s = 1$, to first order, as $\omega(s)$. This function is called $D(s)$ below.

The proof is long and of a computational character. We shall break it down to a number of lemmas. We first prove that the set of singularities of $\omega(s)$ in the half-plane $\mathrm{Re}(s) > 0$, is contained in the set of singularities of $(1 - p^s - q^s)^{-1}$. From this we have the following result.

Lemma 5. The function

$$\pi(s) = \omega(s)(1 - p^s - q^s)$$

is analytic for $\mathrm{Re}(s) > 0$ and uniformly bounded in any half-plane $\mathrm{Re}(s) \geqq \eta > 0$.

Proof. For $\mathrm{Re}(s) > 1$, the sum expressing $\omega(s)$ is absolutely convergent, so that we can regroup terms in the expression of $\pi(s)$. We have

$$\pi(s) = \sum_{\sigma \in H} r(\sigma(0))(p^s; q^s)^\sigma - \sum_{\sigma \in H} r(\sigma(0))p^s(p^s; q^s)^\sigma$$

(3.19)

$$- \sum_{\sigma \in H} r(\sigma(0))q^s(p^s; q^s)^\sigma.$$

We first transform the second and third sums in (3.19):

(3.20) $$\sum_{\sigma \in H} r(\sigma(0))p^s(p^s; q^s)^\sigma = \sum_{\tau \in H\sigma_1} r(\tau\sigma_1^{-1}(0))(p^s; q^s)^\tau$$

(3.21) $$\sum_{\sigma \in H} r(\sigma(0))q^s(p^s; q^s) = \sum_{\tau \in H\sigma_2} r(\tau\sigma_2^{-1}(0))(p^s; q^s)^\tau;$$

we then use the decomposition $H = \{\varepsilon\} \cup H\sigma_1 \cup H\sigma_2$ in the first sum in (3.19), and group terms with those of (3.20), (3.21), so that

$$\pi(s) = r(0) - \sum_{\tau \in H\sigma_1} (r(\tau\sigma_1^{-1}(0)) - r(\tau(0)))(p^s; q^s)^\tau$$

(3.22)

$$- \sum_{\tau \in H\sigma_2} (r(\tau\sigma_2^{-1}(0)) - r(\tau(0)))(p^s; q^s)^\tau.$$

Equation (3.22) is valid for $\mathrm{Re}(s) > 1$; now using the observations

$$\tau(a) - \tau(b) = (p; q)^\tau(a - b)$$

$$r(\tau(a)) - r(\tau(b)) = O((p; q)^\tau),$$

uniformly in $|\tau|$ for fixed a and b, we see that the sums in (3.22) are

$$O\left(\sum_{\tau \in H} |p^{s+1}; q^{s+1})^\tau|\right) = O([1 - |p^{s+1}| - |q^{s+1}|]^{-1})$$

and therefore converge for $\mathrm{Re}(s) > 0$. Thus (3.22) provides the analytic continuation of $\pi(s)$ to the left of $s = 1$ and the lemma is established.

The next stage of the proof of Theorem 2 is to obtain the main terms in the expansion of $\omega(s)$ around $s = 1$. To that purpose, we decompose the sum expressing $\omega(s)$:

$$\omega(s) = \sum_{n \geq 0} \sum_{\sigma \in H_n} r(\sigma(0))(p^s; q^s)^\sigma.$$

Define the sequence of functions over the interval $[0, \lambda/q]$:

$$(3.23) \quad \phi_n^{(s)}(u) = \sum_{\substack{\sigma(0) < u \\ \sigma \in H_n}} (\hat{p}^s; \hat{q}^s)^\sigma \quad \text{with} \quad \hat{p}^s = \frac{p^s}{p^s + q^s}; \quad \hat{q}^s = \frac{q^s}{p^s + q^s}.$$

The function $\phi_n^{(s)}(u)$ is thus the cumulative distribution function of the discrete probability distribution which to the point $\sigma(0)$, with $\sigma \in H_n$, associates the probability $(\hat{p}^s; \hat{q}^s)^\sigma$. The expression of $\omega(s)$ then becomes

$$(3.24) \quad \omega(s) = \sum_{n \geq 0} (p^s + q^s)^n \int_{u \geq 0} r(u) \, d\phi_n^{(s)}(u),$$

where the integral is a Riemann–Stieltjes integral taken on R, although for $u > \lambda/q$ there is no contribution. As we shall see, when n tends to ∞, $\phi_n^{(s)}$ tends to a limit, and the value of this limit for $s = 1$ gives the main term in the local expression of $\omega(s)$.

Lemma 6. For δ such that $p < \delta < 1$, let $D(\delta)$ be the domain

$$D(\delta) = \{s \in C : |\hat{p}^s| < \delta \quad \text{and} \quad |\hat{q}^s| < \delta\}.$$

Then, for each s in $D(\delta)$, there exists a function $\phi_\infty^{(s)}(u)$ defined on R, such that

$$|\phi_n^{(s)}(u) - \phi_\infty^{(s)}(u)| \leq A\delta^n \quad \text{for all } u \text{ in } R \text{ and for some } A \in R.$$

In particular, $\phi_\infty^{(1)}(u)$ has the explicit form

$$(3.25) \qquad \phi_\infty^{(1)}(u) = \begin{cases} 0 & \text{if} \quad u \leq \dfrac{\lambda}{p} \\[2ex] \dfrac{u - \dfrac{\lambda}{p}}{\dfrac{\lambda}{q} - \dfrac{\lambda}{p}} & \text{if} \quad \dfrac{\lambda}{p} \leq u \leq \dfrac{\lambda}{q} \\[2ex] 1 & \text{if} \quad \dfrac{\lambda}{q} \leq u. \end{cases}$$

Proof. From the definition (3.23) of the $\phi_n^{(s)}$, using the decomposition

$H_{n+1} = \sigma_1 H_n \cup \sigma_2 H_n$, we have

$$\phi_{n+1}^{(s)}(u) = \sum_{\substack{\sigma \in H_n \\ \sigma_1 \sigma(0) < u}} \hat{p}^s(\hat{p}^s; \hat{q}^s)^\sigma + \sum_{\substack{\sigma \in H_n \\ \sigma_2 \sigma(0) < u}} \hat{q}^s(\hat{p}^s; q^s)^\sigma,$$

and, since σ_1, σ_2 are monotone increasing functions,

(3.26) $\phi_{n+1}^{(s)}(u) = \hat{p}^s \phi_n^{(s)}(\sigma_1^{-1}(u)) + \hat{q}^s \phi_n^{(s)}(\sigma_2^{-1}(u)),$

with

$$\phi_0^{(s)}(u) = 0 \quad \text{if} \quad u < 0; \qquad \phi_0^{(s)}(u) = 1 \quad \text{if} \quad 0 \leq u.$$

