Fifth Colloquium on Mathematics and Computer Science DMTCS proc. **AI**, 2008, 295–310

Analysis of an algorithm catching elephants on the Internet

Yousra Chabchoub[1], Christine Fricker[1], Frédéric Meunier[2] and Danielle Tibi[3]

[1]*INRIA, Domaine de Voluceau, 78 153 Le Chesnay CX, France*
[2]*Université Paris Est, LVMT, Ecole des Ponts et Chaussées, 6 et 8 avenue Blaise Pascal, Cité Descartes, Champs sur Marne 77455 Marne-la-Vallée*
[3]*Université Paris 7, UMR 7599, 2 Place Jussieu, 75251 Paris Cedex 05*

The paper deals with the problem of catching the elephants in the Internet traffic. The aim is to investigate an algorithm proposed by Azzana based on a multistage Bloom filter, with a refreshment mechanism (called *shift* in the present paper), able to treat on-line a huge amount of flows with high traffic variations. An analysis of a simplified model estimates the number of false positives. Limit theorems for the Markov chain that describes the algorithm for large filters are rigorously obtained. The asymptotic behavior of the stochastic model is here deterministic. The limit has a nice formulation in terms of a $M/G/1/C$ queue, which is analytically tractable and which allows to tune the algorithm optimally.

Keywords: attack detection; Bloom filter; coupon collector; elephants and mice; network mining

Introduction

One traditionally distinguishes two kinds of flows in the Internet traffic: long flows, called *elephants*, which are the less numerous (typically 5-10%), and short flows, called *mice*, which are the most numerous. The convention is to fix a threshold C and to call elephant any flow having more than C packets, and mouse any flow having strictly less than C packets. For various reasons, detections of attacks, pricing, statistics, it is an important task to be able to "catch" the elephants, that means to be able to get the list of all elephants, with their IP addresses, flowing through a given router. We emphasize the fact that this problem is distinct from the one consisting only in providing statistical estimates for the traffic. Even a simple on-line counting of the number of distinct flows reveals to be difficult due to the high throughput of the traffic. There is a wide literature on algorithms for fast estimations of cardinality (i.e. of the number of distinct elements in a set with repeated elements) of huge data sets (see [6], [9] and [11]). A similar question consists in finding the k most frequent flows – the so-called "icebergs" (see [4] and [12]). If one asks for the proportion of elephants or the size distribution of elephants, it is possible to use the Adaptive Sampling algorithm proposed by Wegman and analyzed by Flajolet [8], which provides a sample of the flows independently from their size. This sample can then be used to compute statistics for the elephants (and actually for all the flows). For counting the elephants, Gandouet and Jean-Marie have proposed in [10] an algorithm based on sampling, thus requiring a knowledge on the flow size distribution, which reduces its application. For the target applications, a unique elephant hidden in a huge traffic of mice – which does not exist from a statistical point of view – has to be detected.

For this purpose, an algorithm based on Bloom filters has been already presented by Estan and Varghese [7] in 2002. As explained here, a Bloom filter allows elephants to accumulate but due to huge traffic, collisions occur and mice can be detected as elephants. Collisions will be controlled by taking several stages and cleverly flushing the filter. More precisely, the principle of this latter algorithm is the following. The IP header of each packet is hashed with k independent hashing functions in a k-multi-stage filter. Counters are incremented and when a counter reaches C, the corresponding flow is declared as an elephant and its packets are counted to give eventually its size. The problem is that in fact, under a heavy Internet traffic, the multistage filter quickly gets totally filled. To avoid this problem, Estan and Varghese propose to periodically (every 5 seconds) reinitialize the filter to zero. But, without any a priori knowledge about the traffic (intensity, flows arrival rate...), 5 seconds can either be too long (in which case the filter can be saturated) or too short (a lot of long flows can be missed). Therefore the accuracy of the algorithm closely depends on the characteristics of the traffic trace.

In order to settle this problem, Azzana [2] introduces an adaptative refreshment mechanism, that we will call *shift*, in the multi-stage filter algorithm. It is an efficient method to adapt the algorithm to traffic variations: The filter is refreshed with a frequency depending on the current traffic intensity. Moreover the filter is not reinitialized to zero, but a softer technique is used to avoid missing some elephants. The main difference with the Estan and Varghese algorithm is its ability to deal with traffic variations. Azzana shows in [2] that the refreshment mechanism improves notably the efficiency and accuracy of the algorithm (see Section 3 for practical results). Parameters, as the filter size or related to the refreshment mechanism, are experimentally optimized. Azzana proposes some elements of analysis for this algorithm. Our purpose is to go further and to provide analytical results when the filter is large.

