

Ascending Runs in Dependent Uniformly Distributed Random Variables: Application to Wireless Networks

Nathalie Mitton · Katy Paroux · Bruno Sericola · Sébastien Tixeuil

Received: 19 February 2008 / Revised: 3 June 2008 /
Accepted: 6 June 2008 / Published online: 8 July 2008
© Springer Science + Business Media, LLC 2008

Abstract We analyze in this paper the longest increasing contiguous sequence or maximal ascending run of random variables with common uniform distribution but not independent. Their dependence is characterized by the fact that two successive random variables cannot take the same value. Using a Markov chain approach, we study the distribution of the maximal ascending run and we develop an algorithm to compute it. This problem comes from the analysis of several self-organizing protocols designed for large-scale wireless sensor networks, and we show how our results apply to this domain.

N. Mitton
Laboratoire d'Informatique fondamentale, UMR CNRS 8022, INRIA Lille, Nord Europe,
Université de Lille 1, 59650 Villeneuve d'Asq cedex, France
e-mail: nathalie.mitton@inria.fr

K. Paroux (✉)
Laboratoire de Mathématiques, UMR CNRS 6623,
Université de Franche-Comté, 25030 Besançon cedex, France
e-mail: katy.paroux@univ-fcomte.fr

B. Sericola
INRIA Rennes—Bretagne Atlantique,
35042 Campus de Beaulieu, Rennes cedex, France
e-mail: bruno.sericola@inria.fr

S. Tixeuil
Laboratoire d'Informatique de Paris VI, UMR CNRS 7606,
INRIA Saclay, Ile de France, Université Paris VI, 75005 Paris, France
e-mail: sebastien.tixeuil@lri.fr

Present Address:

K. Paroux
INRIA Rennes—Bretagne Atlantique,
35042 Campus de Beaulieu, Rennes cedex, France
e-mail: katy.paroux@inria.fr

Keywords Markov chains · Maximal ascending run · Self-stabilization · Convergence time

PACS G.3 · G.2 · C.2

Mathematics Subject Classifications (2000) 60J10 · 60J20 · 68R05

1 Introduction

Let $X = (X_n)_{n \geq 1}$ be a sequence of identically distributed random variables on the set $S = \{1, \dots, m\}$. As in Louchard (2002), we define an ascending run as a contiguous and increasing subsequence in the process X . For instance, with $m = 5$, among the 20 first following values of X : 23124342313451234341, there are 8 ascending runs and the length of maximal ascending run is 4. More formally, an ascending run of length $\ell \geq 1$, starting at position $k \geq 1$, is a subsequence $(X_k, X_{k+1}, \dots, X_{k+\ell-1})$ such that

$$X_{k-1} > X_k < X_{k+1} < \dots < X_{k+\ell-1} > X_{k+\ell},$$

where we set $X_0 = \infty$ in order to avoid special cases at the boundary. Under the assumption that the distribution is discrete and the random variables are independent, several authors have studied the behaviour of the maximal ascending run, as well as the longest non-decreasing contiguous subsequence. The main results concern the asymptotic behaviour of these quantities when the number of random variables tends to infinity, see for example Frolov and Martikainen (1999) and Eryilmaz (2006) and the references therein. Note that these two notions coincide when the common distribution is continuous. In this case, the asymptotic behaviour is known and does not depend on the distribution, as shown in Frolov and Martikainen (1999).

We denote by M_n the length of the maximal ascending run among the first n random variables. The asymptotic behaviour of M_n hardly depends on the common distribution of the random variables X_k , $k \geq 1$. Some results have been established for the geometric distribution in Louchard and Prodinger (2003) where an equivalent of the law of M_n is provided and previously in Csaki and Foldes (1996) where the almost-sure convergence is studied, as well as for Poisson distribution.

In Louchard (2005), the case of the uniform distribution on the set $\{1, \dots, s\}$ is investigated. The author considers the problem of the longest non-decreasing contiguous subsequence and gives an equivalent of its law when n is large and s is fixed. The asymptotic equivalent of $\mathbb{E}(M_n)$ is also conjectured.

Others related problems such as waiting time distributions for fixed simple and compound patterns have been studied in Fu and Chang (2002) and Fu and Lou (2003) using probability generating functions.