Let us consider the four regions

$$R_1 = \left\{ u \mid u < \frac{\lambda}{p} \right\}; \qquad R_2 = \left\{ u \mid \frac{\lambda}{p} \leq u < 2\lambda \right\}$$

$$R_3 = \left\{ u \mid 2\lambda \leq u < \frac{\lambda}{q} \right\}; \qquad R_4 = \left\{ u \mid \frac{\lambda}{q} \leq u \right\};$$

the substitutions σ_1^{-1}, σ_2^{-1} operate on these regions as follows:

(3.27) $\begin{cases} \sigma_1^{-1}(R_1) \subset R_1 & \sigma_2^{-1}(R_1) \subset R_1 \\ \sigma_1^{-1}(R_2) \subset R_1 & \sigma_2^{-1}(R_2) \subset R_2 \cup R_3 \\ \sigma_1^{-1}(R_3) \subset R_2 \cup R_3 & \sigma_2^{-1}(R_3) \subset R_4 \\ \sigma_1^{-1}(R_4) \subset R_4 & \sigma_2^{-1}(R_4) \subset R_4. \end{cases}$

In particular, each element of $[\lambda/p; \lambda/q[$ has only one image by σ_1^{-1}, σ_2^{-1} in the interval. For any $I \in R$, we define the norm

$$\|f\|_I = \sup \{ |f(u)| : u \in I \}.$$

Let n_0 be such that $\sigma_1 \sigma_2^{n_0}(0) > \lambda/p$; then

(3.28) $\forall n > n_0 \quad \text{and} \quad \forall \sigma \in H_n : \sigma_1 \sigma(0) > \dfrac{\lambda}{p}.$

Working on the four cases of (3.27), we first find that for all $n > n_0$,

$$\|\phi_{n+1}^{(s)} - \phi_n^{(s)}\|_{R_1} \leq \hat{p}^s \|\phi_n^{(s)} - \phi_{n-1}^{(s)}\|_{\sigma^{-1}R_1} + \hat{q}^s \|\phi_n^{(s)} - \phi_{n-1}^{(s)}\|_{\sigma_2^{-1}R_1}$$

$$\leq \delta \|\phi_n^{(s)} - \phi_{n-1}^{(s)}\|_{R_1}.$$

Since the functions in the first norm vanish, using (3.28), thus

$$\|\phi_{n+1}^{(s)} - \phi_n^{(s)}\|_{R_1} \leq A_1 \delta^n \quad \text{for some finite } A_1.$$

Similarly,

$$\|\phi_{n+1}^{(s)} - \phi_n^{(s)}\|_{R_2} \leq \delta \|\phi_n^{(s)} - \phi_{n-1}^{(s)}\|_{R_2 \cup R_3},$$

$$\|\phi_{n+1}^{(s)} - \phi_n^{(s)}\|_{R_3} \leq \|\phi_n^{(s)} - \phi_{n-1}^{(s)}\|_{R_4} + \delta \|\phi_n^{(s)} - \phi_{n-1}^{(s)}\|_{R_2 \cup R_3}.$$

Also, since $\sigma(0) < \lambda/q$ for all finite σ, $\phi_n^{(s)}(u)$ is 1, when $u \in R_4$, hence

$$\|\phi_{n+1}^{(s)} - \phi_n^{(s)}\|_{R_4} = 0.$$

Therefore, for some A_2 and $n > n_0$,

(3.29) $\|\phi_n^{(s)} - \phi_{n-1}^{(s)}\|_R \leqq A_2 \delta^n.$

Hence for each u, the sequence $\{\phi_n^{(s)}(u)\}$ is a Cauchy sequence, so that it converges to a limit $\phi_\infty^{(s)}(u)$. From (3.29) follows that, for some A:

(3.30) $\|\phi_n^{(s)} - \phi_\infty^{(s)}\| < A\delta^n,$

as was to be proved.

Using this result in conjuction with Equation (3.26), one sees that $\phi_\infty^{(s)}$ satisfies the equation

(3.31) $\phi_\infty^{(s)}(u) = \hat{p}^s \phi_\infty^{(s)}(\sigma_1^{-1}(u)) + \hat{q}^s \phi_\infty^{(s)}(\sigma_2^{-1}(u))$

with the boundary conditions

(3.32) $\phi_\infty^{(s)}(u) = 0$ if $u \leqq \dfrac{\lambda}{p},$ $\phi_\infty^{(s)}(u) = 1$ if $\dfrac{\lambda}{q} \leqq u;$

and it is easy to check that (3.31), (3.32) when $s = 1$, are satisfied by the piecewise linear function of the statement of Lemma 6.

Thus $\phi_\infty^{(1)}(u)$ is nothing but the cumulative distribution function associated with the uniform distribution on $[\lambda/p, \lambda/q]$. We now proceed to use this result in conjunction with the expression (3.24) for $\omega(s)$. Define

(3.33) $D(s) = \omega(s) - \dfrac{1}{1 - p^s - q^s} \dfrac{1}{\dfrac{\lambda}{q} - \dfrac{\lambda}{p}} \displaystyle\int_{\lambda/p}^{\lambda/q} r(u)\, du$

which, *a priori*, is only analytic for $\mathrm{Re}(s) > 1$. Using (3.24), Equation (3.33) becomes

$$D(s) = \sum_{n \geqq 0} (p^s + q^s)^n \int_{u \geqq 0} r(u)\, d(\phi_n^{(s)}(u) - \phi_\infty^{(1)}(u))$$

and integrating by parts

(3.34) $D(s) = \displaystyle\sum_{n \geqq 0} (p^s + q^s)^n \int_{u \geqq 0} r'(u)(\phi_\infty^{(1)}(u) - \phi_n^{(s)}(u))\, du.$

Using the decomposition

$$\phi_\infty^{(1)}(u) - \phi_n^{(s)}(u) = \phi_\infty^{(1)}(u) - \phi_n^{(1)}(u) + \phi_n^{(1)}(u) - \phi_n^{(s)}(u),$$

we have

$$D(s) = D_1(s) + D_2(s),$$

with

(3.35) $$D_1(s) = \sum_{n \geq 0} (p^s + q^s)^n \int_{u \geq 0} r'(u)(\phi_\infty^{(1)}(u) - \phi_n^{(1)}(u))\, du,$$

(3.36) $$D_2(s) = \sum_{n \geq 0} (p^s + q^s)^n \int_{u \geq 0} r'(u)(\phi_n^{(1)}(u) - \phi_n^{(s)}(u))\, du.$$

Using the geometric convergence (cf. (3.30)) of Lemma 6, we see that $D_1(s)$ is analytic in a neighbourhood of $s = 1$, and we may write

(3.37a)

$$D_1(1) = \sum_{n \geq 0} \int_{u \geq 0} r'(u)(\phi_\infty^{(1)}(u) - \phi_n^{(1)}(u))\, du,$$

$$D_1(1) = \int_{u \geq 0} r'(u) w_1(u)\, du$$

with

(3.37b) $$w_1(u) = \sum_{n \geq 0} (\phi_\infty^{(1)}(u) - \phi_n^{(1)}(u))\, du.$$

Furthermore, we can show the following.

