A probabilistic analysis of a leader election algorithm
Hanene Mohamed

To cite this version:

HAL Id: hal-00130117
https://hal.inria.fr/hal-00130117v2
Submitted on 17 Aug 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A probabilistic analysis of a leader election algorithm

Hanène Mohamed†

†INRIA Rocquencourt, Domaine de Voluceau, BP 105, F-78153 Le Chesnay, France

A leader election algorithm is an elimination process that divides recursively into two subgroups an initial group of \( n \) items, eliminates one subgroup and continues the procedure until a subgroup is of size 1. In this paper the biased case is analyzed. We are interested in the cost of the algorithm, i.e. the number of operations needed until the algorithm stops. Using a probabilistic approach, the asymptotic behavior of the algorithm is shown to be related to the behavior of a hitting time of two random sequences on \([0, 1]\).

Keywords: Election Algorithm. Randomized Selection Algorithm. Distributed Systems. Asymptotic Oscillating Behavior. Probabilistic de-Poissonization.

1 Introduction

A single-hop network is a distributed system of \( n \) nodes, also called stations, sharing a common communication channel which can transmit only one message per time unit. In the special case of collision detection, the channel is ternary feedback; each station sending a message to the network can simultaneously listen to the channel and detect: a collision when at least there are two broadcast attempts, a silence when no station sends message, or a success when exactly one station sends its message. A single-hop network with collision detection is called multiple access channel.

Consider a multiple access channel of \( n \) stations which has to elect a leader to control and organize the network. Because of links or stations failures, the leader may be temporarily out of service. Such failure can be detected by a silence, in which case the system stops normal operations and initiates the election process: the system has to identify a new leader in a reasonable execution time. We are interested in the cost of the algorithm, i.e. the number of operations needed to find a leader.

1.1 Leader election problem

We assume that the size \( n \) of the multiple access channel is unknown. Moreover, each station is assumed to have a unique identifying number ID. To elect a leader among themselves, stations have to use the same algorithm. The case \( n \in \{0, 1\} \) is trivial, \( n \) is assumed to be greater than 2. Let us recall the basic one:

— Deterministic Initialization: At the first time unit, each station send a message with its ID number to the common channel. As \( n \geq 2 \), all stations detect a collision.

— Randomized Selection Process: Each station \( S \) generates independently a Bernoulli random variable \( B_S \) with parameter \( p \). Only which obtains \( B_S = 1 \) is allowed to send its message again during the next time unit.

For a station \( S \), there are two cases:

1. If \( B_S = 1 \), station \( S \) will be called Active; \( S \) sends again its message to the channel and can detect
   — a success; only station \( S \) is trying transmission, then all the other stations receive its ID’s message and \( S \) obtains the status of leader. The protocol is finished.
   — a collision; station \( S \) is not the only candidate to be leader, and so has to generate \( B_S \) again.
2. Otherwise, station $S$ becomes Non Active; it remains candidate to be leader, listens to the channel but does not participate to the transmission. So it can detect

— a success; only one station $S' \neq S$ is trying transmission. The other stations (including $S$) detect its ID. So $S'$ obtains the status of leader. The protocol is finished.

— a collision; although station $S$ is not participating to the selection process, there are at least 2 Active stations. So, station $S$ is eliminated.

— a silence; all stations are Non Active, so station $S$ has to generate $B_S$ again to send or not its ID’s message to the channel.

That is, at the end of the protocol, a single station remains Active and becomes the leader of the system.

This splitting process using a Bernoulli random variable was also used in the tree protocol of Capetanakis and Tsybakov. For a survey, see Mathys and Flajolet (1985).

The example below illustrates the election process applied to a group of 4 stations \{A, B, C, D\}. In this case, the leader $A$ is elected in 4 times units.

<table>
<thead>
<tr>
<th>time units</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Stations</td>
<td>A BC D</td>
<td>A BC</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Non Active Stations</td>
<td>D</td>
<td>A BC</td>
<td>B C</td>
<td></td>
</tr>
<tr>
<td>Eliminated Stations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channel feedback</td>
<td>Collision</td>
<td>Collision</td>
<td>Silence</td>
<td>Success</td>
</tr>
</tbody>
</table>

Fig. 1: Election of the leader $A$; $H_4 = 4$. Incomplete tree structure.

**Definition 1 (Algorithm Cost)** It is the number of rounds needed to find a leader. Denote by $H_n$ the algorithm cost when the size of the network is $n$.

Such a randomized elimination algorithm has various applications in distributed systems like cellular phones and wireless communication networks. In mobile Ad-hoc networks, failures occur when mobile nodes move out of transmission range. The unstable topology of the network makes leader election problem more complex. For more details, see Malpani et al. (2000). Electing a leader in a computer network is fundamental to supervise communication and synchronization. See Fill et al. (1996). It is also studied in a context of radio networks. For an interesting survey on randomized communication in this context, see Chlebus (2001). For more elaborate leader election algorithms on radio network with no collision detection, see Lavault et al. (2003).