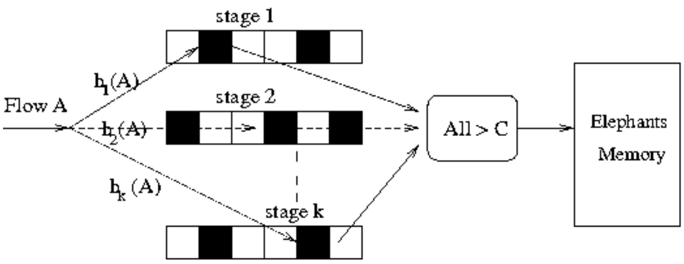

Fig. 1: The multi-stage filter

Description of the algorithm

The algorithm designed by Azzana uses a *Bloom filter with counters* (defined below) and involves four parameters in input: the number k of stages in the Bloom filter, the number m of counters in each stage, the maximum value C of each counter, i.e. the size threshold C to be declared as an elephant and the *filling rate* r.

A Bloom filter is a set of k stages, each of these stages being a set of m counters, initially at 0 and taking values in $\{0, \dots, C\}$. Together with the k stages $F_1, \dots, F_k$, one supposes that k hashing functions $h_1, \dots, h_k$ are given, one for each stage. We make the (strong) assumption that these hashing functions are independent, which implies that k is small ($k = 10$ is probably the upper limit). Each hashing function h_i maps the part of the IP header of a packet indicating the flow to which it belongs, to one of the counters of stage F_i.

The algorithm works on-line on the stream, processing the packets one after the other. Flows identified as elephants are stored in a list $\mathcal{E}$. When a packet is processed, it is first checked if it belongs to a flow already identified as an elephant (that is a flow already in $\mathcal{E}$). Indeed, in this case, there is no interest in mapping it to the k counters, and the algorithm simply forgets this packet. If not, it is mapped by the hashing functions on one counter per stage and it increases these counters by one, except for those that have already reached C, in which case they remain at C. When, after processing some packet, all the k counters are at C, the flow is declared to be an elephant and stored in the dedicated memory $\mathcal{E}$. When the proportion of nonzero counters reaches r in the whole set of km counters, one decreases all nonzero counters by one. This last operation is called the ***shift***. See Figure 1 for an illustration.

Motivation of the algorithm

Packets of a same flow hit the same k counters, but two distinct flows may also increase the same counter in one or several stages. The idea of using several stages where flows are mapped independently and uniformly, intends to reduce the probability of collisions between flows. The shift is crucial in the sense that it prevents the filters to be completely saturated, that is, to have many counters with high values. Without the shift operation, mice would be very quickly mapped to counters equal to C and declared as elephants. The algorithm would have a finite *lifetime* because when the filter is saturated, nothing can be detected.

False positive and false negative

A *false positive* is a mouse detected as an elephant by the algorithm. A *false negative* is an elephant not declared as such (hence considered as a mouse) by the algorithm. Generally, a false negative is worse than a false positive. Think of an attack: One does not want to miss it, and a false alarm has less serious consequences than a successful attack.

In our context, a false positive is a mouse one packet of which is mapped onto counters all $\geq C - 1$. A false negative is due to the shift, and if it happens, it means that there were at least $f - C + 1$ shifts during the transmission time of some elephant of size f. If shifts do not occur too often, a false negative is then an elephant whose packets are broadcast at a slow rate.

Intuitively, and it will be confirmed by the forthcoming analysis, if the parameters (actually r) are chosen so as to maintain counters at low values, then shifts occur often, and if one tries to decrease the shift frequency, then the counters tend to have high values. Therefore, a compromise has to be found between these two properties (frequency of the shifts, height of the counters), which translates into a compromise between false positives and false negatives. This last compromise depends on the applications.

A Markovian representation

In this paper, we will mainly focus our analysis on the one-stage filter case when the traffic is made up only of mice of size 1. The aim is to estimate the proportion of false positives. From this analysis, we will then derive results for the general case. Let us now introduce our main notations.

We thus assume that $k = 1$ and that all flows have size 1. Throughout the paper, $W_n^m(i)$ denotes the proportion of counters having value i just before the nth shift in a filter with m counters. According to this notation, $W_n^m(0)$ is close to $1-r$ and $\sum_{i=0}^{C} W_n^m(i) = 1$. Notice that the nth shift exactly decreases the number of nonzero counters by $mW_n^m(1)$. An important part of our analysis will consist in *estimating* $W_n^m(C-1) + W_n^m(C)$. Indeed, it gives an upper bound on the probability that a flow is declared as an elephant (that is a false positive) between the $(n-1)$th and the nth shifts, since, according to our assumptions, there is no elephant at all.

In this framework ($k = 1$ and flows of size one), the algorithm has a simple description in terms of urns and balls. Each flow is a ball thrown at random into one of m urns (each urn being one of the m counters). When a ball falls into an urn with C balls, it is immediately removed, in order to have at most C balls in each urn. When the proportion of non empty urns reaches r, one ball is removed in every non empty urn.