Lemma 7. The function $D_2(s)$ is analytic in a neighbourhood of $s = 1$.

Proof. We use an indirect argument: since $D_2(s) = D(s) - D_1(s)$ is the difference of two functions meromorphic at $s = 1$, it is meromorphic there and has at most a simple pole at $s = 1$ by Lemma 5. We propose to prove that as $s \to 1+$

$$D_2(s) = o\left(\frac{1}{s-1}\right).$$

which will establish that $D_2(s)$ is regular at $s = 1$. We write first, using (3.35), (3.34), (3.23):

$$D_2(s) = \sum_{n \geq 0} d_n(s), \qquad d_n(s) = (p^s + q^s)^n \sum_{v \in H_n} r(v(0))[(\hat{p}^s; \hat{q}^s)^v - (p; q)^v].$$

We estimate the d_n's using the decomposition

$$H_n = H_k \times H_l \quad \text{with} \quad k = k(n) = [\sqrt{n}], \qquad l = l(n) = n - [\sqrt{n}],$$

and writing

$$d_n(s) = (p^s + q^s)^n \sum_{\sigma \in H_k} \sum_{\tau \in H_l} r(\sigma\tau(0))[(\hat{p}^s; \hat{q}^s)^{\sigma\tau} - (p; q)^{\sigma\tau}].$$

Since r is assumed to be continuously differentiable and $\sigma(\tau(0)) - \sigma(0) = (p; q)^\sigma \tau(0)$, we write $r(\sigma\tau(0)) = r(\sigma(0)) + O((p; q)^\sigma)$ uniformly in σ, τ. Hence

$$d_n(s) = (p^s + q^s)^n \sum_{\sigma \in H_k} \sum_{\tau \in H_l} r(\sigma(0))[(\hat{p}^s; \hat{q}^s)^{\sigma\tau} - (p; q)^{\sigma\tau}]$$

$$+ (p^s + q^s)^n O\left(\sum_{\sigma \in H_k} \sum_{\tau \in H_l} ((\hat{p}^s; \hat{q}^s)^{\sigma\tau} + (p; q)^{\sigma\tau})(p; q)^\sigma \right).$$

Using the fact that

$$\sum_{\tau \in H_l} (\hat{p}^s; \hat{q}^s)^\tau = 1 \quad \text{and} \quad \sum_{\tau \in H_l} (p; q)^{\sigma\tau} = (p; q)^\sigma,$$

we get

$$d_n(s) = (p^s + q^s)^n \sum_{\sigma \in H_k} r(\sigma(0))[(\hat{p}^s; \hat{q}^s)^\sigma - (p; q)^\sigma]$$

$$+ (p^s + q^s)^n O\left[\sum_{\sigma \in H_k} ((\hat{p}^s; \hat{q}^s)^\sigma (p; q)^\sigma + (p^2; q^2)^\sigma) \right].$$

For $s = 1 + \varepsilon$, the term in the second line is

$$O(p^2 + q^2)^k(p^{1+\varepsilon} + q^{1+\varepsilon})^{n-k}(1 + \varepsilon k(n)), \qquad \varepsilon > 0$$

and thus is $O(M^{\sqrt{n}})$ for some $M : 0 < M < 1$.

We now estimate the first term in $d_n(s)$. Computing derivatives, one finds that

$$(\hat{p}^s; \hat{q}^s)^\sigma - (p; q)^\sigma = (p; q)^\sigma[k(n)h(p; q) + |\sigma_1|\log p + |\sigma_2|\log q]$$

$$+ (s - 1) + O(s - 1)^2,$$

uniformly in σ.

We have thus proved that

$$d_n(s) = O(p^s + q^s)^n k(n)(s - 1) + O(M^{\sqrt{n}}).$$

Since, as $x \to 1^-$

$$\sum_n [\sqrt{n}]x^n = O(1 - x)^{-\frac{3}{2}},$$

we obtain $D_2(s) = O(s - 1)^{-\frac{1}{2}}$ and $D_2(s)$ is analytic at $s = 1$ by our preceding remarks.

With Lemma 7 and result (3.37), we can conclude that $D(s)$ is analytic at $s = 1$, so that extracting from the second term in (3.33) the coefficient of $(s - 1)^{-1}$ we have the residue of $\omega(s)$ around $s = 1$; we define the difference $E(s)$,

(3.38) $$E(s) = \omega(s) - \frac{1}{s - 1} \frac{1}{h(p, q)\left(\dfrac{\lambda}{q} - \dfrac{\lambda}{p}\right)} \int_{\lambda/p}^{\lambda/q} r(u)\, du$$

which admits around 1 an expansion of the form

$$a_0 + a_1(s-1) + a_2(s-2)^2 + \cdots.$$

To complete the proof of Theorem 2, we now look for an explicit expression of a_0. From Equations (3.33), (3.38) we see that

$$(3.39) \quad a_0 = E(1) = D(1) + \lim_{s \to 1} \left(\frac{1}{1-p^s-q^s} - \frac{1}{(s-1)h(p,q)} \right) \frac{1}{\frac{\lambda}{q} - \frac{\lambda}{p}} \int_{\lambda/p}^{\lambda/q} r(u)\,du$$

where the existence of $D(1)$ is guaranteed by Lemma 7. Dealing first with the limiting term, the value in the parenthesis equals

$$\frac{d}{ds} \frac{s-1}{1-p^s-q^s} \quad \text{at} \quad s=1,$$

which yields

$$\frac{p \log^2 p + q \log^2 q}{2h^2(p,q)}.$$

Proceeding to compute $D(1)$ we use Equation (3.34) and the decomposition

$$\phi_\infty^{(1)}(u) - \phi_n^{(s)}(u) = \phi_\infty^{(s)}(u) - \phi_n^{(s)}(u) + \phi_\infty^{(1)}(u) - \phi_\infty^{(s)}(u)$$