1.2 Splitting process and tree structure

Formally, the algorithm starts with a group of $n$ items which is divided in two subgroups. The probability that an item is sent into the left subgroup is $p$. This subgroup will be divided by the same process. The other items will be ignored. If the left subgroup is empty, the algorithm restarts from the previous level.

This distributed algorithm is a randomized elimination process with a natural binary tree structure (Fig 1). At the root of the associated tree, is the initial group of items. In the first split, it generates two nodes: the
left one will be split by the same process, the right one is a terminal node, also called leaf, which will not be treated by the algorithm except when the left node is empty. Only in this case, the right node will be split into two one.

Thus, this tree structure can be represented as an incomplete tree in which only one side is developed. We define the height of the associated tree as the length of the path from the root to the leader which is the longest root-to-leaf path in the tree (see Fill et al. (1996)). Then the algorithm cost is equivalently the height of the associated tree. Fig.1 illustrates this equality.

1.3 Previous works

It is known that the average cost of the leader election algorithm is of logarithmic order in \( n \) with an oscillating behavior. See Prodinger (1993) for the unbiased case \( p = 1/2 \), Janson and Szpankowski (1997) for the biased one \( p \neq 1/2 \).

Consider the Poisson model of the leader election problem, i.e. the election process applied to a network with random size following a Poisson process \( \mathcal{N}_x \) (see 1.6). Let \( h \) the Poisson transform of the sequence of average cost of the algorithm \( \{E(H_n)\}_{n \geq 0} \).

**Definition 2 (Poisson transform)** For \( x > 0 \), the Poisson transform of the sequence \( \{E(H_n)\} \) is the function \( h \) defined by

\[
h(x) = E(H_{\mathcal{N}_x}) = \sum_{n=2}^{+\infty} E(H_n) \frac{x^n}{n^5} e^{-x}.
\]

Then, function \( h \) is solution of a functional equation, called basic functional equation associated to the algorithm

\[
h(x) = h(px) + h(qx) e^{-px} + f(x), \quad \text{where } p + q = 1
\]

and \( f \) is a given function. Equation 1 is the starting point of all studies made on this algorithm.

The unbiased case

When the splitting process follows a Bernoulli random variable of parameter 1/2, the leader election algorithm is called symmetrical. Observe that, for the unbiased case, the functional equation 1 is solved by direct iteration. In fact, the Poisson transform \( h \) verifies

\[
h(x) = h(x/2) \left(1 + e^{-x/2}\right) + f(x),
\]

which can be rewritten \( g(x) = g(x/2) + f(x)/(1 - e^{-x}) \) where \( g(x) = h(x)/(1 - e^{-x}) \).

The first analysis of the leader election algorithm was proposed by Prodinger (1993). He investigated different parameters of interest such as the height, called depth in his paper, the size of the associated tree, i.e. the number of nodes. Using combinatorial techniques, he established exact expressions and asymptotic formulas for these quantities for the symmetrical case. So, it is shown that for an initial group of size \( n \), the algorithm stops on average after about \( \log_2 n \) steps. Using complex analysis techniques like Mellin and inverse Mellin transform, Fill et al. (1996) studied the asymptotic behavior of the first two moments of the algorithm cost \( H_n \). Moreover, they obtained the exact expression and asymptotic behavior of the distribution of \( H_n \) and they have shown that a limit distribution for the centered algorithm cost \( H_n - \lfloor \log_2 n \rfloor \) does not exist. For a survey on Mellin transform, see Flajolet et al. (1995).

The biased case

If the splitting process is biased, i.e. the probability that an item is sent into the left subgroup is \( p \neq 1/2 \), the algorithm is called asymmetrical. Studies on biased case become more rare. An asymmetric leader election algorithm was investigated by Janson and Szpankowski (1997) using complex analysis techniques. The asymptotic behavior of the first two moments of the algorithm cost \( H_n \) is given in term of the sequence of their exact values \( \{E(H_j)\}_{j \in \mathbb{N}} \) computed numerically from two recurrence equations.

This implicit dependence is due to the asymmetry of the functional equation 1 obtained by Poissonization. The coefficient \( e^{-p x} \) makes more complex the establishment of an iterative scheme such as in the context of a protocol for a multi-access broadcast channel (see Fayolle et al. (1986)). Applying the Mellin transform to equation 1 without solving it yields this dependence.
1.4 Related leader election algorithms

Leader election algorithm in network of fixed size

Consider a simple algorithm for leader election algorithm in the context of communication network; at each level, the probability \( p \) for a station to send its message depends on the number \( n \) of stations remaining in the elimination process; \( p = 1/n \). Expected run time is \( O(1) \) but it is clear that is necessary to know the number of active stations in advance, or at least to estimate it. See [Prodinger (1993)] for the unbiased case, Ward and Szpankowski (2004) for the biased case, Prodinger (1993) for the unbiased one. Let \( M \) obtain the same result as the moderator continue the process. See Ward and Szpankowski (2004) for the elimination process; each of participants and the moderator throws independently a coin and only those who consider a variant of the leader election algorithm by introducing a moderator who determines the elimination process; see [Ward and Szpankowski (2004)] for the biased case. [Prodinger (1993)] for the unbiased one. Let \( M_n \) the number of participants remaining in the last non trivial round from an initial group of \( n \) items. It is asymptotically equivalent to the multiplicity of phrases in the LZ’77 data compression scheme.

