Ascending runs in dependent uniformly distributed random variables: Application to wireless networks

Nathalie Mitton, Katy Paroux, Bruno Sericola, Sébastien Tixeuil

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Ascending runs in dependent uniformly distributed random variables: Application to wireless networks

Nathalie Mitton , Katy Paroux , Bruno Sericola , Sébastien Tixeuil

N° 6443

February 2008

Thèmes COM et NUM

ISSN 0249-6399 ISRN INRIA/RR--6443--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Ascending runs in dependent uniformly distributed random variables: Application to wireless networks

Nathalie Mitton *, Katy Paroux †, Bruno Sericola ‡, Sébastien Tixeuil §

Thèmes COM et NUM — Systèmes communicants et Systèmes numériques
Équipes-Projets Dionysos, Grand Large et Pops

Rapport de recherche n° 6443 — February 2008 — 12 pages

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.

Key-words: Markov chains, maximal ascending run, self-stabilization, convergence time.

* INRIA Lille - Nord Europe/LIP6(USTL,CNRS), nathalie.mitton@inria.fr
† Université de Franche-Comté, katy.paroux@univ-fcomte.fr
‡ INRIA Rennes - Bretagne Atlantique, bruno.sericola@inria.fr
§ INRIA Saclay - Île-de-France/LIP6, sebastien.tixeuil@lri.fr

Centre de recherche INRIA Rennes – Bretagne Atlantique
IRISA, Campus universitaire de Beaulieu, 35042 Rennes Cedex
Téléphone : +33 2 99 84 71 00 — Télécopie : +33 2 99 84 71 71

Sous-suites croissantes contiguës de variables aléatoires dépendantes uniformément distribuées: application aux réseaux sans fil

Résumé : Nous analysons dans cet article la plus longue sous-suite croissante contiguë d'une suite de variables aléatoires de même distribution uniforme mais non indépendantes. Leur dépendance est caractérisée par le fait que deux variables successives ne peuvent prendre la même valeur. En utilisant une approche markovienne, nous étudions la distribution de la plus longue sous-suite croissante contiguë et nous développons un algorithme pour la calculer. Ce problème provient de l'analyse de plusieurs protocoles auto-organisants pour les réseaux de capteurs sans fil à grande échelle, et nous montrons comment nos résultats s'appliquent à ce domaine.

Mots-clés : Chaînes de Markov, sous-suites croissantes contiguës, auto-stabilisation, temps de convergence.

1 Introduction

Let $X = (X_n)_{n\geqslant 1}$ be a sequence of identically distributed random variables on the set $S = \{1, \ldots, m\}$. As in [8], 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 $\ell \geqslant 1$, starting at position $k \geqslant 1$, is a subsequence $(X_k, X_{k+1}, \ldots, X_{k+\ell-1})$ such that

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

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 [6] and [4] 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 [6].

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, $k \geqslant 1$. Some results have been established for the geometric distribution in [10] where an equivalent of the law of M_n is provided and previously in [1] where the almost-sure convergence is studied, as well as for Poisson distribution.

In [9], the case of the uniform distribution on the set $\{1, \ldots, s\}$ 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.

In this paper, we consider a sequence $X = (X_n)_{n\geqslant 1}$ of integer random variables on the set $S = \{1, \ldots, m\}$, with $m \geqslant 2$. The random variable X_1 is uniformly distributed on S and, for $n \geqslant 2$, 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 $n \geqslant 1$,

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

By induction over n and unconditioning, we get, for every $n \geqslant 1$ 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, 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 $X_1, \ldots, X_n$ 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 [10], 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 $Y = (V_k, L_k)_{k \geqslant 1}$, where V_k is the value of the first element of the kth ascending run and L_k is its length. 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 $k \geqslant 2$, the random variable V_k takes its values in $\{1, \ldots, m-1\}$. Moreover $V_1 = X_1 = m$ implies that $L_1 = 1$. Thus, for any $\ell \geqslant 2$, (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 $L_k \geqslant 2$, which implies that $(1, 1)$ is not a state of Y. Moreover $V_k = i$ implies that $L_k \leqslant m - i + 1$.

According to this behaviour, we have

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

where

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

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

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

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

$$\psi_\ell(i) = \mathbb{P}(L_k \geqslant \ell | V_k = i). \quad (3)$$

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 $(j, \lambda) \in E$ 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 $k \geqslant 1$ then for any $(v_k, \ell_k), \ldots, (v_1, \ell_1) \in E \cup \{(m, 1)\}$, we denote by A_k the event :

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

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

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

and

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

By induction, we obtain

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

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

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

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

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

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

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

where the third equality follows from the Markov property. ■

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

Lemma 2. *For every $i,j,\ell \in S$, we have*

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

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

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

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

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

It is well-known, see for instance [5], 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 (4) can be written as

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

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

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

Adding these two terms, we get

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

which completes the proof of the first relation.