we see that $D(s)$ can also be written in the form:

$$D(s) = \sum_{n \geq 0} (p^s + q^s)^n \int r'(u)(\phi_\infty^{(s)}(u) - \phi_n^{(s)}(u))\,du$$

$$+ (1-p^s-q^s)^{-1} \int r'(u)(\phi_\infty^{(1)}(u) - \phi_\infty^{(s)}(u))\,du,$$

for $\mathrm{Re}(s) > 1$. When $s \to 1^+$, the first term converges to $D_1(1)$ given by expression (3.37). The second term is thus another form of $D_2(1)$, so that it has a limit when $s \to 1^+$ which is equal to:

$$D_2(1) = \int r'(u)w_2(u)\,du$$
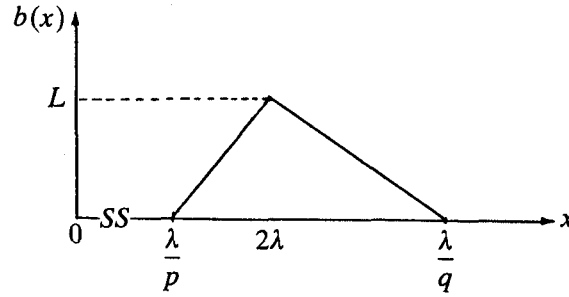
with

$$w_2(u) = -\frac{1}{h(p,q)} \left[\frac{\partial}{\partial s} \phi_\infty^{(s)}(u) \right]_{s=1}.$$

To obtain $w_2(u)$, explicitly we differentiate the functional equation satisfied by $\phi_\infty^{(s)}$ and find:

$$(3.40) \quad \frac{\partial}{\partial s} \phi_\infty^{(s)}(u) = p^s \frac{\partial}{\partial s} \phi_\infty^{(s)}(\sigma_1^{-1}(u)) + q^s \frac{\partial}{\partial s} \phi_\infty^{(s)}(\sigma_2^{-1}(u))$$

$$+ L(s)\phi_\infty^{(s)}(\sigma_1^{-1}(u)) - L(s)\phi_\infty^{(s)}(\sigma_2^{-1}(u))$$

with $L(s) = p^s q^s \log p/q$ and $L = L(1) < 1$ for all p, q.

Figure 2. The graph of the triangular function $b(x)$

Let $b(x)$ be the triangular function

(3.41) $$b(x) = L[\phi_\infty^{(1)}(\sigma_1^{-1}(x)) - \phi_\infty^{(1)}(\sigma_2^{-1}(x))].$$

The graph of $b(x)$ is displayed in Figure 2. Initializing (3.40) at $s = 1$, and integrating it to solve this functional equation, we find that

(3.42) $$w_2(u) = - \sum_{\sigma \in H} (p; q)^\sigma b(\sigma^{-1}(u))/h(p, q).$$

Equation (3.42) defines a function which is nowhere differentiable and which is a superposition of triangular functions of smaller and smaller supports and amplitudes. Functions of a similar nature are not uncommon in the analysis of algorithms.

The function $w_1(u)$ in (3.37b) can also be expressed in a similar way:

(3.43) $$w_1(u) = \sum_{\sigma \in H} (p; q)^\sigma a(\sigma^{-1}(u)),$$

where $a(u)$ is the piecewise-linear function

$$a(u) = \phi_\infty^{(1)}(u) - \phi_0^{(1)}(u)$$

the graph of which is displayed in Figure 3.

These calculations complete the proof of Theorem 2 with $w(u) = w_1(u) + w_2(u)$, using an 'explicit' expression through (3.42) and (3.43).

3.4. *Evaluation of the inverse Mellin transform.* Theorem 2 thus gives expressions for the first terms in the expansion of the functions $\theta(-s)$, $\xi(-s)$ appearing in Lemma 4, around their singularity at $s = -1$. By the Mellin
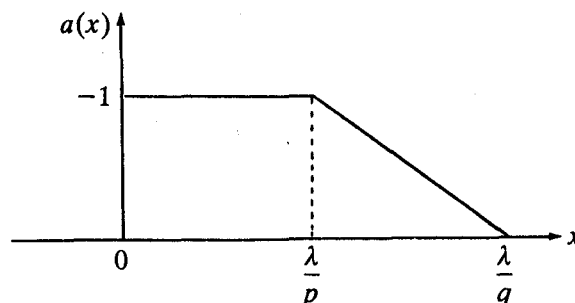
Figure 3. The graph of function $a(x)$

inversion theorem (3.16),

$$\beta(x) = \frac{1}{2i\pi} \int_{-3/2-i\infty}^{-3/2+i\infty} \beta^*(s) x^{-s} \, ds.$$

To carry out the integration we have to establish a general property on the poles of $\beta^*(s)$ in the right half-plane in order to complete the integration contour in the above integral to a closed curve. Let

$$f(s) = p^s + q^s - 1,$$

and

$$Z(f) = \{s \mid \text{Re}(s) > 0; f(s) = 0\}.$$

The set $Z(f)$ contains by Lemma 5 the poles of $\omega(s)$ is the right half-plane. We prove that $Z(f)$ is uniformly discrete in the following sense.

Lemma 8. There exists a real number $\delta > 0$ such that

$$\forall s, s' \in Z(f) \quad |s - s'| > \delta.$$

Proof. Assume on the contrary that the lemma is not satisfied. Then for all $\delta > 0$, there exists an s in $Z(f)$ and an a with $|a| < \delta$ such that

$$p^s + q^s = 1 \quad \text{and} \quad p^{s+a} + q^{s+a} = 1.$$

Eliminating q^s, we should have

(3.44)
$$p^s = \frac{1 - q^s}{p^a - q^a}.$$

Now for small a, a local expansion shows that

$$\frac{1 - q^a}{p^a - q^a} = -\frac{\log q}{\log \frac{p}{q}} + O(a).$$

The function

$$h(x) = -\frac{\log(1-x)}{\log \frac{x}{1-x}}$$

satisfies $h(x) > 1$ for all $x \in (0, 1/2)$.

Thus the right-hand side of (3.44) is of modulus strictly larger than 1 for a small enough. There is thus a contradiction in (3.44) since for $\text{Re}(s) > 0$, $|p^s| < 1$; this establishes the lemma.