1.5 Overview

In a previous paper on splitting algorithms, [Mohamed and Robert (2005)] proposed a direct approach based on a probabilistic reformulation of a basic functional equation associated to such algorithms. The purpose of this work is to apply the techniques used by [Mohamed and Robert (2005)] to analyze an additive quantity in the context of an incomplete tree structure. In Section 2 a similar series formula for the average cost \( E(H_n) \) is given by Proposition 1. The asymptotic behavior of the algorithm is studied in Section 3 and reformulated on the behavior of some stopping time \( \tau \). Theorem 2 presents a new representation of the asymptotic oscillations of the algorithm. In Section 4, the distribution of the algorithm cost is investigated.

Using the binary decomposition of the interval \([0,1]\), the exact expression of the distribution of \( H_n \) is established. Proposition 3 is a slight variation of the asymptotic formula given by [Janson and Szpankowski (1997)] for the distribution of the algorithm cost \( H_n \) in the biased case.

1.6 Notations

Throughout this paper, \((t_n)_{n \geq 1}\) is a non decreasing random variables sequence such that

- \( t_1 \) follows an exponential distribution with parameter 1,
- \((t_{n+1} - t_n)\) is a sequence of i.i.d. random variables exponentially distributed with parameter 1.

For \( x \geq 0 \), let \( \mathcal{N}_x \) be the number of \( t_n \) in the interval \([0, x] \). It is a r. v. with Poisson distribution.

2 Average Cost of The Algorithm

2.1 Algorithm cost

The algorithm cost is the number of steps needed to find a leader, or equivalently the height of the associated tree. Denote by \( H_n \) this quantity when the size of the initial group of items is \( n \), then, for \( n \geq 2 \), this random variable verifies a recurrence relation:

\[
H_n \overset{\text{dist.}}{=} 1 + H_{1,S_n} 1_{\{S_n \neq 0\}} + H_{2,n} 1_{\{S_n = 0\}},
\]

with the boundary conditions \( H_0 = H_1 = 0 \), where \((B_i(p))_{1 \leq i \leq n}\) are \( n \) independent Bernoulli variables of parameter \( p \),

\[
S_n = \sum_{i=1}^{n} B_i(p),
\]

for \((m,n) \in \mathbb{N}^2\), \( H_{1,m} \) and \( H_{2,n} \) are independent and, for \( i = 1, 2 \), the variable \( H_{i,m} \) has the same distribution as \( H_m \). So, for \( n \geq 0 \), the recurrence equation for the sequence \( (H_n) \) can be rewritten

\[
H_n \overset{\text{dist.}}{=} 1 + H_{S_n} + H_n 1_{\{S_n = 0\}} - 1_{\{n \leq 1\}}. \tag{2}
\]
2.2 Poissonization

Consider the Poisson model, i.e. the size of the initial group of items is random following a Poisson process $N_α$ of intensity $1$ on the interval $[0,x]$. The following proposition gives a useful representation of the Poisson transform of the average cost of the algorithm.

**Proposition 1** For $x > 0$,

$$E(H_{N_α}) = E \left( \sum_{j=0}^{+\infty} \frac{1}{\pi_j} \cdot 1_{\{a_{j+1} > x, \pi_j \leq x(a_j + \pi_j)\}} \right),$$

where $(A_j, B_j)$ is a sequence of i.i.d. realizations of a couple of random variable $(A, B)$ with distribution

$$P(A = p, B = 0) = p, \quad P(A = q, B = p) = q,$$

$\pi_0 = 1$, $\alpha_0 = 0$ and, for $i \geq 1$,

$$\pi_i = \prod_{j=0}^{i-1} A_j, \quad \alpha_i = \sum_{j=0}^{i-1} \pi_j B_j.$$

**Proof:** Let $h$ the Poisson transform of the average cost (see Definition 2). Then, the recurrence equation (2) for the sequence $(H_n)_{n \geq 0}$ becomes

$$h(x) = h(px) + h(qx) e^{-px} + 1 - (1 + x)e^{-x}.$$

Following the approach of Mohamed and Robert (2005), direct iteration becomes possible using a probabilistic formulation of the last equation as below

$$h(x) = E \left( \frac{h(Ax)}{A} e^{-Bx} \right) + f(x),$$

(3)