In this paper, we consider a sequence $X = (X_n)_{n \geq 1}$ of integer random variables on the set $S = \{1, \dots, m\}$, with $m \geq 2$. The random variable X_1 is uniformly distributed on S and, for $n \geq 2$, X_n is uniformly distributed on S with the constraint $X_n \neq X_{n-1}$. This process may be seen as the drawing of balls, numbered from 1 to m in an urn where at each step the last ball drawn is kept outside the urn. Thus we have, for every $i, j \in S$ and $n \geq 1$,

$$\mathbb{P}(X_1 = i) = \frac{1}{m} \text{ and } \mathbb{P}(X_n = j | X_{n-1} = i) = \frac{1_{\{i \neq j\}}}{m-1}.$$

By induction over n and unconditioning, we get, for every $n \geq 1$ and $i \in S$,

$$\mathbb{P}(X_n = i) = \frac{1}{m}.$$

Hence the random variables X_n are uniformly distributed on S but are not independent. Using a Markov chain approach different from that of Fu and Chang (2002), we study the distribution of the maximal ascending run and we develop an algorithm to compute it. This problem comes from the analysis of self-organizing protocols designed for large-scale wireless sensor networks, and we show how our results apply to this domain.

The remainder of the paper is organized as follows. In the next section, we use a Markov chain approach to study the behavior of the sequence of ascending runs in the process X . In Section 3, we analyze the hitting times of an ascending run of fixed length and we obtain the distribution of the maximal ascending M_n over the n first random variables X_1, \dots, X_n using a Markov renewal argument. An algorithm to compute this distribution is developed in Section 4 and Section 5 is devoted to the practical implications of this work in large-scale wireless sensor networks.

2 Associated Markov Chain

The process X is obviously a Markov chain on S . As observed in Louchard and Prodinger (2003), we can see the ascending runs as a discrete-time process having two components: the value taken by the first element of the ascending run and its length. We denote this process by $Y = (V_k, L_k)_{k \geq 1}$, where V_k is the value of the first element of the k th ascending run and L_k is its length. Going over the example used in the Introduction, we obtain

$$(V_1, L_1) = (2, 2), (V_2, L_2) = (1, 3), (V_3, L_3) = (3, 2), (V_4, L_4) = (2, 2), \dots$$

The state space of Y is a subset S^2 we shall precise now.

Only the first ascending run can start with the value m . Indeed, as soon as $k \geq 2$, the random variable V_k takes its values in $\{1, \dots, m - 1\}$. Moreover $V_1 = X_1 = m$ implies that $L_1 = 1$. Thus, for any $\ell \geq 2$, (m, ℓ) is not a state of Y whereas $(m, 1)$ is only an initial state that Y will never visit again.

We observe also that if $V_k = 1$ then necessarily $L_k \geq 2$, which implies that $(1, 1)$ is not a state of Y . Moreover $V_k = i$ implies that $L_k \leq m - i + 1$.

According to this behaviour, we have

$$Y_1 \in E \cup \{(m, 1)\} \text{ and for } k \geq 2, Y_k \in E,$$

where

$$E = \{(i, \ell) \mid 1 \leq i \leq m - 1 \text{ and } 1 \leq \ell \leq m - i + 1\} \setminus \{(1, 1)\}.$$

We define the following useful quantities for any $i, j, \ell \in S$ and $k \geq 1$:

$$\Phi_\ell(i, j) = \mathbb{P}(V_{k+1} = j, L_k = \ell \mid V_k = i), \tag{1}$$

$$\varphi_\ell(i) = \mathbb{P}(L_k = \ell \mid V_k = i), \tag{2}$$

$$\psi_\ell(i) = \mathbb{P}(L_k \geq \ell \mid V_k = i). \tag{3}$$

Theorem 1 *The process Y is a homogeneous Markov chain with transition probability matrix P , which entries are given for any $(i, \ell) \in E \cup \{(m, 1)\}$ and $(j, \lambda) \in E \cup \{(m, 1)\}$ by*

$$P_{(i,\ell),(j,\lambda)} = \frac{\Phi_\ell(i, j)\varphi_\lambda(j)}{\varphi_\ell(i)}.$$

Proof We exploit the Markov property of X , rewriting events for Y as events for X . For every $(j, \lambda) \in E$ and taking $k \geq 1$ then for any $(v_k, \ell_k), \dots, (v_1, \ell_1) \in E \cup \{(m, 1)\}$, we denote by A_k the event :

$$A_k = \{Y_k = (v_k, \ell_k), \dots, Y_1 = (v_1, \ell_1)\}.$$

We have to check that

$$\mathbb{P}(Y_{k+1} = (j, \lambda) | A_k) = \mathbb{P}(Y_2 = (j, \lambda) | Y_1 = (v_k, \ell_k)).$$

First, we observe that

$$A_1 = \{Y_1 = (v_1, \ell_1)\} = \{X_1 = v_1 < \dots < X_{\ell_1} > X_{\ell_1+1}\},$$

and

$$\begin{aligned} A_2 &= \{Y_2 = (v_2, \ell_2), Y_1 = (v_1, \ell_1)\} \\ &= \{X_1 = v_1 < \dots < X_{\ell_1} > X_{\ell_1+1} = v_2 < \dots < X_{\ell_1+\ell_2} > X_{\ell_1+\ell_2+1}\} \\ &= A_1 \cap \{X_{\ell_1+1} = v_2 < \dots < X_{\ell_1+\ell_2} > X_{\ell_1+\ell_2+1}\}. \end{aligned}$$