The argument used in Lemma 8 can actually be used to prove that all

elements of $Z(f)$ are simple zeros of $f(s)$ and that for every $\delta > 0$ there exists an $\eta > 0$ such that

$$\forall s', \ \mathrm{Re}(s') > 0, \qquad \forall s \in Z(f) : |s - s'| > \delta \qquad f(s') > \eta.$$

One then sees that for some fixed small enough $\varepsilon > 0$ such that the minimal distance between points in $Z(f)$ is larger than 4ε and for each integer n there exists a closed contour Γ_n with the following properties.

(i) Γ_n consists of four curves:

$$\Gamma_n = \Gamma_n^1 + \Gamma_n^2 + \Gamma_n^3 + \Gamma_n^4$$

with

$$\Gamma_n^1 = \{-\tfrac{3}{2} + it \mid t \in [-n, +n]\}$$
$$\Gamma_n^2 = \{z \mid \mathrm{Re}\,(z) \in [-\tfrac{3}{2}, -\varepsilon]; \ \mathrm{Im}\,(z) \in [n, n + 2\varepsilon]\}$$
$$\Gamma_n^3 = \{z \mid \mathrm{Re}\,(z) \in [-3\varepsilon, \varepsilon]; \ \mathrm{Im}\,(z) \in [-n - 2\varepsilon, n + 2\varepsilon]\}$$
$$\Gamma_n^4 = \{z \mid \mathrm{Re}\,(z) \in [-\tfrac{3}{2}, -\varepsilon]; \ \mathrm{Im}\,(z) \in [-n - 2\varepsilon, -n]\}.$$

(ii) Each point in Γ_n is at a distance at least ε from a zero of $f(-s)$.

Such a contour can be constructed by distorting a rectangular contour so as to avoid the zeros of $f(-s)$. We can thus assume that Γ_n is rectifiable and has length $O(n)$. Figure 4 displays the shape of such a contour. We now consider the integral

$$I_n = \frac{1}{2i\pi} \int_\Gamma \beta^*(s) x^{-s} \, ds$$

with Γ_n oriented clockwise, and let n tend to ∞.

By the complement formula for the gamma function, $\Gamma(z)$ has an exponential decrease at ∞ along imaginary lines. On the other hand, along all of Γ_n,
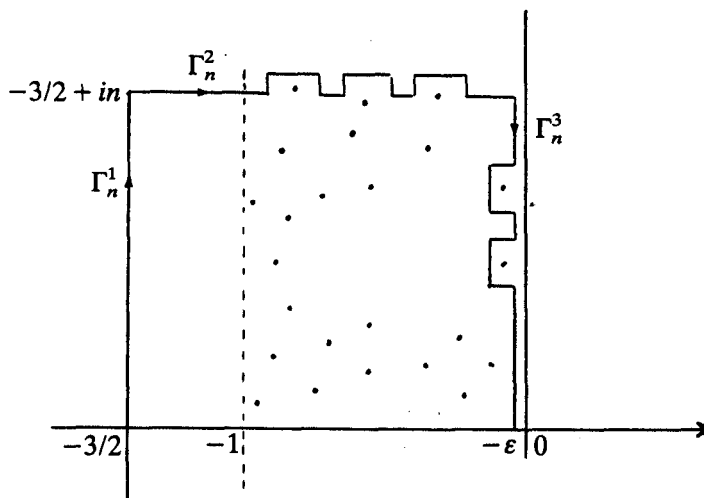

Figure 4. A schema showing the contour Γ_n in the upper half-plane

$\theta(-s)$ and $\xi(-s)$ are bounded by Lemma 5 and the remarks following it. Thus, the integral

$$(3.45) \qquad \int_{\Gamma_n^i} \beta^*(s) x^{-s} \, ds.$$

taken along Γ_n^2 and Γ_n^4 tends to 0 exponentially fast. Integrals (3.45) along Γ_n^1 and Γ_n^3 tends to limits that are respectively $\beta(x)$ and a function of x which is $O(x^{3\varepsilon})$ as x gets large. Taking residues into account by the Cauchy theorem, we therefore get

$$\beta(x) = -\sum \operatorname{Re} s(\beta^*(s) x^{-s}) + O(x^{3\varepsilon})$$

the sum being extended to all the poles that lie inside the contour. Because of the exponential decrease of $\Gamma(s)$ at $i\infty$ and the uniform discreteness of the set of zeros of $f(s)$, the sum of the residues is absolutely convergent.

Note that while $(1 - p^{-s} - q^{-s})^{-1}$ does have poles with $\operatorname{Re}(s) > -1$ (for certain p), $\beta^*(s)$ does not, as can be shown by bounding (3.17).

To conclude the computation of $\beta(s)$, we estimate the residue of $\beta^*(s)$ at $s = -1$. Introducing the notations $\omega(s) = (s - 1)^{-1}\mu(r) + v(r) + o(s - 1)$ with

$$(3.46) \qquad \mu(r) = \frac{1}{h(p, q)} \frac{1}{\dfrac{\lambda}{q} - \dfrac{\lambda}{p}} \int_{\lambda/p}^{\lambda/q} r(u) \, du$$

$$(3.47) \qquad v(r) = \int_0^{\lambda/q} r'(u) w(u) \, du + \frac{p \log^2 p + q \log^2 q}{2\left(\dfrac{\lambda}{q} - \dfrac{\lambda}{p}\right) h^2(p, q)} \int_{\lambda/p}^{\lambda/q} r(u) \, du$$

we have around $s = -1$:

$$\theta(-s) = \mu(\exp(-u))(s + 1)^{-1} + v(\exp(-u)) + o(s + 1)$$

$$\xi(-s) = -\mu(u \exp(-u))(s + 1)^{-1} + v(u \exp(-u)) + o(s + 1)$$

$$\Gamma(s + 1) = \frac{1}{s + 1} - \gamma + o(s + 1)$$

$$\Gamma(s) = \frac{-1}{s + 1} + (\gamma - 1) + o(s + 1)$$

$$x^{-s} = x(1 - (s + 1) \log x + o(s + 1)^2).$$

With these expansions, we see that $\beta^*(s)$ has only a simple pole at $s = -1$.