where $f(x) = 1 - (1 + x)e^{-x}$ and $(A, B)$ is couple of random variables with distribution

$$P(A = p, B = 0) = p, \quad P(A = q, B = p) = q.$$

Let the sequence of i.i.d realizations $(A_i, B_i)_{i \in \mathbb{N}}$ of the couple of random variables $(A, B)$. We introduce some notations; for $x \geq 0$, $X_0 = x$, $Y_0 = 0$, and for $n \in \mathbb{N},$

$$X_{n+1} = A_n X_n, \quad Y_{n+1} = B_n X_n.$$

By iterations of equation (3), one gets at the $(n+1)^{th}$ stage

$$h(x) = E \left( \frac{h(X_{n+1})}{\prod_{i=0}^n A_i} e^{-\sum_{i=0}^{n+1} Y_i} \right) + E \left( \sum_{i=0}^n e^{-\sum_{j=0}^n Y_j} \frac{f(X_i)}{\prod_{j=0}^{i-1} A_j} \right).$$

Since $h'(0) = 0$ and, almost surely, $\lim_{n \to +\infty} X_{n+1} = 0$, then, one obtains

$$h(x) = E \left( \sum_{i=0}^{+\infty} \frac{1}{\pi_i} (1 - (1 + \pi_i)e^{-\pi_i x}) e^{-\alpha_i x} \right),$$

where $\pi_0 = 1$, $\alpha_0 = 0$ and, for $i \geq 1$,

$$\pi_i = \prod_{j=0}^{i-1} A_j, \quad \alpha_i = \sum_{j=0}^{i-1} \pi_j B_j.$$

As the sequences $(\alpha_i)$ and $(\alpha_i + \pi_i)$ are, almost surely, in the interval $[0, 1]$, the function $h$ can be represented as follows

$$h(x) = E \left( \sum_{i=0}^{+\infty} \frac{1}{\pi_i} 1_{\{t_1 > \alpha_i x, t_2 < (\alpha_i + \pi_i)x\}} \right).$$

(4)

The proposition has been proved. □

From now on, throughout the paper, we conserve the notations introduced in this proof.
2.3 de-Poissonization

The next step is the probabilistic de-Poissonization of \( \tau \) following the method of [Robert (2005)] to obtain the expression of the average cost \( \mathbb{E}(H_n) \).

**Proposition 2 (Probabilistic representation of the average cost)** For \( n \geq 2 \),

\[
\mathbb{E}(H_n) = \mathbb{E} \left( \sum_{i=0}^{\tau(U_{1,n}, U_{2,n}) - 1} \frac{1}{\pi_i} \right),
\]

where, for \( 0 < x < y < 1 \), \( \tau(x, y) = \min(\nu(x); \mu(y)) \) with

\[
\nu(x) = \inf \{ i \geq 1 : \alpha_i > x \},
\]

\[
\mu(y) = \inf \{ i \geq 1 : \alpha_i + \pi_i < y \},
\]

and \( U_{1,n} \) is the \( i \)th smallest variables of \( n \) independent, uniformly distributed random variables on \( [0, 1] \) independent of the sequence \( (A_j, B_j)_{j \geq 0} \).

**Proof:** For \( x > 0 \), by decomposing with respect to the number of points of the Poisson process \( (N_x) \) in the interval \( [0, x] \), one gets, for \( 0 < a < b < 1 \),

\[
\mathbb{P}(t_1 > ax, t_2 < bx) = \sum_{n=2}^{+\infty} \mathbb{P}(t_1 > ax, t_2 < bx, N_x = n) \mathbb{P}(N_x = n).
\]

For \( n \geq 2 \), conditionally on the event \( \{N_x = n\} \), the couple of variables \( (t_1, t_2) \) has the same distribution as the couple \( (axU_{1,n}, bxU_{2,n}) \) of the two smallest random variables of \( n \) uniformly distributed random variables on \( [0, x] \). So, we get the identity

\[
\mathbb{P}(t_1 > ax, t_2 < bx) = \mathbb{E} \left( \sum_{n=2}^{+\infty} 1 \{U_{1,n} > a, U_{2,n} < b\} \frac{x^n}{n!} e^{-x} \right).
\]

Due to the independence of the sequence \( (A_i, B_i) \) and \( (t_1, t_2) \), and using the Fubini’s Theorem, one gets

\[
\mathbb{E}(H_{N_x}) = \sum_{n=2}^{+\infty} \left( \mathbb{E} \left( \sum_{i=0}^{+\infty} 1 \{U_{1,n} > \alpha_i, U_{2,n} < (\alpha_i + \pi_i)\} \right) \right) \frac{x^n}{n!} e^{-x}.
\]