The second relation follows from expression (3) by writing

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

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

Note that the matrix Φ defined by

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

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

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

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

3 Hitting times and maximal ascending run

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

$$T_r = \inf\{k \geqslant r\,;\, X_{k-r+1} < \cdots < X_k\}.$$

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

Theorem 3. *For $2 \leqslant r \leqslant m$, we have*

$$\mathbb{P}(T_r \leqslant n | V_1 = i) = \begin{cases} 0 \quad \textit{if} \quad 1 \leqslant n \leqslant r-1 \\ \\ \psi_r(i) + \displaystyle\sum_{\ell=1}^{r-1} \sum_{j=1}^{m} \Phi_\ell(i,j) \mathbb{P}(T_r \leqslant n - \ell | V_1 = j) \quad \textit{if} \quad n \geqslant r. \end{cases} \tag{5}$$

Proof. Since $T_r \geqslant r$, we have, for $1 \leqslant n \leqslant r-1$,

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

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

$$\mathbb{P}(T_r \leqslant n, L_1 \geqslant r | V_1 = i) = \mathbb{P}(L_1 \geqslant r | V_1 = i) = \psi_r(i). \tag{6}$$

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

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

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

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

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

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

■

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

$$M_n \geqslant r \Longleftrightarrow T_r \leqslant n,$$

which implies

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

4 Algorithm

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

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

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.

input : m, n
output : $\mathbb{E}(M_h)$ for $h = 1, \ldots, n$.
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** $\mathbb{E}(M_h) = \frac{1}{m} \sum_{r=1}^{m\wedge h} \mathbb{1}^t W_{h,r}$ **endfor**

Table 1: Algorithm for the distribution and expectation computation of M_n.

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* [2, 3]. 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 [13]. 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 [7], and local coloring at distance 2 has successively been used to self-organize a wireless network into more manageable clusters [12].

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 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 [11], the stabilization time of the coloring protocol was theoretically analyzed while the stabilization time of a particular client protocol (the clustering scheme of [12]) 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 [11], 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 m.

Results show 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 [11].

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 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 [11].

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

Results shows that for a fixed number of nodes n, the expected length of the maximal ascending run converges to 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.

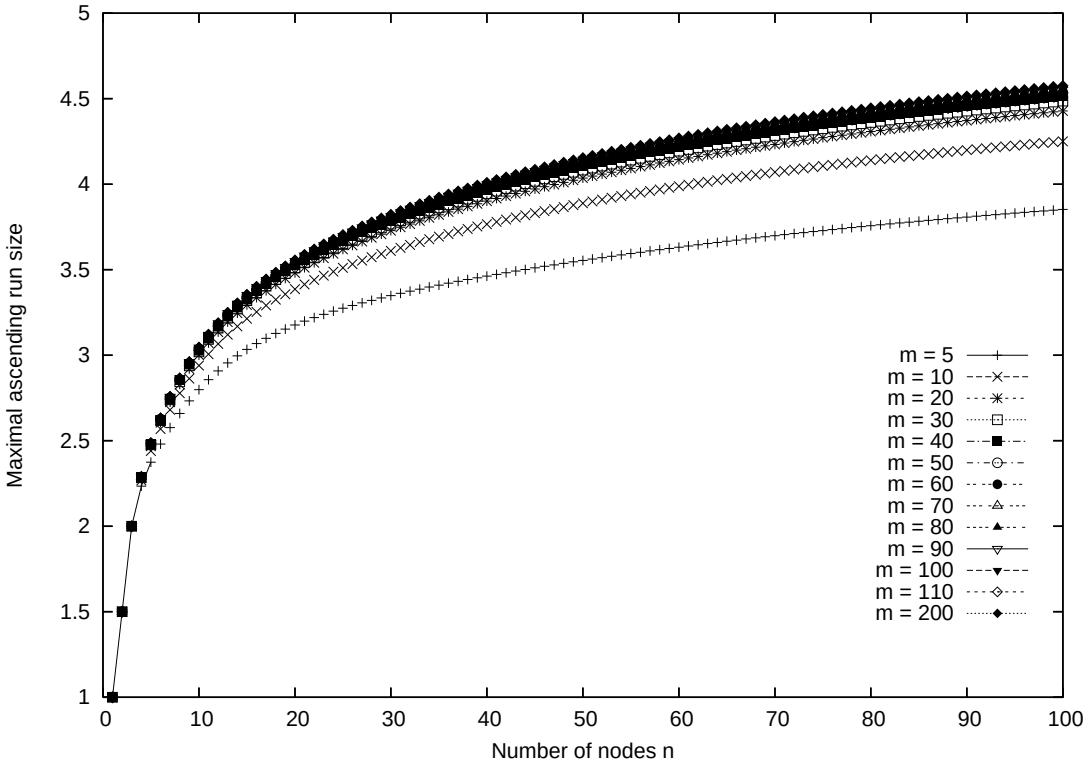

Figure 1: Expected length of the maximal ascending run as a function of the number of nodes.