The residue there is found to be equal to $-A$ where

$$(3.48) \quad A = (\exp(-\lambda/p) - \exp(-\lambda/q))(GK(1 - 2\gamma) + 2H(K - 1))$$

$$+ \int [Ku \exp(-u) - (2K - 1) \exp(-u)]w(u) \, du$$

where

$$G = \frac{1}{\left(\dfrac{\lambda}{q} - \dfrac{\lambda}{p}\right)h(p, q)}, \qquad H = G\frac{p \log^2 p + q \log^2 q}{2h(p, q)}$$

and K is given in (2.7). Similarly, let χ be a pole of $(1 - p^{-s} - q^{-s})^{-1}$ with $\mathrm{Re}(s) \geqq -1$, the residue of $\beta^*(s)$ at χ is $-a(\chi)$ where

$$(3.49) \quad a(\chi) = \frac{1}{p^\chi \log p + q^\chi \log q} [\pi_1(\chi)(\Gamma(\chi) + K\Gamma(1 + \chi) + K\pi_2(\chi)\Gamma(\chi))],$$

and π_1 and π_2 are the π-functions of Lemma 5 (in (3.19)) associated with $r(u) = \exp(-u)$, $u \exp(-u)$. Note that this collection of poles may contain superfluous ones. The leading term, however, is genuine.

With these calculations, we have the following result.

Theorem 3. The average time to resolve an n-collision satisfies

$$(3.50) \quad \alpha_n = \frac{2A}{1 + 2D(\lambda)} n + \frac{2}{1 + 2D(\lambda)} \sum_\chi a(\chi)n^{-\chi} + O(n^{1-\eta}),$$

the sum being extended to χ's satisfying

$$1 - p^{-\chi} - q^{-\chi} = 0; \qquad -1 \leqq \mathrm{Re}(\chi) < -1 + \eta; \qquad \chi \neq -1$$

for any sufficiently small $\eta > 0$.

The sum in the expression of the theorem is a bounded fluctuating function with an amplitude small compared to the value of A, typically less by several orders of magnitude. Its asymptotic nature depends on very specific arithmetical properties of numbers p and q. In the sum

$$(3.51) \quad \sum a(\chi)n^{-\chi}$$

for large n, the $a(\chi)$ have an exponential decrease in $|\mathrm{Im}(\chi)|$ while the $n^{-\chi}$ increase with $|\mathrm{Re}(\chi)|$. Estimating the order of (3.51) thus necessitates determining the relation between $\mathrm{Re}(\chi)$ and $\mathrm{Im}(\chi)$ for the leftmost χ's that are poles of $(1 - p^{-s} - q^{-s})^{-1}$. Setting

$$\mathrm{Re}(\chi) = -1 + \varepsilon; \qquad \mathrm{Im}(\chi) = t,$$

we look for the solutions of

(3.52) $p^{1-\varepsilon} \exp{(it \log p)} + q^{1-\varepsilon} \exp{(it \log q)} = 1$

for small positive ε. Simplifying (3.52), it decomposes into

(3.53) $p^{1-\varepsilon} \cos{(t . \log p)} + q^{1-\varepsilon} \cos{(t . \log q)} = 1$

(3.54) $p^{1-\varepsilon} \sin{(t . \log p)} + q^{1-\varepsilon} \sin{(t . \log q)} = 0.$

With α, β denoting the principal values of $t \log p$ and $t \log q$ in $[-\pi, \pi)$, (3.54) yields

(3.55) $$\frac{\sin \alpha}{\sin \beta} = -\frac{q^{1-\varepsilon}}{p^{1-\varepsilon}},$$

so that α and β are of opposite sign. For ε small enough, from (3.53), (3.55) we see that α, β must be small and in the limit, when $\varepsilon = 0$. (3.53) can only be satisfied by $\alpha = \beta = 0$.

Local expansions show that for some constants A, B,

(3.56) $|\alpha| \leq A \varepsilon^{\frac{1}{2}}; \qquad |\beta| \leq B \varepsilon^{\frac{1}{2}}.$

Since α and β are principal values of $t \log p$, $t \log q$, we have

$$t \log p = 2a\pi + \alpha$$

$$t \log q = 2b\pi + \beta$$

for integral a, b. Thus eliminating t and using (3.56), one must have

(3.57) $$\left| \rho - \frac{a}{b} \right| < \frac{D\varepsilon^{\frac{1}{2}}}{b} \quad \text{where} \quad \rho = \frac{\log p}{\log q}.$$

Since ρ is linearly related to t, this represents a relation between ε and t, i.e. Re (χ) and Im (χ). Note that as $\varepsilon \to 0$ values of a and b that will satisfy (3.57) will increase, and provide a vanishing contribution to the sum unless the following holds.

Corollary. If $\rho = \log p / \log q$ is rational, i.e. $\rho = d/r$ with $(d, r) = 1$, one has

(3.58) $$\alpha_n = \frac{2A}{1 + 2D(\lambda)} n + n . P(r . \log_p n) + o(n^{1-\eta})$$

for some $\eta > 0$, with $P(u)$ a Fourier series of u with mean value 0.

The proof relies on the fact that in this case (3.52) admits solutions with $\varepsilon = 0$ (this is also apparent from (3.57)). Note that in this case the Dirichlet series have a pole-free strip right of Re $(s) = -1$.

While in general (when ρ is not rational) the fluctuating function (3.51) is

$o(n)$, and the limit $\lim_{n\to\infty} \alpha_n/n$ exists, this is *not* the case in (3.58); the latter represents a situation where the ratio α_n/n continues to oscillate (with a minute amplitude) around the 'mean value' given by the leading term.

4. Evaluating the mean CRI duration when $p = q = \frac{1}{2}$

When $p = \frac{1}{2}$ the computational structures of Sections 2 and 3 simplify considerably, even though one expects no qualitative difference in the behavior of the underlying processes. Note that this case comes under the corollary to Theorem 3. Solving the equation $D(\lambda) = -\frac{1}{2}$, to obtain the range of stability still involves numerical search. Indeed, substituting $p = \frac{1}{2}$ in g_n and k_n of (2.24), recalling that now $\sigma(z)$, when $|\sigma| = i$, is

$$\sigma_i(z) = \sigma_i(0) + z/2^i = 2\lambda(1 - 1/2^i) + z/2^i,$$

one obtains

$$(4.1) \quad D(\lambda) = K \exp(-2\lambda) \sum_{i \geq 0} 2^i \exp(2\lambda/2^i)\{\exp(-\lambda/2^i)(1 - \lambda/2^i)$$

$$-1 - 2(\lambda/2^i)^2 + 2(\lambda/2^i)\}$$

a rather finicky series that converges fast to $-1/2$ for $\lambda = \lambda_{\max} = 0.360177147+$, in good agreement with the results reported in Fayolle and Hofri (1982).