The identification of the representation of the Poisson transform (see Definition 2) \( \mathbb{E}(H_{N_x}) \) and the last identity gives the following formula for \( n \geq 2 \)

\[
\mathbb{E}(H_n) = \mathbb{E} \left( \sum_{i=0}^{+\infty} 1 \{U_{1,n} > \pi_i, U_{2,n} < (\alpha_i + \pi_i)\} \right).
\]

Since, almost surely, the sequence \( (\alpha_i)_{i \geq 0} \) is increasing to a random variable \( \alpha \in [0, 1] \) and the sequence \( (\alpha_i + \pi_i)_{i \geq 0} \) is decreasing to the same random variable, the following equality holds

\[
\{ i \geq 0 : U_{1,n} > \pi_i, U_{2,n} < \alpha_i + \pi_i \} = [0, \tau(U_{1,n}, U_{2,n}) - 1],
\]

where the hitting time \( \tau \) is defined as above. \( \square \)

3 Asymptotic Analysis of The Average Cost

3.1 Two random sequences and one hitting time

It is clear that the key of the analysis of the asymptotic behavior of the algorithm is the hitting time \( \tau \) written on the two random sequences \( (\alpha_i)_{i \geq 0} \) and \( (\alpha_i + \pi_i)_{i \geq 0} \). Let \( (\gamma_i) \) the sequence of random variables defined by

\[
(\gamma_i)_{i \in \mathbb{N}} = \{ j \geq 0 : B_j = p \}.\]
These are the moments of jump of the sequence \((\alpha_i)_{i \geq 0}\) and conversely the instants of stopping for the other one, \((\pi_i + \alpha_i)_{i \geq 0}\). It is clear that these moments can be recursively defined as below: \(\gamma_0 = G_0\) and \(\gamma_{n+1} = 1 + \gamma_n + G_{n+1}\). where \((G_n)_{n \geq 0}\) is a sequence of \(i.i.d\) \(v\) with a geometric distribution \(Geo(q)\)

\[
P(\text{Geo}(q) = k) = q^k p^k.
\]

So, it is easy to see that

\[
\nu(x) \in \{1 + \gamma_i : i \in \mathbb{N}\}, \quad \mu(y) \notin \{1 + \gamma_i : i \in \mathbb{N}\}.
\]

Using a discussion on the position of the hitting time \(\tau\) in comparison with the sequence \(\gamma\), we establish the following lemma which will be proved in the Appendix 5.

**Lemma 1**

\[
\mathbb{E} \left( \sum_{i=0}^{\tau(x,y)-1} \frac{1}{\pi_i} \right) = [\log_p(y)] + \left( [\log_p(\rho(\log_p(y))y)] - [\log_p(y)] \right) \mathbf{1}_{\Omega(x,y)}
\]

\[
+ \mathbb{E} \left( \sum_{i=1+\lceil \log_p(\rho(\log_p(y))y) \rceil}^{\tau(x,y)-1} \frac{1}{\pi_i} \mathbf{1}_{\{\gamma_0=\lceil \log_p(\rho(\log_p(y))y) \rceil; \gamma_1=\lceil \log_p(\rho(\log_p(y))y) \rceil\}} \right) \mathbf{1}_{\Omega(x,y)}.
\]

where \(\Omega(x,y) = \{(x,y) \in (0,1]^2 : [\log_p(y)] = [\log_p(x)]\}\) and \(\rho\) is a periodic function with magnitude 1 defined for \(z > 0\) by

\[
\rho(z) = \frac{1 - p^{1-(z)}}{1-p}, \quad \{z\} = z - \lfloor z \rfloor \text{ is the fractional part of } z.
\]

### 3.2 Asymptotic fluctuations phenomena

**Theorem 1 (Asymptotic behavior of the average cost)** The average cost \(\mathbb{E}(H_n)\) admits the following asymptotic formula

\[
\mathbb{E}(H_n) = - \log_p(\gamma_0) + \mathbb{E} \left( [\log_p(t_2)] \right) + F(\log_p(\gamma_0)) + R(n),
\]

where \(F\) is a periodic function defined for all \(z > 0\) by

\[
F(z) = \int_0^\infty y(1 - p^{1-(\log_p(y-z))}) \left( \log_p \left( 1 - \frac{p^{1-(\log_p(y-z))}}{1-p} \right) + \log_p \left( y - z \right) - \log_p \left( y - z \right) \right) e^{-y} dy,
\]

\[
\Omega_n = \Omega(U_{1,n}, U_{2,n}) \text{ and } R(n) \text{ is a rest discussed in Section 3.3 defined by}
\]

\[
R(n) = \mathbb{E} \left( \sum_{i=1+\lceil \log_p(\rho(\log_p(y))y) \rceil}^{\tau(U_{1,n}, U_{2,n}) - 1} \frac{1}{\pi_i} \mathbf{1}_{\{\gamma_0=\lceil \log_p(\rho(\log_p(y))y) \rceil; \gamma_1=\lceil \log_p(\rho(\log_p(y))y) \rceil\}} \right) \mathbf{1}_{\Omega_n}.
\]