By induction, we obtain

$$A_k = A_{k-1} \cap \{X_{\ell(k-1)+1} = v_k < \dots < X_{\ell(k)} > X_{\ell(k)+1}\},$$

where $\ell(k) = \ell_1 + \dots + \ell_k$. Using this remark and the fact that X is a homogeneous Markov chain, we get

$$\begin{aligned} \mathbb{P}(Y_{k+1} = (j, \lambda) | A_k) &= \mathbb{P}(V_{k+1} = j, L_{k+1} = \lambda | A_k) \\ &= \mathbb{P}(X_{\ell(k)+1} = j < \dots < X_{\ell(k)+\lambda} > X_{\ell(k)+\lambda+1} | \\ &\quad X_{\ell(k-1)+1} = v_k < \dots < X_{\ell(k)} > X_{\ell(k)+1}, A_{k-1}) \\ &= \mathbb{P}(X_{\ell(k)+1} = j < \dots < X_{\ell(k)+\lambda} > X_{\ell(k)+\lambda+1} | \\ &\quad X_{\ell(k-1)+1} = v_k < \dots < X_{\ell(k)} > X_{\ell(k)+1}) \\ &= \mathbb{P}(X_{\ell_k+1} = j < \dots < X_{\ell_k+\lambda} > X_{\ell_k+\lambda+1} | \\ &\quad X_1 = v_k < \dots < X_{\ell_k} > X_{\ell_k+1}) \\ &= \mathbb{P}(V_2 = j, L_2 = \lambda | V_1 = v_k, L_1 = \ell_k) \\ &= \mathbb{P}(Y_2 = (j, \lambda) | Y_1 = (v_k, \ell_k)). \end{aligned}$$

We now have to show that

$$\mathbb{P}(Y_{k+1} = (j, \lambda) | Y_k = (v_k, \ell_k)) = \mathbb{P}(Y_2 = (j, \lambda) | Y_1 = (v_k, \ell_k)).$$

Using the previous result, we have

$$\begin{aligned}
 \mathbb{P}(Y_{k+1} = (j, \lambda) \mid Y_k = (v_k, \ell_k)) &= \frac{\mathbb{P}(Y_{k+1} = (j, \lambda), Y_k = (v_k, \ell_k))}{\mathbb{P}(Y_k = (v_k, \ell_k))} \\
 &= \frac{\sum_{\substack{(v_i, \ell_i) \in E \\ i=1,2,\dots,k-1}} \mathbb{P}(Y_{k+1} = (j, \lambda), Y_k = (v_k, \ell_k), A_{k-1})}{\sum_{\substack{(v_i, \ell_i) \in E \\ i=1,2,\dots,k-1}} \mathbb{P}(Y_k = (v_k, \ell_k), A_{k-1})} \\
 &= \frac{\sum_{\substack{(v_i, \ell_i) \in E \\ i=1,2,\dots,k-1}} \mathbb{P}(Y_{k+1} = (j, \lambda) \mid A_k) \mathbb{P}(A_k)}{\sum_{\substack{(v_i, \ell_i) \in E \\ i=1,2,\dots,k-1}} \mathbb{P}(A_k)} \\
 &= \mathbb{P}(Y_2 = (j, \lambda) \mid Y_1 = (v_k, \ell_k)).
 \end{aligned}$$

We have shown that Y is a homogeneous Markov chain over its state space. The entries of matrix P are then given, for every $(j, \lambda) \in E$ and $(i, \ell) \in E \cup \{(m, 1)\}$ by

$$\begin{aligned}
 P_{(i,\ell),(j,\lambda)} &= \mathbb{P}(V_{k+1} = j, L_{k+1} = \lambda \mid V_k = i, L_k = \ell) \\
 &= \mathbb{P}(V_{k+1} = j \mid V_k = i, L_k = \ell) \mathbb{P}(L_{k+1} = \lambda \mid V_{k+1} = j, V_k = i, L_k = \ell) \\
 &= \mathbb{P}(V_{k+1} = j \mid V_k = i, L_k = \ell) \mathbb{P}(L_{k+1} = \lambda \mid V_{k+1} = j) \\
 &= \frac{\mathbb{P}(V_{k+1} = j, L_k = \ell \mid V_k = i)}{\mathbb{P}(L_k = \ell \mid V_k = i)} \varphi_\lambda(j) \\
 &= \frac{\Phi_\ell(i, j) \varphi_\lambda(j)}{\varphi_\ell(i)},
 \end{aligned}$$

where the third equality follows from the Markov property. □

We give the expressions of $\varphi_\lambda(j)$ and $\Phi_\ell(i, j)$ for every $i, j, \ell \in S$ in the following lemma.