Now α_n can be evaluated also from the solution given in Lemma 1—for $p = 1/2$ it is relatively painless, or approximated by (3.48). The first again gives a series which is amenable for evaluation:

$$(4.2) \quad \alpha_n = 1 - \frac{2K \exp(-2\lambda)}{1 + 2D(\lambda)} \sum_{i \geq 0} 2^i \exp(2\lambda/2^i)[(1 - 2\lambda/2^i)((1 - 2^{-i})^n + n2^i - 1))$$

$$+ n2^{-i}((1 - 2^{-i})^{n-1} - 1)].$$

This too agrees satisfyingly with the values obtained in Fayolle and Hofri (1982). From Equation (3.48) the first term provides $\exp(-2\lambda)/(1 - 2\lambda) \log 2$. For the integral we need $w(u)$. Now for $p = q$, $w_2(u) \neq 0$ for one point only, $u = 2\lambda$, and its contribution to the integral vanishes. The value of $w_1(u)$ is $\sum_{i \geq 0} a(\sigma_i^{-1}(u))$, and since a vanishes outside $(0, 2\lambda)$ we obtain the contribution of the second term in (3.48)

$$= \sum_{i \geq 0} \int_{2\lambda(1-2^{-i})}^{\lambda} \exp(-u)(2K - 1 - Ku)\, du$$

$$= 2K\lambda \exp(-2\lambda) \sum_{i \geq 0} 2^{-i} \exp(2\lambda/2^i)$$

hence

$$(4.3) \qquad A = K \exp(-2\lambda) \left\{ \frac{1}{\log 2} + 2\lambda \sum_{i \geq 0} 2^{-i} \exp(2\lambda/2^i) \right\}.$$

For this particular choice of p the set $\{\chi\}$ in (3.50) is known; it comprises $\chi_k = -1 + 2\pi i k / \log 2$, with k going over all the integers. But due to cancellations the sum gives an increase in α_n that is about five to seven orders of magnitude below the main term, and we neglected it in the numerical evaluations. Comparing the values of α_n according to (4.2) and (4.3) showed consistent overestimate when using the 'asymptotic' value. The difference depends on λ: for $n = 10$ and λ in the range $(0\cdot1, 0\cdot36)$ it varied between $11\cdot7\%$ to $7\cdot7\%$. For $n = 20$, it was $2\cdot3\%$ to $3\cdot7\%$.

5. The ergodicity condition

Theorem. The necessary and sufficient condition for a stable transmission process (i.e. $\alpha_n < \infty$ for all n) is $\lambda < \lambda_{\max}$ where $\lambda_{\max}$ is the smallest positive root of

$$(5.1) \qquad\qquad\qquad 1 + 2D(\lambda) = 0$$

(see Equation (2.19)).

We note that this condition implies a finite delay for each packet with probability 1.

Proof. The proof relies on results of Section 3 together with standard properties of Markov chains (see for example Çinlar (1975)). We shall deal with it in some detail as it touches upon a number of issues which are significant in this type of model.

The state of the system comprises the aggregate stack of all the active transmitters at time t and can be represented by a vector of variable length Y

$$(5.2) \qquad\qquad \bar{N}(t) = \{[N(0), N(1), \cdots, N(Y)], Y\}$$

where Y is a random variable, and $N(i)$ is the number of active users at level i of their stack. The time index t denotes slot count.

Note. $\{\bar{N}(t)\}_{t \geq 1}$ form a first-order irreducible aperiodic Markov chain, with countable state space. Such a chain has the same type for all its states (i.e. all are recurrent or all are transient). α_n is the mean time of first visit from the state $\{[n], 0\}$ to the state $\{[0], 0\}$, and the above statement implies that all α_n, $n \geq 2$ are finite or infinite together. Moreover, since a CRI starts with the

new packets that were generated in a single slot,

$$\Psi(\lambda) = \sum_{n=0} \alpha_n \frac{\lambda^n}{n!} \exp(-\lambda) = \frac{1}{P_0}, \quad \text{where} \quad P_0 = \sum_n \Pr\{\bar{N}(t) = \{[n], 0\}\}.$$

Sufficiency. From Theorem 1, we have that the functional equation of $\Psi(z)$ (which yields the α_n) has a unique entire solution, and Theorem 3 provides that it is finite unless $\lambda \in \{\lambda_0\} \cup \lambda_1$, where $\{\lambda_0\}$ is the set of roots of (5.1) and λ_1 is a root of $(\lambda/q)\exp(-\lambda/q) - (\lambda/p)\exp(-\lambda/p) = 0$. We shall denote by λ_{max} the least of them all; for all p, $\min\{\lambda_0\} < \lambda_1$, and therefore, for $0 \leqq \lambda < \lambda_{max}$, $\alpha(z)$ is analytic and its uniqueness assures that those α_n are indeed $E(L_n)$. On the contrary, if $E(L_n)$ did not exist it would imply a singularity for $\alpha(z)$ in the λ plane, which it cannot have there.

Necessity. The thing to prove here is that α_n that result from the solution (3.23) for $\lambda > \lambda_{max}$ have no probabilistic interpretation, i.e. they are not $E(L_n)$. (For the rest of this section we denote $E(L_n)$ by l_n, to emphasize its (possible) distinction from α_n, defined for the time being as $[z^n]n!\alpha(z)$). We precede the proof by the following.

Observation 1. The means l_n, $n \geqq 2$ are monotonically increasing in λ.

Proof of Observation 1. Admittedly this observation appears obvious, but we wish to show its consistency *with the recurrence relations*. A preliminary result is an even more 'obvious' observation.

Observation 2. $l_{n+1} \geqq l_n$. This observation is proved only for $l_n < \infty$, and uses the note above that this inequality also implies $l_{n+1} < \infty$. We note trivially $l_2 > l_1 = l_0$. The proof employs a procedure of examining each realization of L_{n+1}, up to a suitable slot, and then adding the expected duration of the rest of the realization.

To a group of n colliders add one tagged packet and observe the evolution of the CRA till the next slot the tagged packet is in level 0. Up to this point the CRA behaved precisely as it would to produce L_n. In this stopping slot, level 0 may contain

  (1) One packet,
  (2) Two packets,
  (3) More than two packets.

In case (1) the tagged packet exits the system, and whether the CRA now terminates or not, this L_{n+1} equals (the corresponding) L_n.

In case (2) the tagged packet converts an L_1 to an L_2 that will end later. The rest of L_{n+1} has the same tail, in probability, as the corresponding L_n, and equal expected duration, thus in summary here $l_{n+1} > l_n$.

In case (3) L_n as well as L_{n+1} proceed with a collision, and again we follow

the CRA until the tagged packet reappears in level 0. The requirement $l_n < \infty$ implies that, with probability 1, event (3) will occur a finite number of times: thus, with probability 1, we conclude the process in an event that allows to state $l_{n+1} \geqq l_n$.