**Proof:** Using Lemma 1, one gets

\[
\mathbb{E}(H_n) = \mathbb{E} \left( [\log_p(U_{2,n})] \right) + \mathbb{E} \left( [\log_p(\rho[\log_p(U_{2,n})U_{2,n}])] - [\log_p(U_{2,n})] \right) \mathbf{1}_{\Omega_n}
\]

\[
+ \mathbb{E} \left( \sum_{i=1+\lceil \log_p(\rho(\log_p(y))y) \rceil}^{\tau(U_{1,n}, U_{2,n}) - 1} \frac{1}{\pi_i} \mathbf{1}_{\{\gamma_0=\lceil \log_p(\rho(\log_p(y))y) \rceil; \gamma_1=\lceil \log_p(\rho(\log_p(y))y) \rceil\}} \right) \mathbf{1}_{\Omega_n}.
\]

The only not neglect terms are

\[
T_1(n) = \mathbb{E} \left( [\log_p(U_{2,n})] \right) \text{ and } T_2(n) = \mathbb{E} \left( [\log_p(\rho[\log_p(U_{2,n})U_{2,n}])] - [\log_p(U_{2,n})] \right) \mathbf{1}_{\Omega_n}.
\]

As \(n\) goes to infinity, \(nU_{2,n}\) converges in distribution to a random variable \(t_2\) which is a sum of two \(i.i.d\) exponential random variables with parameter 1. Then, the first term satisfies

\[
T_1(n) = \mathbb{E}([\log_p(t_2) - \log_p(\gamma_0)] + O\left(\frac{1}{n}\right).
\]
Let $D$, function of $-\log_p(n)$, the difference
\[
D(-\log_p(n)) = \mathbb{E}(\|\log_p(t_2) - \log_p(n)\|) - (\mathbb{E}(\|\log_p(t_2)\|) - \log_p(n))
\]
It is easy to check that $D(z) = D\{\{z\}\} - \lfloor z \rfloor$, then $\lim_{n \to +\infty} n D(-\log_p n) = \lim_{n \to +\infty} p^{-z} D(z) = 0$, and one gets
\[
T_1(n) = -\log_p(n) + \mathbb{E}(\|\log_p(t_2)\|) + O\left(\frac{1}{n}\right).
\]
The last term $T_2(n)$ is asymptotically equivalent to $F(\log_p(n))$ where $F$ is defined by \[5\] In fact
\[
|F(\log_p(n)) - T_2(n)| \leq \int_0^1 (1 - \frac{y}{n})n^{-2} e^{-y} dy + \int_1^\infty \log_p(\rho(\log_p(y/n))) y e^{-y} dy + \frac{1}{n} F(\log_p(n)) + 2e^{-n}
\]
Observe that
\[
\int_n^\infty \log_p(\rho(\log_p(y))) y e^{-y} dy = n^2 \int_1^\infty \log_p(\rho(\log_p(y))) y e^{-ny} dy.
\]
By decomposition on the sequence of intervals $[\lfloor p^{k+1}, p^k \rfloor]$, the last integral is dominated by a geometric sum and the following inequality holds for $n > 2$
\[
\int_1^\infty \log_p(\rho(\log_p(y))) y e^{-ny} dy \leq \frac{p^{n-2}}{1-p}.
\]
Then,
\[
F(\log_p(n)) - \mathbb{E} \left( \left[ \|\log_p(\rho(\log_p(U_{2,n}))U_{2,n})\| - \|\log_p(U_{2,n})\| \right] \mathbf{1}_{\Omega_n} \right) = O\left(\frac{1}{n}\right).
\]
This ends the proof. \(\square\)

### 3.3 Estimation of the rest

The final step is to estimate the rest $\mathcal{R}(n)$ defined by \[6\]. For $x, y \in [0, 1]$, $K > k > 0$
\[
\mathbb{E} \left( \sum_{i=k}^{\tau(x,y)-1} \frac{1}{\pi_i} \mathbf{1}_{\{\gamma_0 = k; \gamma_1 = K\}} \right) \leq (1 - \delta)^K \sqrt{\mathbb{E} \left( \frac{1}{\delta^2} \tau(x,y) \right)}
\]
where $\delta = \min(p, q)$. The following result is admitted.