For m fixed, $(W_n^m)_{n\in\mathbb{N}} = \left((W_n^m(i))_{i=0,\dots,C}\right)_{n\in\mathbb{N}}$ is an ergodic Markov chain on some finite state space. Its invariant probability measure π_m is the distribution of some variable W_∞^m. For $C = 2$, the first non-trivial case, even the expression of the transition matrix P_m of the Markov chain is combinatorially quite complicated and an expression for π_m seems out of reach. In practice, the number m of counters per stage is large. This suggests to look at the limiting behavior of the algorithm when m tends to ∞. We use as far as possible the Markovian structure of the algorithm in order to derive rigorous limit theorems and analytical expressions for the limiting regime. This is the longest and most technical part of the paper, which also contains the main result, from a mathematical point of view.

Main results

The model considered in the paper describes the collisions between mice in order to evaluate the number of false positives due to these collisions. In a one-stage filter where all flows are mice of size 1, the Markov chain $(W_n^m)_{n\in\mathbb{N}}$ describes the evolution of the counters observed just before shift times. The main result is that, when m is large, the random vector W_∞^m converges in distribution to some deterministic value $\overline{w}$.

This result is not quite completely proved. The way to proceed is classical for large Markovian models (see for example [5] and [1]). The idea is to study the convergence of the process over finite times. It is shown that the Markov chain given by the empirical distributions $(W_n^m)_{n\in\mathbb{N}}$ converges to a deterministic dynamical system $w_{n+1} = F(w_n)$, which has a unique fixed point $\overline{w}$. The situation is analogous in discrete time to the study by Antunes and al. [1]. A Lyapunov function for F would allow to prove the convergence in distribution of W_∞^m. Such a Lyapunov function is exhibited in the particular case $C = 2$. The dynamical system provides a limiting description of the original chain which stationary behavior is then described by $\overline{w}$. The fixed point $\overline{w}$ has the following interpretation.

The fixed point $\overline{w}$ is identified as the invariant probability measure $\mu_{\overline{\lambda}}$ of the number of customers in an $M/G/1/C$ queue where service times are 1 and arrival rate is some $\overline{\lambda}$ satisfying the fixed point equation

$$\mu_{\overline{\lambda}}(0) = 1 - r$$

or equivalently

$$\overline{\lambda} = \log\left(1 + \frac{\mu_{\overline{\lambda}}(1)}{1-r}\right).$$

As a byproduct, the stationary time between two shifts divided by m converges in distribution to the constant $\overline{\lambda}$. Thus the inter-shift time (closely related to the number of false negatives) and the probability of false positives are respectively approximated by $\overline{\lambda}m$ and bounded by $\mu_{\overline{\lambda}}(C-1)+\mu_{\overline{\lambda}}(C)$ when m is large.

When mice have general size distribution, the previous model is extended to an approximated model where packets of a given mouse arrive simultaneously. The involved quantity is the invariant measure of an $M/G/1/C$ queue with arrivals by batches with distribution the mouse size distribution. In the case of size 1 mice, the multi-stage filter case is investigated.

Even if $\mu_{\overline{\lambda}}$ is not explicit, which complicates the exhibition of a Lyapunov function, the quantities $\overline{\lambda}$ and $\mu_{\overline{\lambda}}(C-1)+\mu_{\overline{\lambda}}(C)$ can be numerically computed. It appears that the latter quantity is an increasing function of r (as r varies from 0 to 1). Hence, given the mouse size distribution, one can numerically determine the values of r for which the algorithm performs well.

Section 1 is the most technical part of the paper. It investigates the one-stage filter in case of size 1 flows. In Section 2, this analysis is generalized to a general mouse size distribution in a simplified model and to a multi-stage filter. Then Section 3 is devoted to discussing the performance of the algorithm, to experimental results and improvements (validated through an implementation).

1 The Markovian urn and ball model

In this section, C is fixed and we consider the sequence $(W_n^m)_{n\in\mathbb{N}}$, where W_n^m denotes the vector of the proportions of urns with $0,\dots,C$ balls just before the nth shift time. For $m\geq 1$, $(W_n^m)_{n\in\mathbb{N}}$ is an ergodic Markov chain on the finite state space

$$\mathcal{P}_m^{(r)}=\left\{w=(w(0),\dots,w(C))\in\left(\frac{\mathbb{N}}{m}\right)^{C+1},\sum_{i=0}^{C}w(i)=1\text{ and }\sum_{i=1}^{C}w(i)=\frac{\lceil rm\rceil}{m}\right\},$$

(where $\lceil rm\rceil$ denotes the smallest integer larger or equal to rm) with transition matrix P_m defined as follows: If $W_n^m=w\in\mathcal{P}_m^{(r)}$, then W_{n+1}^m, distributed according to $P_m(w,.)$, is the empirical distribution of m urns when, starting with distribution w, one ball is removed from every non empty urn and then balls are thrown at random until $\lceil rm\rceil$ urns are non empty again, balls overflowing the capacity C being rejected. The required number of thrown balls is

$$\tau_n^m=\sum_{l=\lceil rm\rceil-W_n^m(1)m}^{\lceil rm\rceil-1}Y_l,\tag{1}$$

where $Y_l,\ l\in\mathbb{N}$ are independent random variables with geometrical distributions on $\mathbb{N}^*$ with respective parameters l/m, i.e. $\mathbb{P}(Y_l=k)=(l/m)^{k-1}(1-l/m),\ k\geq 1$.