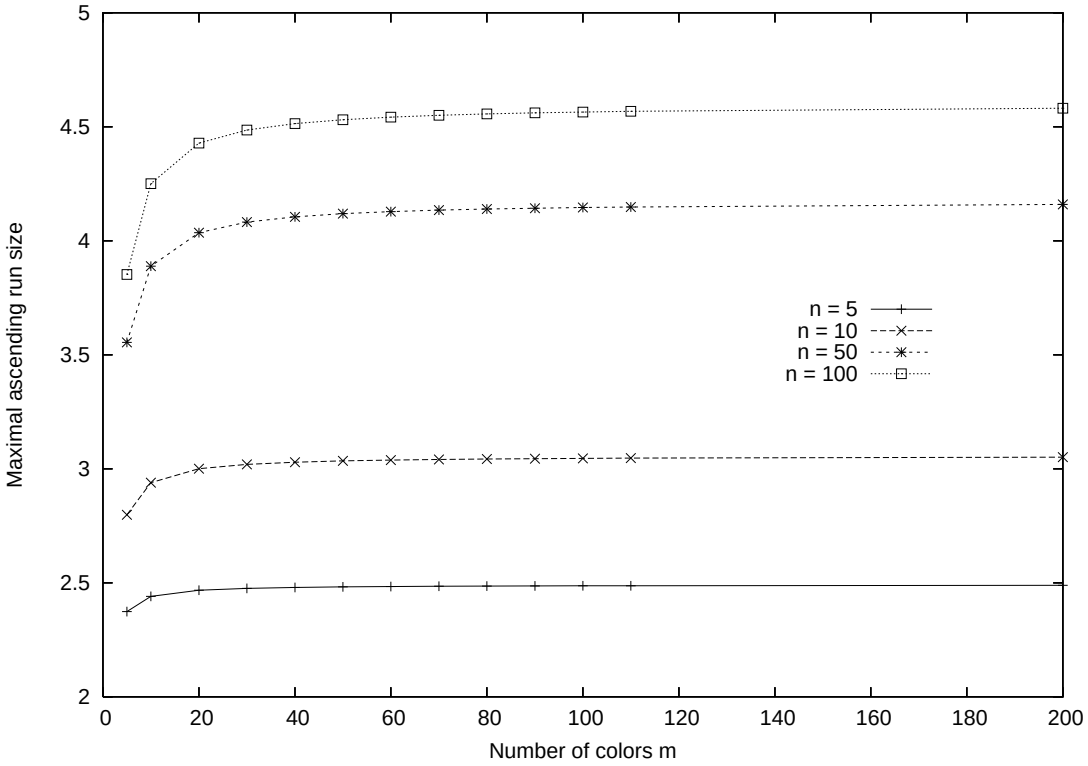

Figure 2: Expected length of the maximal ascending run as a function of the number of colors.

References

[1] E. Csaki and A. Foldes. On the length of the longest monotone block. *Stud. Sci. Math. Hungarian*, 31:35–46, 1996.

[2] E. W. Dijkstra. Self-stabilizing systems in spite of distributed control. *Commun. ACM*, 17(11):643–644, 1974.

[3] S. Dolev. *Self Stabilization*. MIT Press, 2000.

[4] S. Eryilmaz. A note on runs of geometrically distributed random variables. *Discrete Mathematics*, 306:1765–1770, 2006.

[5] D. Foata and A. Fuchs. *Calcul des probabilités*. Masson, 1996.

[6] A. N. Frolov and A. I. Martikainen. On the length of the longest increasing run in Rd. *Statistics and Probability Letters*, 41(2):153–161, 1999.

[7] T. Herman and S. Tixeuil. 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)*, number 3121 in Lecture Notes in Computer Science, pages 45–58, Turku, Finland, July 2004. Springer-Verlag.

[8] G. Louchard. Runs of geometrically distributed random variables: a probabilistic analysis. *J. Comput. Appl. Math.*, 142(1):137–153, 2002.

[9] G. Louchard. Monotone runs of uniformly distributed integer random variables: a probabilistic analysis. *Theoretical Computer Science*, 346(2–3):358–387, 2005.

[10] G. Louchard and H. Prodinger. Ascending runs of sequences of geometrically distributed random variables: a probabilistic analysis. *Theoretical Computer Science*, 304:59–86, 2003.

[11] N. Mitton, E. Fleury, I. Guérin-Lassous, B. Sericola, and S. Tixeuil. Fast convergence in self-stabilizing wireless networks. In *12th International Conference on Parallel and Distributed Systems (ICPADS'06)*, Minneapolis, Minnesota, USA, July 2006.

[12] N. Mitton, E. Fleury, I. Guérin-Lassous, and S. Tixeuil. Self-stabilization in self-organized multihop wireless networks. In *WWAN'05*, Columbus, Ohio, USA, 2005.

[13] S. Tixeuil. *Wireless Ad Hoc and Sensor Networks*, chapter Fault-tolerant distributed algorithms for scalable systems. ISTE, October 2007. ISBN: 978 1 905209 86.

Centre de recherche INRIA Rennes – Bretagne Atlantique
IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399