**Conjecture 1** The hitting time $\tau$ satisfies
\[
\sup_{x \in [0, 1]} \mathbb{E} \left( \frac{1}{\delta^2} \tau(x,x) \right) < \infty.
\]

**Remark 1** Conjecture \[7\] is an intuitive restriction on the exponential moment of the hitting time $\tau$. It is supported by some simulations (Fig. \[2\]) of $x \to \mathbb{E} \left( \frac{1}{\delta^2} \tau(x,x) \right)$ using Monte-Carlo techniques. Observe that, for the unbiased case (Fig. \[2\]), the maximum corresponds to numerical values of $x$ around 0.5 which is, on average, the limit $\alpha$ of the two random sequences $(\alpha_k)$ and $(\alpha_k + \pi_k)$. This maximum is of order of $10^{14}$, which is reasonable since it implies that
\[
\mathbb{E}(\tau) \leq 14 \log_2(10) \approx 23.25
\]
For the biased case (Fig. \[3\]), since $\delta = 0.2$, a maximum of the order of $10^{80}$ is acceptable; $\mathbb{E}(\tau) \leq 57.22$.

Since, for $0 \leq x < y \leq 1$, $\tau(x,y) \leq \max(\tau(x,x), \tau(y,y))$, then, using Conjecture \[1\] we obtain
\[
\mathcal{R}(n) \leq C \mathbb{E} \left( (1 - \delta)^{\|\log_p(\rho(U_{2,n})U_{2,n})\|} \right),
\]
where $C = \sup_{x \in [0, 1]} \sqrt{\mathbb{E} \left( \frac{1}{\delta^2} \tau(x,x) \right)}$. Using the same method as for the function $F$, one gets
\[
\mathbb{E} \left( (1 - \delta)^{\|\log_p(\rho(U_{2,n})U_{2,n})\|} \right) \leq \mathbb{E} \left( (1 - \delta)^{\|\log_p(U_{2,n})\|} \right) \sim \left(\frac{1}{n}\right)^{\log_p(1-\delta)}.
\]
This gives
\[
\mathcal{R}(n) = O\left(\frac{1}{n^{\log_p(1-\delta)}}\right).
\]
Conclusion
\[
\mathbb{E}(H_n) = -\log_p(n) + \mathbb{E} \left( \|\log_p(t_2)\| \right) + F(\log_p(n)) + O\left(\frac{1}{n^{\log_p(1-\delta)}}\right).
\]
A probabilistic analysis of a leader election algorithm

Fig. 2: Unbiased case: simulations of $x \rightarrow E \left( 4^{\tau(x,x)} \right)$.

Fig. 3: Biased case: simulations of $x \rightarrow E \left( \frac{1}{4^\tau(x,x)} \right)$.
4 Algorithm Cost Distribution

It is more appropriate to use these notations $p_0 = p$, $p_1 = q$, to define recursively the sequence of intervals $(I_k^n)$ associated to the binary decomposition of the interval $[0, 1]$ in the base $(p_0, p_1)$

\[
\begin{align*}
I_k^0 &= [0, 1] \\
I_k^{n+1} &= (I_{k-1}^{n+1})_+ + p_{k-2[2k/2]} I_{k/2}^n,
\end{align*}
\]

where $(I)_+$ denotes the right extremity of the interval $I$. Let $|I|$ the length of the interval $I$, then

\[
(I_k^n)_+ = \sum_{i=0}^{k} |I_k^n|.
\]

Let $n \in \mathbb{N}$ and $0 \leq k \leq 2^{n+1} - 1$. Consider the binary decomposition of $k$ at the stage $n$

\[
k = a_0 + a_1 2 + \ldots + a_n 2^n, \quad \text{for } 0 \leq i \leq n, \ a_i \in \{0, 1\}.
\]

Then, the length of the interval $I_k^{n+1}$ is

\[
|I_k^{n+1}| = \prod_{i=0}^{n} p_{a_i}.
\]

For $k \in \mathbb{N}$, $x > 0$, one gets the following identity

\[
\{H_{N_x} > k\} = \{\exists 0 \leq i < 2^k : N(xI_k^i) = \ldots = N(xI_{k-1}^i) = 0, \ N(xI_k^i) \geq 2\}.
\]

So

\[
\mathbb{P}(H_{N_x} \leq k) = e^{-x} + x \sum_{i=0}^{2^k-1} |I_k^{i}| e^{-(I_k^{i})_+}.
\]  

Let us define the sequence of probability measures $(\mu_k)$ by

\[
\mu_k(t) = \sum_{i=0}^{2^k-1} |I_k^{i}| \delta_{(I_k^{i})_+}(t).
\]

Then, equation (7) can be rewritten as

\[
\mathbb{P}(H_{N_x} \leq k) = e^{-x} + x \int_0^1 e^{-xt} d\mu_k(t).
\]  

Using a probabilistic de-Poissonization of equation (8) as done for Proposition 2, we obtain the exact distribution of $H_n$.

**Proposition 3** For $n \geq 2$,

\[
\mathbb{P}(H_n \leq k) = n \int_0^1 (1 - t)^{n-1} d\mu_k(t),
\]

where the probability measure $\mu_k$ is described as above.

Using this identity

\[
1 - n(1 - t)^{n-1} = 1 - nt (1 - t)^{n-1} - n(1 - t)^n,
\]

the following result is immediate.