Let F be defined on $\mathcal{P}=\left\{w\in\mathbb{R}_+^{C+1},\sum_{i=0}^{C}w(i)=1\right\}$ by

$$F(w)=T_C\left(s(w)*\mathcal{P}_{\lambda(w)}\right)\tag{2}$$

where

$$s : w \mapsto (w(0) + w(1), w(2), \dots, w(C), 0) \text{ on } \mathcal{P}$$

$$T_C : \mathcal{P}(\mathbb{N}) = \left\{ (w_n)_{n\in\mathbb{N}}, \sum_{i=0}^{+\infty} w_i = 1 \right\} \to \mathcal{P},\ w \mapsto \left(w(0), \dots, w(C-1), \sum_{i\geq C} w(i) \right)$$

$$\lambda : \mathcal{P} \to \mathbb{R}^+,\ w \mapsto \log\left(1 + \frac{w(1)}{1-r}\right)$$

and $\mathcal{P}_\lambda$ is the Poisson distribution with parameter λ. Notice that F maps $\mathcal{P}$ to itself and, also by definition of λ, $\mathcal{P}^{(r)} \overset{\text{def}}{=} \left\{ w \in \mathbb{R}_+^{C+1}, \sum_{i=0}^{C} w(i) = 1 \text{ and } \sum_{i=1}^{C} w(i) = r \right\}$ to itself.

1.1 Convergence to a dynamical system

We prove the convergence of $(W_n^m)_{n\in\mathbb{N}}$ to the dynamical system given by F as m tends to $+\infty$. The following lemma is the key argument. The uniform convergence stated below appears as the convenient way to express the convergence of $P_m(w,.)$ to $\delta_{F(w)}$ in order to prove both the convergence of $(W_n^m)_{n\in\mathbb{N}}$, and, later on, the convergence of the stationary distributions.

Define $\| x \| = \sup_{i=0}^{C} |x_i|$ for $x \in \mathbb{R}^{C+1}$.

Lemma 1 *For* $\varepsilon > 0$*,*

$$\sup_{w\in\mathcal{P}_m^{(r)}} P_m(w, \{w' \in \mathcal{P}_m^{(r)} : ||w' - F(w)|| > \varepsilon\}) \underset{m\to+\infty}{\longrightarrow} 0.$$

Proof: The first step is to prove that, for $\varepsilon > 0$,

$$\sup_{w\in\mathcal{P}_m^{(r)}} \mathbb{P}_w \left(\left| \frac{\tau_1^m}{m} - \lambda(w) \right| > \varepsilon \right) \underset{m\to\infty}{\to} 0 \tag{3}$$

where $\lambda(w) = \log\left(1 + \dfrac{w(1)}{1-r}\right)$ and $\mathbb{P}_w(.)$ denotes $\mathbb{P}(.|W_0^m = w)$. By Bienaymé-Chebyshev's inequality, it is enough to prove that

$$\sup_{w\in\mathcal{P}_m^{(r)}} \left| \mathbb{E}_w \left(\frac{\tau_1^m}{m} \right) - \lambda(w) \right| \underset{m\to\infty}{\to} 0 \tag{4}$$

and

$$\sup_{w\in\mathcal{P}_m^{(r)}} \mathrm{Var}_w \left(\frac{\tau_1^m}{m} \right) \underset{m\to\infty}{\to} 0. \tag{5}$$

By equation (1), as $\mathbb{E}(Y_l) = 1/(1 - l/m)$, using a change of index,

$$\mathbb{E}_w \left(\frac{\tau_1^m}{m} \right) = \sum_{l=\lceil rm\rceil - w(1)m}^{\lceil rm\rceil - 1} \frac{1}{m-l} = \sum_{j=m-\lceil rm\rceil+1}^{m-\lceil rm\rceil + w(1)m} \frac{1}{j}. \tag{6}$$

A comparison with integrals leads to the following inequalities:

$$\log \frac{1-\frac{\lceil rm\rceil}{m}+w(1)+\frac{1}{m}}{1-\frac{\lceil rm\rceil}{m}+\frac{1}{m}} \le \mathbb{E}_w\left(\frac{\tau_1^m}{m}\right) \le \log \frac{1-\frac{\lceil rm\rceil}{m}+w(1)}{1-\frac{\lceil rm\rceil}{m}}.$$