Lemma 1 *For every $i, j, \ell \in S$, we have*

$$\begin{aligned}
 \Phi_\ell(i, j) &= \frac{\binom{m-i}{\ell-1}}{(m-1)^\ell} 1_{\{m-i \geq \ell-1\}} - \frac{\binom{j-i}{\ell-1}}{(m-1)^\ell} 1_{\{j-i \geq \ell-1\}}, \\
 \psi_\ell(i) &= \frac{\binom{m-i}{\ell-1}}{(m-1)^{\ell-1}} 1_{\{m-i \geq \ell-1\}}, \\
 \varphi_\ell(i) &= \frac{\binom{m-i}{\ell-1}}{(m-1)^{\ell-1}} 1_{\{m-i \geq \ell-1\}} - \frac{\binom{m-i}{\ell}}{(m-1)^\ell} 1_{\{m-i \geq \ell\}}.
 \end{aligned}$$

Proof For every $i, j, \ell \in \mathcal{S}$, it is easily checked that $\Phi_\ell(i, j) = 0$ if $m < i + \ell - 1$. If $m \geq i + \ell - 1$, we have

$$\begin{aligned} \Phi_\ell(i, j) &= \mathbb{P}(V_2 = j, L_1 = \ell | V_1 = i) \\ &= \mathbb{P}(i < X_2 < \dots < X_\ell > X_{\ell+1} = j | X_1 = i) \\ &= \mathbb{P}(i < X_2 < \dots < X_\ell, X_{\ell+1} = j | X_1 = i) \\ &\quad - \mathbb{P}(i < X_2 < \dots < X_\ell < X_{\ell+1} = j | X_1 = i) 1_{\{j > i + \ell - 1\}}. \end{aligned} \tag{4}$$

We introduce the sets $G_1(i, j, \ell, m)$, $G_2(i, j, \ell, m)$, $G(i, \ell, m)$ and $H(\ell, m)$ defined by

$$\begin{aligned} G_1(i, j, \ell, m) &= \{(x_2, \dots, x_{\ell+1}) \in \{i + 1, \dots, m\}^\ell; x_2 < \dots < x_\ell \neq x_{\ell+1} = j\}, \\ G_2(i, j, \ell, m) &= \{(x_2, \dots, x_{\ell+1}) \in \{i + 1, \dots, m\}^\ell; x_2 < \dots < x_\ell = x_{\ell+1} = j\}, \\ G(i, \ell, m) &= \{(x_2, \dots, x_\ell) \in \{i + 1, \dots, m\}^{\ell-1}; x_2 < \dots < x_\ell\}, \\ H(\ell, m) &= \{(x_2, \dots, x_{\ell+1}) \in \{1, \dots, m\}^\ell; i \neq x_2 \neq \dots \neq x_{\ell+1}\}. \end{aligned}$$

It is well-known, see for instance Foata and Fuchs (1996), that

$$|G(i, \ell, m)| = \binom{m-i}{\ell-1}.$$

Since $|G_2(i, j, \ell, m)| = |G(i, \ell - 1, j - 1)|$, the first term in Eq. 4 can be written as

$$\begin{aligned} \mathbb{P}(i < X_2 < \dots < X_\ell, X_{\ell+1} = j | X_1 = i) &= \frac{|G_1(i, j, \ell, m)|}{|H(\ell, m)|} \\ &= \frac{|G(i, \ell, m)| - |G_2(i, j, \ell, m)|}{|H(\ell, m)|} \\ &= \frac{|G(i, \ell, m)| - |G(i, \ell - 1, j - 1)|}{|H(\ell, m)|} \\ &= \frac{\binom{m-i}{\ell-1} - \binom{j-i-1}{\ell-2} 1_{\{j-i \geq \ell-1\}}}{(m-1)^\ell}. \end{aligned}$$

The second term is given, for $j > i + \ell - 1$, by

$$\mathbb{P}(i < X_2 < \dots < X_\ell < X_{\ell+1} = j | X_1 = i) = \frac{|G(i, \ell, j-1)|}{|H(\ell, m)|} = \frac{\binom{j-i-1}{\ell-1}}{(m-1)^\ell}.$$

Adding these two terms, we get

$$\begin{aligned} \Phi_\ell(i, j) &= \frac{\binom{m-i}{\ell-1} 1_{\{m-i \geq \ell-1\}} - \binom{j-i-1}{\ell-2} 1_{\{j-i \geq \ell-1\}} - \binom{j-i-1}{\ell-1} 1_{\{j-i \geq \ell\}}}{(m-1)^\ell} \\ &= \frac{\binom{m-i}{\ell-1} 1_{\{m-i \geq \ell-1\}} - \binom{j-i}{\ell-1} 1_{\{j-i \geq \ell-1\}}}{(m-1)^\ell}, \end{aligned}$$

which completes the proof of the first relation.