**Corollary 1** For $k \in \mathbb{N}$,

\[
\mathbb{P}(H_n > k) \sim \int_0^1 \mathbb{P}(U_{2,n} < t) \ d\mu_k(t), \text{ as } n \text{ goes to infinity}.
\]

where $U_{2,n}$ is the second smallest random variable of $n$ uniformly distributed random variables on $[0, 1]$. 
5 Appendix

We present the proof of Lemma 1. Recall the sequence of random variables $(\gamma_i)$ defined by

$$(\gamma_i)_{i \in \mathbb{N}} = \{j \geq 0 : B_j = p\}.$$  

**Proof of Lemma 1**: First, note that

$$(\nu(x) \geq 2 + \gamma_0) \Leftrightarrow (\gamma_0 \geq \lfloor \log_p(x) \rfloor)$$

$$(\mu(y) \leq \gamma_0) \Leftrightarrow (\gamma_0 \geq \lfloor \log_p(y) \rfloor).$$

Denote by $\Omega_0$ the following set

$$\Omega_0 = \Omega(x,y) := \{(x,y) \in (0,1]^2 : \lfloor \log_p(y) \rfloor = \lfloor \log_p(x) \rfloor\}.$$  

By decomposing the function $\Phi$ with respect to $\Omega_0$, one gets this formula

$$E\left( \sum_{i=0}^{\tau(x,y)-1} \frac{1}{\pi_i} \right) = E\left( \sum_{i=0}^{\gamma_0} \frac{1}{p^i} 1_{\{\gamma_0 < \lfloor \log_p(y) \rfloor\}} \right)$$

$$+ \mathbf{P} \left( \gamma_0 = \lfloor \log_p(y) \rfloor \right) \sum_{i=0}^{\lfloor \log_p(y) \rfloor} \frac{1}{p^i} + E\left( \sum_{i=0}^{\mu(y)-1} \frac{1}{p^i} 1_{\{\gamma_0 \geq \lfloor \log_p(y) \rfloor\}} \right)$$

$$+ E\left( \sum_{i=\lfloor \log_p(y) \rfloor}^{\lfloor \log_p(y) \rfloor-1} \frac{1}{p^i} 1_{\{\gamma_0 = \lfloor \log_p(y) \rfloor\}} \right) 1_{\Omega_0}.$$  

Since

$$\mu(y)(\gamma_0 \geq \lfloor \log_p(y) \rfloor) = \inf \{i \geq 1, p^i < y \} = \lfloor \log_p(y) \rfloor,$$

then, by simple calculations, one gets

$$E\left( \sum_{i=0}^{\tau(x,y)-1} \frac{1}{\pi_i} \right) = \lfloor \log_p(y) \rfloor + E\left( \sum_{i=\lfloor \log_p(y) \rfloor}^{\tau(x,y)-1} \frac{1}{\pi_i} 1_{\{\gamma_0 = \lfloor \log_p(y) \rfloor\}} \right) 1_{\Omega_0}.$$  

A second discussion on $\gamma_1$ implies that, on the set $(\gamma_0 = \lfloor \log_p(y) \rfloor, \Omega_0),$

$$(\nu(x) \geq 2 + \gamma_1) \Leftrightarrow (\gamma_1 \geq \lfloor \log_p(\rho(\log_p(x))x) \rfloor)$$

$$(\mu(y) \leq \gamma_1) \Leftrightarrow (\gamma_1 \geq 1 + \lfloor \log_p(\rho(\log_p(y)y) \rfloor),$$

where $\rho$ is a periodic function with magnitude 1 defined by $\rho(z) = (1 - p^{1-z})/(1 - p).$ Moreover, $\rho$ is decreasing on $[0,1]$, so on the set $\Omega_0,$

$$\rho(\log_p(x)x) < \rho(\log_p(y)y).$$

Let $\Omega_1 = \Omega(\rho(\log_p(x)x), \rho(\log_p(y)y)).$ Then

$$E\left( \sum_{i=0}^{\tau(x,y)-1} \frac{1}{\pi_i} \right) = \lfloor \log_p(y) \rfloor + \lfloor \log_p(\rho(\log_p(y)y)) \rfloor - \lfloor \log_p(y) \rfloor \right) 1_{\Omega(x,y)}$$

$$+ E\left( \sum_{i=1+\lfloor \log_p(\rho(\log_p(y)y) \rfloor}^{\tau(x,y)-1} \frac{1}{\pi_i} 1_{\{\gamma_0 = \lfloor \log_p(y) \rfloor; \gamma_1 = \lfloor \log_p(\rho(\log_p(y)y) \rfloor\}} \right) 1_{\Omega(x,y)}.$$  

This ends the proof.

### Acknowledgements

I wish to thank Philippe Robert for his very useful discussions and recommendations, Christine Fricker for her comments on a preliminary version and Mohamed Kamel Eddine Mrad for numerical simulations and graphs.
References