It is then easy to show that the two extreme terms tend to $\lambda(w) = \log(1 + w(1)/(1 - r))$, uniformly in $w(1) \in [0, 1]$. This gives (4). For (5), as $\mathrm{Var}(Y_l) = (l/m)/(1 - l/m)^2$, by the same change of index,

$$\mathrm{Var}_w\left(\frac{\tau_1^m}{m}\right) = \frac{1}{m}\sum_{j=m-\lceil rm\rceil+1}^{m-\lceil rm\rceil+w(1)m} \frac{m-j}{j^2} = \sum_{j=m-\lceil rm\rceil+1}^{m-\lceil rm\rceil+w(1)m} \frac{1}{j^2} - \frac{1}{m}\mathbb{E}_w\left(\frac{\tau_1^m}{m}\right). \tag{7}$$

The first term of the right-hand side is bounded independently of w by $\sum_{j=m-\lceil rm\rceil+1}^{+\infty} 1/j^2$, which tends to 0 as m tends to $+\infty$. The second term tends to 0 uniformly in w using (4) together with the uniform bound $\lambda(w) \le \log(1 + 1/(1 - r))$.

To obtain the lemma, it is then sufficient to prove that, for each $\varepsilon > 0$,

$$\sup_{w\in\mathcal{P}_m^{(r)}} \mathbb{P}_w\left(||W_1^m - F(w)|| > \varepsilon, \left|\frac{\tau_1^m}{m} - \lambda(w)\right| \le \frac{\varepsilon}{2}\right) \underset{m\to\infty}{\to} 0. \tag{8}$$

Since W_1^m and $F(w)$ are probability measures on $\{0, \dots, C\}$, to get (8), it is sufficient to prove that for $j \in \{0, \dots, C - 1\}$,

$$\sup_{w\in\mathcal{P}_m^{(r)}} \mathbb{P}_w\left(|W_1^m(j) - F(w)(j)| > \varepsilon, \left|\frac{\tau_1^m}{m} - \lambda(w)\right| \le \frac{\varepsilon}{2}\right) \underset{m\to\infty}{\to} 0. \tag{9}$$

Let $w \in \mathcal{P}_m^{(r)}$. Define the following random variables: For $1 \le i \le m$, N_i^m (respectively $\widetilde{N}_i^m(w)$) is the number of additional balls in urn i when τ_1^m (respectively $m\lambda(w)$) new balls are thrown in the m urns. One can construct these variables from the same sequence of balls (i.e. of i.i.d. uniform on $\{1, \dots, m\}$ random variables), meaning that balls are thrown in the same locations for both operations until stopping. This provides a natural coupling for the N_i's and $\widetilde{N}_i$'s. Let $j \in \{0, \dots, C-1\}$ be fixed. Given $W_0^m = w$, as $j \le C - 1$, the capacity constraint does not interfere and $W_1^m(j)$ can be represented as

$$W_1^m(j) = \frac{1}{m}\sum_{k=0}^{j}\sum_{i\in I_{w,k}^m} 1_{\{N_i^m=j-k\}} \tag{10}$$

where $I_{w,k}^m$ is the set of urns with k balls in some configuration of m urns with distribution $s(w)$, so that card $I_{w,k}^m = ms(w)(k)$. The sum over i is exactly the number of urns that contains k balls after the removing of one ball per urn, and having j balls after new balls have been thrown. By coupling, on the event $\{W_0^m = w,\ |\tau_1^m/m - \lambda(w)| \le \varepsilon/2\}$, the following is true:

$$\mathrm{card}\{i, N_i^m \ne \widetilde{N}_i^m(w)\} \le \frac{\varepsilon}{2}m \tag{11}$$

thus, denoting $\widetilde{W}_1^m(j) = \frac{1}{m}\sum_{k=0}^{j}\sum_{i\in I_{w,k}^m} 1_{\{\widetilde{N}_i^m(w)=j-k\}}$, on the same event,

$$|W_1^m(j) - \widetilde{W}_1^m(j)| \leq \frac{\varepsilon}{2}.$$

To prove equation (9), it is then sufficient to show that

$$\sup_{w\in\mathcal{P}_m^{(r)}} \mathbb{P}_w\left(|\widetilde{W}_1^m(j) - F(w)(j)| > \varepsilon\right) \underset{m\to\infty}{\longrightarrow} 0.$$