The second relation follows from expression (3) by writing

$$\begin{aligned} \psi_\ell(i) &= \mathbb{P}(L_1 \geq \ell | V_1 = i) \\ &= \mathbb{P}(i < X_2 < \dots < X_\ell | X_1 = i) 1_{\{m-i \geq \ell-1\}} \\ &= \frac{|G(i, \ell, m)|}{|H(\ell - 1, m)|} \\ &= \frac{\binom{m-i}{\ell-1}}{(m-1)^{\ell-1}} 1_{\{m-i \geq \ell-1\}}. \end{aligned}$$

The third relation follows from definition (2) by writing $\varphi_\ell(i) = \psi_\ell(i) - \psi_{\ell+1}(i)$. \square

Note that the matrix Φ defined by

$$\Phi = \sum_{\ell=1}^m \Phi_\ell$$

is obviously a stochastic matrix, which means that, for every $i = 1, \dots, m$, we have

$$\sum_{\ell=1}^m \varphi_\ell(i) = 1,$$

and

$$\sum_{\ell=1}^m \sum_{j=1}^m \Phi_\ell(i, j) = \sum_{\ell=1}^m \varphi_\ell(i) = 1.$$

3 Hitting Times and Maximal Ascending Run

For every $r = 1, \dots, m$, we denote by T_r the hitting time of an ascending run of length at least equal to r . More formally, we have

$$T_r = \inf\{k \geq r; X_{k-r+1} < \dots < X_k\}.$$

It is easy to check that we have $T_1 = 1$ and $T_r \geq r$. The distribution of T_r is given by the following theorem.

Theorem 2 For $2 \leq r \leq m$, we have

$$\mathbb{P}(T_r \leq n | V_1 = i) = \begin{cases} 0 & \text{if } 1 \leq n \leq r - 1 \\ \psi_r(i) + \sum_{\ell=1}^{r-1} \sum_{j=1}^m \Phi_\ell(i, j) \mathbb{P}(T_r \leq n - \ell | V_1 = j) & \text{if } n \geq r. \end{cases} \tag{5}$$

Proof Since $T_r \geq r$, we have, for $1 \leq n \leq r - 1$,

$$\mathbb{P}(T_r \leq n | V_1 = i) = 0.$$

Let us assume from now that $n \geq r$. Since $L_1 \geq r$ implies that $T_r = r$, we get

$$\mathbb{P}(T_r \leq n, L_1 \geq r | V_1 = i) = \mathbb{P}(L_1 \geq r | V_1 = i) = \psi_r(i). \quad (6)$$

We introduce the random variable $T_r^{(p)}$ defined by hitting time of an ascending run length at least equal to r when counting from position p . Thus we have

$$T_r^{(p)} = \inf\{k \geq r; X_{p+k-r} < \cdots < X_{p+k-1}\}.$$

We then have $T_r = T_r^{(1)}$. Moreover, $L_1 = \ell < r$ implies that $T_r = T_r^{(L_1+1)} + \ell$, which leads to

$$\begin{aligned} \mathbb{P}(T_r \leq n, L_1 < r | V_1 = i) &= \sum_{\ell=1}^{r-1} \mathbb{P}(T_r \leq n, L_1 = \ell | V_1 = i) \\ &= \sum_{\ell=1}^{r-1} \mathbb{P}(T_r^{(L_1+1)} \leq n - \ell, L_1 = \ell | V_1 = i) \\ &= \sum_{\ell=1}^{r-1} \sum_{j=1}^m \mathbb{P}(T_r^{(L_1+1)} \leq n - \ell, V_2 = j, L_1 = \ell | V_1 = i) \\ &= \sum_{\ell=1}^{r-1} \sum_{j=1}^m \Phi_\ell(i, j) \mathbb{P}(T_r^{(L_1+1)} \leq n - \ell | V_2 = j, L_1 = \ell, V_1 = i) \\ &= \sum_{\ell=1}^{r-1} \sum_{j=1}^m \Phi_\ell(i, j) \mathbb{P}(T_r^{(L_1+1)} \leq n - \ell | V_2 = j) \\ &= \sum_{\ell=1}^{r-1} \sum_{j=1}^m \Phi_\ell(i, j) \mathbb{P}(T_r \leq n - \ell | V_1 = j), \end{aligned} \quad (7)$$

where the fifth equality follows from the Markov property and the last one from the homogeneity of Y . Putting together relations (6) and (7), we obtain

$$\mathbb{P}(T_r \leq n | V_1 = i) = \psi_r(i) + \sum_{\ell=1}^{r-1} \sum_{j=1}^m \Phi_\ell(i, j) \mathbb{P}(T_r \leq n - \ell | V_1 = j). \quad \square$$