Armed with this inequality we proceed to prove observation 1, with the following scenario: at slot 1 an n-collision occurred; the arrival rate during this slot is changed from λ to $\tilde{\lambda} > \lambda$, and in the following slots it reverts to λ. Let the quantities that correspond to this scenario, where they differ from the usual (unmodified) ones, be denoted by a tilde. Writing

(5.3) $$\tilde{l}_n = 1 + E_I[E_{\tilde{X}}(\tilde{l}_{I+\tilde{X}}) + E_Y(\tilde{l}_{n-I+Y})]$$

(5.4) $$= 1 + E_I[E_{\tilde{X}}(\tilde{l}_{I+\tilde{X}}) + E_Y(l_{n-I+Y})],$$

where (5.4) results from (5.3) by noting that once $\tilde{X}$ is determined both components depend on future samples of X only. To conclude that $\tilde{l}_n > l_n$, it suffices to show: $E_{\tilde{X}}(\tilde{l}_{I+\tilde{X}}) > E_X(l_{I+X})$, for any value of I. We note that viewing $\tilde{a}(j)$ and $a(j)$ as functions of j, there exists a unique value j_0 that satisfies

$$\tilde{a}(j) - a(j) \begin{cases} \leqq 0, & j \leqq j_0 \\ \geqq 0, & j > j_0 \end{cases}, \quad \text{where} \quad a(j) = e^{-\lambda}\frac{\lambda^j}{j!}.$$

Actually $[\lambda] \leqq j_0 \leqq [\tilde{\lambda}]$. We have to show $G = \sum_{j \geqq 0} [\tilde{a}(j) - a(j)]l_{I+j} \geqq 0$. Splitting the sum at j_0

$$G = \sum_{j=0}^{j_0} [\tilde{a}(j) - a(j)]l_{I+j} + \sum_{j>j_0} [\tilde{a}(j) - a(j)]l_{I+j},$$

we note that the terms in brackets are $\leqq 0$ in the first sum and $\geqq 0$ in the second. Thus observation 2 provides that each sum in the right-hand side will decrease if we replace l_{I+j} by l_{I+j_0}, hence

$$G > \sum_{j \geqq 0} [\tilde{a}(j) - a(j)]l_{I+j_0},$$

but now the right-hand side vanishes ($\sum a(j) = \sum \tilde{a}(j) = 1$), and hence

(5.5) $$\tilde{l}_n > l_n.$$

Taking (5.5) as the basis for induction, assume it holds if the change $\lambda \to \tilde{\lambda} > \lambda$ persists for k slots. Assume the change is not extended to $k + 1$ slots. Writing the equivalent of (5.3) with obvious change of notation

(5.6) $$\tilde{l}_n(k + 1) = 1 + E_I\{E_{\tilde{X}}[\tilde{l}_{I+\tilde{X}}(k) + E_{\tilde{Y}}[\tilde{l}_{n-I+\tilde{Y}}(k - L_{I+\tilde{X}})]]\},$$

we observe that both $\tilde{l}$ on the right-hand side are larger, by the induction hypothesis, than the corresponding l, with the first one satisfying a strict inequality, hence

$$(5.7) \qquad \tilde{l}_n(k+1) > l_n$$

and this will hold for all k. This completes the proof of Observation 1.

Since l_n diverges to ∞ as $\lambda \to \lambda_{\max}$ they remain infinite for $\lambda > \lambda_{\max}$ by this observation. Thus the necessity is established as well.

The ergodicity condition $\lambda < \lambda_{\max}$ is equivalent, by (2.23a), to $\lambda_{\max} = \inf\{\lambda \mid 0 < \Psi(\lambda) < \infty\}$ and the stochastic interpretation of $\Psi(\lambda)$ is easy: since

$$\Psi(\lambda) = \exp(-\lambda) \sum_{n=0}^{\infty} \frac{\lambda^n \alpha_n}{n!},$$

we see that $\Psi(\lambda)$ represents the expected value of the mean collision resolution interval (in the wide sense, because α_0 and α_1 are taken into account).

6. Conclusion

As said in Section 1, other schemes can be proposed to get a higher $\lambda_{\max}$. They essentially try to save 'doomed' slots. This can lead for example to the following recursive relationships for the L_n:

$$L_n = 1 + \begin{cases} L_{I+X} + L_{n-I+y} & \text{if } I + X \neq 0 \\ L_n & \text{if } I + X = 0, \end{cases}$$

(see Massey (1981) and Fayolle and Hofri (1982)).

The functional equation for the generating function of the $\alpha_n = E(L_n)$ is then non-symmetrical with respect to p and $q = 1 - p$. Moreover, there is a term involving $\alpha(qz)$. Nevertheless, the same analysis can be applied, although the details become more involved.

Note added in proof. In a companion paper, the present authors jointly with Philippe Jacquet provide a delay analysis for this protocol (Fayolle et al. (1985)).

References

ÇINLAR, E. (1975) *Introduction to Stochastic Processes.* Prentice Hall, Englewood Cliffs, NJ.

DAVIES, B. (1978) *Integral Transforms and their Applications.* Springer-Verlag, New York.

DOETSCH, G. (1955) *Handbuch der Laplace Tranformation.* Birkauser, Basel.

FAYOLLE, G. (1975) Étude du comportement d'un canal radio partagé entre plusieurs utilisateurs. Thèse Docteur Ingénieur, Université Paris VI.

FAYOLLE, G. AND HOFRI, M. (1982) On the capacity of a collision channel under stack-based collision resolution algorithms. Dept of Computer Science, Technion.

KNUTH, D. E. (1973) *The Art of Computer Programming, Vol. 3: Sorting and Searching*. Addison-Wesley, Reading, Mass.

MASSEY, J. L. (1981) Collision resolution algorithms and random access communication. In *Multiuser Communications*, ed. G. Longo, CISM Courses and Lectures **265**, Springer-Verlag, New York, 73–137.

TSYBAKOV, B. A. AND VVEDENSKAYA, N. D. (1980) Random multiple-access stack algorithm. *Probl. Inf. Trans.* **16**, 230–243.

VVENDSKAYA, N. D. AND TSYBAKOV, B. S. (1983) Random multiple-access of packets on a channel with errors (in Russian). *Probl. Perd. Inf.* **19**, 52–68.

Reference added in proof
FAYOLLE, G., FLAJOLET, P., HOFRI, M. AND JACQUET, P. (1985) Analysis of a stack algorithm for random multiple-access communication. *IEEE Trans. Inf. Theory* **31**, 244–254.