This will result from

$$\begin{array}{rcll} \sup_{w\in\mathcal{P}_m^{(r)}} |\mathbb{E}_w(\widetilde{W}_1^m(j)) - F(w)(j)| & \underset{m\to\infty}{\longrightarrow} & 0 & \text{and} \\ \sup_{w\in\mathcal{P}_m^{(r)}} \mathrm{Var}_w(\widetilde{W}_1^m(j)) & \underset{m\to\infty}{\longrightarrow} & 0. & \end{array} \quad (12)$$

which is quite standard to prove. The key argument, with classical proof, is the following: If L_i^m is the number of balls in urn i when throwing $m\lambda$ balls at random in m urns, if $0 < a < b$, then, for all $(i_1, i_2) \in \mathbb{N}^2$,

$$(i) \sup_{\lambda\in[a,b]} |\mathbb{P}(L_1^m = i_1) - \mathcal{P}_\lambda(i_1)| \underset{m\to\infty}{\longrightarrow} 0,$$

$$(ii) \sup_{\lambda\in[a,b]} |\mathbb{P}(L_1^m = i_1, L_2^m = i_2) - \mathcal{P}_\lambda(i_1)\mathcal{P}_\lambda(i_2)| \underset{m\to\infty}{\longrightarrow} 0.$$

It is applied since $\lambda(w) \in [0, \log(1 + 1/(1-r)]$. It ends the proof. $\square$

Proposition 1 *If W_0^m converges in distribution to $w_0 \in \mathcal{P}_m^{(r)}$ then $(W_n^m)_{n\in\mathbb{N}}$ converges in distribution to the dynamical system $(w_n)_{n\in\mathbb{N}}$ given by the recursion $w_{n+1} = F(w_n),\ n \in \mathbb{N}$.*

Proof: Assume that W_0^m converges in distribution to $w_0 \in \mathcal{P}_m^{(r)}$. Convergence of $(W_0^m, \ldots, W_n^m)$ can be proved by induction on $n \in \mathbb{N}$. By assumption it is true for $n = 0$. Let us just prove it for $n = 1$, the same arguments holding for general n, from the assumed property for $n - 1$. Let g be continuous on the (compact) set $\mathcal{P}_m^{(r)2}$. Since the distribution μ_m of W_0^m has support in $\mathcal{P}_m^{(r)}$,

$$\begin{aligned}\mathbb{E}\left(g(W_0^m, W_1^m)\right) &= \int_{\mathcal{P}_m^{(r)2}} g(w,w')P_m(w,dw')d\mu_m(w) \\ &= \int_{\mathcal{P}_m^{(r)}}\int_{\mathcal{P}_m^{(r)}} \left(g(w,w')P_m(w,dw') - g(w,F(w))\right) d\mu_m(w) + \int_{\mathcal{P}_m^{(r)}} g(w,F(w))d\mu_m(w).\end{aligned}$$

Since $g(.,F(.))$ is continuous on $\mathcal{P}_m^{(r)}$ (F being continuous as can be easily checked), the last integral converges to $g(w_0, w_1)$ by assumption (or case $n = 0$). The first term is bounded in modulus, for each $\eta > 0$, by

$$\begin{aligned}&\sup_{w\in\mathcal{P}_m^{(r)}} \left|\int_{\mathcal{P}_m^{(r)}} g(w,w')P_m(w,dw') - g(w,F(w))\right| \\ &\qquad\qquad \leq 2 \parallel g \parallel_\infty \sup_{w\in\mathcal{P}_m^{(r)}} P_m\left(w, \left\{w' \in \mathcal{P}_m^{(r)}, \parallel w' - F(w) \parallel > \varepsilon\right\}\right) + \eta\end{aligned}$$

where ε is associated to η by the uniform continuity of g on $\mathcal{P}_m^{(r)2}$. By Lemma 1, this is less than 2η for m sufficiently large. Thus, as m tends to $+\infty$,

$$\mathbb{E}\left(g(W_0^m, W_1^m)\right) \to g(w_0, w_1).$$

$\square$

1.2 Convergence of invariant measures

Let, for $m \in \mathbb{N}$, π_m be the stationary distribution of $(W_n^m)_{n\in\mathbb{N}}$. Define P as the transition on $\mathcal{P}^{(r)}$ given by $P(w, .) = \delta_{F(w)}$.

Proposition 2 *Any limiting point π of $(\pi_m)_{m\in\mathbb{N}}$ is a probability measure on $\mathcal{P}^{(r)}$ which is invariant for P i.e. that satisfies $F(\pi) = \pi$.*

Proof: A classical result states that, if P and $P_m,\ m \in \mathbb{N}$, are transition kernels on some metric space E such that, for any bounded continuous f on E, Pf is continuous and $P_m f$ converges to Pf uniformly on E then, for any sequence (π_m) of probability measures such that π_m is invariant under P_m, any limiting point of π_m is invariant under P. Indeed, for any m and any bounded continuous f, $\pi_m P_m f = \pi_m f$. If a subsequence (π_{m_p}) converges weakly to π, then $\pi_{m_p} f$ converges to πf. Writing $\pi_{m_p} P_{m_p} f = \pi_{m_p} Pf + \pi_{m_p}(P_{m_p} f - Pf)$, since Pf continuous (and bounded since f is), the first term $\pi_{m_p} Pf$ converges to πPf and the second term tends to 0 by uniform convergence of $P_m f$ to Pf. Equation $\pi_{m_p} P_{m_p} f = \pi_{m_p} f$ thus gives, in the limit, $\pi Pf = \pi f$ for any bounded continuous f.