For every $n \geq 1$, we define M_n as the maximal ascending run length over the n first values X_1, \dots, X_n . We have $1 \leq M_n \leq m \wedge n$ and

$$M_n \geq r \iff T_r \leq n, \quad (8)$$

which implies

$$\mathbb{E}(M_n) = \sum_{r=1}^{m \wedge n} \mathbb{P}(M_n \geq r) = \sum_{r=1}^{m \wedge n} \mathbb{P}(T_r \leq n) = \frac{1}{m} \sum_{r=1}^{m \wedge n} \sum_{i=1}^m \mathbb{P}(T_r \leq n | V_1 = i).$$

When m is fixed and n gets large, we deduce from Eq. 8 that M_n converges in probability to m . Since $M_{n+1} \leq M_n$, $(M_n)_{n \geq 1}$ is a non-decreasing sequence which is also bounded by m , almost surely. Thus when m is fixed and n gets large, M_n converges almost surely to m .

Table 1 Algorithm for the distribution and expectation computation of M_n

Algorithm

input: m, n
output: $\mathbb{P}(M_h \geq r)$ for $h = 1, \dots, n$ and $r = 1, \dots, m \wedge n$.
 $\mathbb{E}(M_h)$ for $h = 1, \dots, n$.

for $\ell = 1$ **to** m **do** Compute the matrix Φ_ℓ **endfor**
for $r = 1$ **to** m **do** Compute the column vectors ψ_r **endfor**
for $h = 1$ **to** n **do** $W_{h,1} = \mathbb{1}$ **endfor**
for $r = 2$ **to** $m \wedge n$ **do**
 for $h = 1$ **to** $r - 1$ **do** $W_{h,r} = 0$ **endfor**
 for $h = r$ **to** n **do** $W_{h,r} = \psi_r + \sum_{\ell=1}^{r-1} \Phi_\ell W_{h-\ell,r}$ **endfor**
endfor
for $h = 1$ **to** n **do**
 for $r = 1$ **to** $m \wedge h$ **do** $\mathbb{P}(M_h \geq r) = \frac{1}{m} \mathbb{1}^t W_{h,r}$
endfor
 for $h = 1$ **to** n **do** $\mathbb{E}(M_h) = \frac{1}{m} \sum_{r=1}^{m \wedge h} \mathbb{1}^t W_{h,r}$ **endfor**

4 Algorithm

For $r = 1, \dots, m$, we denote by ψ_r the column vector of dimension m which i th entry is $\psi_r(i)$. For $r = 1, \dots, m$, $n \geq 1$ and $h = 1, \dots, n$, we denote by $W_{r,h}$ the column vector of dimension m which i th entry is defined by

$$W_{h,r}(i) = \mathbb{P}(T_r \leq h | V_1 = i) = \mathbb{P}(M_h \geq r | V_1 = i),$$

and we denote by $\mathbb{1}$ the column vector of dimension m with all entries equal to 1. An algorithm for the computation of the distribution and the expectation of M_n is given in Table 1.

The main effort required in this algorithm concerns the computation of the column vectors $W_{h,r}$. Each vector $W_{h,r}$ requires $r - 1$ matrix-vector products of dimension m which gives a complexity of $O(rm^2)$ for each of them. Since r varies from 2 to $m \wedge n$ and h varies from 1 to n , the total computational effort required is $O(m^2n(m \wedge n)^2)$. Note that this algorithm avoids numerical problems since all the computed matrices and vectors, i.e. Φ_ℓ , ψ_r and $W_{h,r}$, have their entries between 0 and 1. This leads to a stable algorithm with very accurate results.

5 Application to Wireless Networks: Fast Self-Organization

Our analysis has important implications in forecast large-scale wireless networks. In those networks, the number of machines involved and the likeliness of fault occurrences prevents any centralized planification. Instead, distributed self-organization must be designed to enable proper functioning of the network. A useful technique to provide self-organization is *self-stabilization* (Dijkstra 1974; Dolev 2000). Self-stabilization is a versatile technique that can make a wireless network withstand any kind of fault and reconfiguration.

A common drawback with self-stabilizing protocols is that they were not designed to handle properly large-scale networks, as the stabilizing time (the maximum amount of time needed to recover from any possible disaster) could be related to the actual size of the network. In many cases, this high complexity was due to the fact that network-wide unique identifiers are used to arbitrate symmetric situations (Tixeuil 2007). However, there exists a number of problems appearing in wireless networks that need only locally unique identifiers.