Here the difficulty is that the P_m's and P are transitions on $\mathcal{P}_m^{(r)}$ and $\mathcal{P}^{(r)}$, which are in general disjoint. To solve this difficulty, extend artificially P_m and P to $\mathcal{P}$ by setting:

$$\begin{aligned} P_m(w, .) &= \delta_{F(w)} \quad \text{for } w \in \mathcal{P} \setminus \mathcal{P}_m^{(r)} \\ P(w, .) &= \delta_{F(w)} \quad \text{for } w \in \mathcal{P} \setminus \mathcal{P}^{(r)}. \end{aligned}$$

The proposition is then deduced from the classical result if we prove that, for each f continuous on $\mathcal{P}$ (notice that then $Pf = f \circ F$ is continuous),

$$\sup_{w \in \mathcal{P}_m^{(r)}} |P_m f(w) - f(F(w))| \underset{m\to\infty}{\longrightarrow} 0,$$

which is straightforward from Lemma 1. The fact that the support of π is in $\mathcal{P}^{(r)}$ is deduced from the portmanteau theorem (see Billingsley [3] p.16) using the sequence of closed sets

$$\mathcal{P}^{(r),n} = \left\{ w \in \mathcal{P}, r \leq \sum_{i=1}^{C} w(i) \leq r + \frac{1}{n} \right\}.$$

$\square$

The fixed points of the dynamical system are the probability measures w on $\mathcal{P}^{(r)}$ such that

$$w = F(w) = T_C(s(w) * \mathcal{P}_{\lambda(w)})$$

where $\lambda(w) = \log(1 + w(1)/(1-r))$. This is exactly the invariant measure equation for the number of customers just after completion times in an $M/G/1/C$ queue with arrival rate $\lambda(w)$ and service times 1, so that it is equivalent to

$$w = \mu_{\lambda(w)} \tag{13}$$

where μ_λ (respectively ν_λ) is the limiting distribution of the process of the number of customers in an $M/G/1/C$ (respectively $M/G/1/\infty$) queue with arrival rate λ and service times 1.

Indeed, it is well-known that this queue has a limiting distribution for $\lambda \in \mathbb{R}^+$ (respectively $0 \leq \lambda < 1$) which is the invariant probability measure of the embedded Markov chain of the number of customers just after completion times. The balance equations here reduce to a recursion system, so that, even when $\lambda \geq 1$, ν_λ is well defined up to a multiplicative constant (which can not be normalized into a probability measure in this case). Moreover, ν_λ is given by the Pollaczek-Khintchine formula for its generating function:

$$\sum_{n \in \mathbb{N}} \nu_\lambda(n) u^n = \nu_\lambda(0) \frac{g_\lambda(u)(u-1)}{u - g_\lambda(u)}, \quad \text{for } |u| < 1 \tag{14}$$

where $g_\lambda(u) = e^{-\lambda(1-u)}$ and for $\lambda < 1$, $\nu_\lambda(0) = 1 - \lambda$ (see for example Robert [13] p176-177). Notice that $\nu_\lambda(n)$ $(n \in \mathbb{N})$ has no closed form. For example, the expressions of the first terms are

$$\begin{aligned}
&\nu_\lambda(1) = \nu_\lambda(0)(e^\lambda - 1), \\
&\nu_\lambda(2) = \nu_\lambda(0)e^\lambda(e^\lambda - 1 - \lambda), \\
&\nu_\lambda(3) = \nu_\lambda(0)e^\lambda \left(\frac{\lambda(\lambda+2)}{2} - (1+2\lambda)e^\lambda + e^{2\lambda} \right)
\end{aligned} \tag{15}$$

where $\nu_\lambda(0) = 1 - \lambda$ if $\lambda < 1$. For the $M/G/1/C$ queue,

$$\mu_\lambda(i) = \frac{\nu_\lambda(i)}{\sum_{l=0}^{C} \nu_\lambda(l)}, \quad i \in \{0, \ldots, C\}. \tag{16}$$

The following proposition characterizes the fixed points of F.