Modeling the network as a graph where nodes represent wireless entities and where edges represent the ability to communicate between two entities (because each is within the transmission range of the other), a local coloring of the nodes at distance d (i.e. having two nodes at distance d or less assigned a distinct color) can be enough to solve a wide range of problems. For example, local coloring at distance 3 can be used to assign TDMA time slots in an adaptive manner (Herman and Tixeuil 2004), and local coloring at distance 2 has successively been used to self-organize a wireless network into more manageable clusters (Mitton et al. 2005).

In the performance analysis of both schemes, it appears that the overall stabilization time is balanced by a tradeoff between the coloring time itself and the stabilization time of the protocol using the coloring (denoted in the following as the *client* protocol). In both cases (TDMA assignment and clustering), the stabilization time of the client protocol is related to the height of the directed acyclic graph induced by the colors. This DAG is obtained by orienting an edge from the node with the highest color to the neighbor with the lowest color. As a result, the overall height of this DAG is equal to the longest strictly ascending chain of colors across

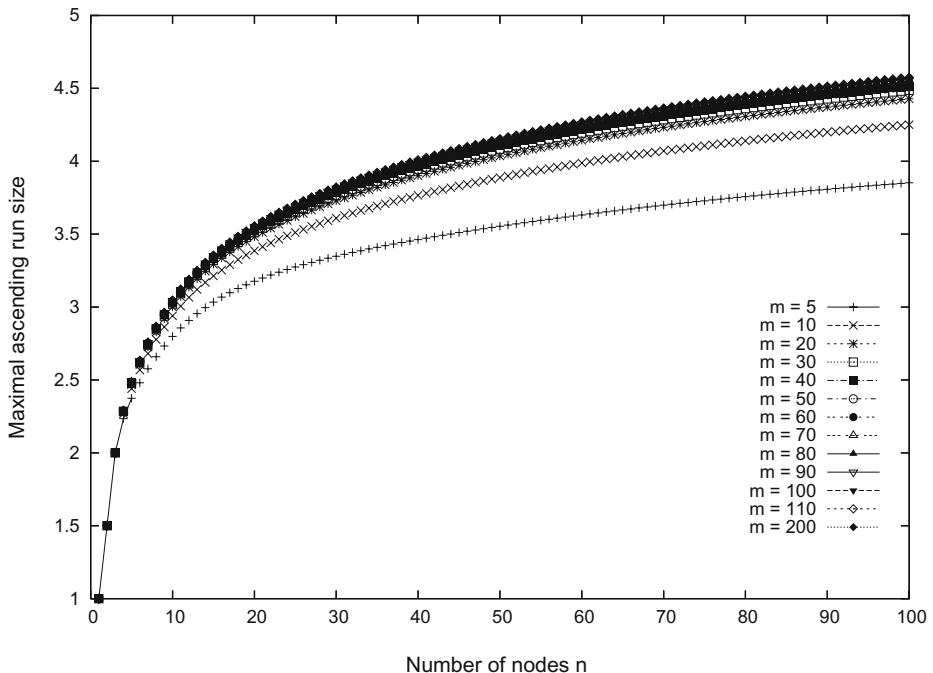


Fig. 1 Expected length of the maximal ascending run as a function of the number of nodes

neighboring nodes. Of course, a larger set of colors leads to a shorter stabilization time for the coloring (due to the higher chance of picking a fresh color), but yields to a potential higher DAG, that could delay the stabilization time of the client protocol.

In Mitton et al. (2006), the stabilization time of the coloring protocol was theoretically analyzed while the stabilization time of a particular client protocol (the clustering scheme of Mitton et al. (2005)) was only studied by simulation. The analysis performed in this paper gives a theoretical upper bound on the stabilization time of all client protocols that use a coloring scheme as an underlying basis. Together with the results of Mitton et al. (2006), our study constitutes a comprehensive analysis of the overall stabilization time of a class of self-stabilizing protocols used for the self-organization of wireless sensor networks. In the remaining of the section, we provide quantitative results regarding the relative importance of the number of used colors with respect to other network parameters.

Figure 1 shows the expected length of the maximal ascending run over a n -node chain for different values of the number m of colors.

Our numerical results point out several interesting behaviors. Indeed, self-organization protocols relying on a coloring process achieve better stabilization time when the expected length of maximal ascending run is short but a coloring process stabilizes faster when the number of colors is high (Mitton et al. 2006).

Figure 1 clearly shows that even if the number of colors is high compared to n ($n << m$), the expected length of maximal ascending run remains short, which is a great advantage. Moreover, even if the number of nodes increases, the expected length of the maximal ascending run remains short and increases very slowly. This

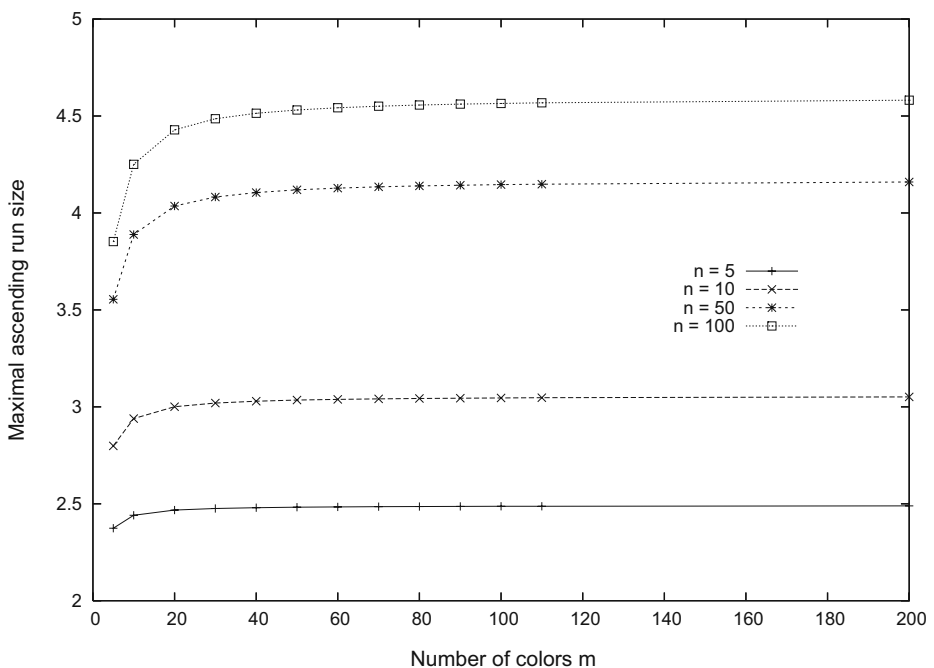


Fig. 2 Expected length of the maximal ascending run as a function of the number of colors

observation demonstrates the scalability properties of a protocol relying on a local coloring process since its stabilization time is directly linked to the length of this ascending run (Mitton et al. 2006).

Figure 2 shows the expected length of maximal ascending run over a n -node chain for different values of n .

Our numerical results indicate that for a fixed number of nodes n , the expected length of the maximal ascending run seems to converge toward a finite value, depending of n . This implies that using a large number of colors does not impact the stabilization time of the client algorithm.

References

- Csaki E, Foldes A (1996) On the length of the longest monotone block. *Stud Sci Math Hung* 31:35–46
- Dijkstra EW (1974) Self-stabilizing systems in spite of distributed control. *Commun ACM* 17(11): 643–644
- Dolev S (2000) *Self stabilization*. MIT Press
- Eryilmaz S (2006) A note on runs of geometrically distributed random variables. *Discrete Math* 306: 1765–1770
- Foata D, Fuchs A (1996) *Calcul des probabilités*. Masson
- Frolov AN, Martikainen AI (1999) On the length of the longest increasing run in \mathbb{R}^d . *Stat Probab Lett* 41(2):153–161
- Fu JC, Chang YM (2002) On probability generating functions for waiting time distributions of compound patterns in a sequence of multistate trials. *J Appl Probab* 39:70–80
- Fu JC, Lou WY (2003) *Distribution theory of runs and patterns and its applications: a finite Markov chain imbedding approach* world scientific, Singapore
- Herman T, Tixeuil S (2004) A distributed TDMA slot assignment algorithm for wireless sensor networks. In: *Proceedings of the first workshop on algorithmic aspects of wireless sensor networks (AlgoSensors'2004)*, no. 3121 in lecture notes in computer science. Springer-Verlag, Turku, Finland, pp 45–58
- Louchard G (2002) Runs of geometrically distributed random variables: a probabilistic analysis. *J Comput Appl Math* 142(1):137–153
- Louchard G (2005) Monotone runs of uniformly distributed integer random variables: a probabilistic analysis. *Theor Comp Sci* 346(2–3):358–387
- Louchard G, Prodinger H (2003) Ascending runs of sequences of geometrically distributed random variables: a probabilistic analysis. *Theor Comp Sci* 304:59–86
- Mitton N, Fleury E, Guérin-Lassous I, Tixeuil S (2005) Self-stabilization in self-organized multihop wireless networks. In: *WWAN'05*. Columbus, Ohio, USA
- Mitton N, Fleury E, Guérin-Lassous I, Sericola B, Tixeuil S (2006) Fast convergence in self-stabilizing wireless networks. In: *12th international conference on parallel and distributed systems (ICPADS'06)*. Minneapolis, MN, USA
- Tixeuil S (2007) *Wireless ad hoc and sensor networks*, chap. Fault-tolerant distributed algorithms for scalable systems. ISTE <http://www.iste.co.uk/index.php?p=a&ACTION=View&id=166>. ISBN: 978 1 905209 86