Proposition 3 *F defined by* (2) *has one unique fixed point, denoted by $\bar{w}$, in $\mathcal{P}^{(r)}$ given by the limiting distribution $\mu_{\overline{\lambda}}$ of the number of customers in an $M/G/1/C$ queue with arrival rate $\overline{\lambda}$ and service times* 1*, where $\overline{\lambda}$ is determined by the implicit equation $\mu_{\overline{\lambda}}(0) = 1 - r$ which is equivalent to*

$$\overline{\lambda} = \log\left(1 + \frac{\mu_{\overline{\lambda}}(1)}{1-r}\right), \tag{17}$$

where μ_λ is given by (16) *and ν_λ by the Pollaczek-Kintchine formula* (14)*.*

Notice that, moreover,

$$r \leq \overline{\lambda} \leq -\log(1-r).$$

The upper bound on $\overline{\lambda}$, obtained from equation (17) using $\mu_{\overline{\lambda}}(1) \leq r$ just says that the stationary mean number of balls between two shifts is less than the mean number of balls thrown until the first shift (starting with empty urns). Moreover $\overline{\lambda} \geq r$, which is clear if $\overline{\lambda} \geq 1$, and obtained if $\overline{\lambda} < 1$ writing (16) for $i = 0$ and using $\sum_{l=0}^{C} \nu_{\overline{\lambda}}(l) \leq 1$ in this equation. This is exactly the fact that the asymptotic stationary mean number of balls $\overline{\lambda}m$ arriving between two shift times is greater than the number of removed balls at each shift, which is $\lceil rm \rceil$. It is due to the losses under the capacity limit C.

Proof: Only the existence and uniqueness result remains to prove. According to (13), w is some fixed point if and only if it is a fixed point of the function

$$\begin{aligned} \mathcal{P}^{(r)} &\longrightarrow \mathcal{P}^{(r)} \\ w &\longmapsto \mu_{\lambda(w)} \end{aligned}$$

with $\lambda(w) = \log(1 + w(1)/(1-r))$. This function being continuous on the convex compact set $\mathcal{P}^{(r)}$, by Brouwer's theorem, it has a fixed point. To prove uniqueness, let w and w' be two fixed points of F in $\mathcal{P}^{(r)}$. By definition of $\mathcal{P}^{(r)}$,

$$\mu_{\lambda(w)}(0) = \mu_{\lambda(w')}(0) = 1 - r. \tag{18}$$

A coupling argument shows that, if $\lambda \leq \lambda'$ then μ_λ is stochastically dominated by $\mu_{\lambda'}$, and in particular,

$$\mu_\lambda(0) + \mu_\lambda(1) \geq \mu_{\lambda'}(0) + \mu_{\lambda'}(1). \tag{19}$$

It can then be deduced that $\lambda(w) = \lambda(w')$. Indeed, if for example $\lambda(w) < \lambda(w')$, by equations (18) and (19),

$$\mu_{\lambda(w)}(1) \geq \mu_{\lambda(w')}(1).$$

thus, using (15) together with (18),

$$\lambda(w) = \log\left(1 + \frac{\mu_{\lambda(w)}(1)}{1-r}\right) \geq \lambda(w') = \log\left(1 + \frac{\mu_{\lambda(w')}(1)}{1-r}\right)$$

which contradicts $\lambda(w) < \lambda(w')$. One finally gets $\lambda(w) = \lambda(w')$, and then by equation (13), $w = w'$. $\square$

A Lyapunov function for the dynamical system given by F on $\mathcal{P}^{(r)}$ is a function $g \geq 0$ on $\mathcal{P}^{(r)}$ such that, for each $w \in \mathcal{P}^{(r)}$, $g(F(w)) \leq g(w)$ with equality if and only if w is the fixed point of F. In the particular case $C = 2$, a Lyapunov function can be exhibited, resulting from a contracting property of F in this case.

Indeed, restricted to $\mathcal{P}^{(r)}$, F is here given by:

$$\begin{aligned} w = (1-r, w(1), w(2) = r - w(1)) \longmapsto F(w) = \Bigg(1 - r, (1-r)\left[\log\left(1 + \frac{w(1)}{1-r}\right) + \frac{r - w(1)}{1 - r + w(1)}\right], \\ 1 - (1-r)\left[\log\left(1 + \frac{w(1)}{1-r}\right) + \frac{1}{1 - r + w(1)}\right]\Bigg). \end{aligned}$$

$\mathcal{P}^{(r)}$ is some one dimensional subvariety of $\mathbb{R}^3$, so that any $w \in \mathcal{P}^{(r)}$ can be identified with its second coordinate $w(1) \in [0, r]$, or equivalently with $\lambda(w) = \log(1 + w(1)/(1-r)) \in [0, \log(1/(1-r))]$.

Using this last parametrization of $\mathcal{P}^{(r)}$, it is easy to show that F rewrites as G, mapping the interval $I=[0,\log(1/(1-r))]$ to itself and defined, for $\lambda\in I$, by $G(\lambda)=\log\left(\lambda+\dfrac{e^{-\lambda}}{1-r}\right)$.

An elementary computation shows that G has derivative on I taking values in the interval $]-1,0]$, which gives the already known existence and uniqueness of a fixed point $\overline{\lambda}$ for G (or F, both assertions being equivalent, and $\overline{\lambda}$ being equal to $\lambda(\overline{w})$). Moreover, the following inequality holds for $\lambda\in I$:

$$|G(\lambda)-\overline{\lambda}|\le|\lambda-\overline{\lambda}|,$$

equality occurring only at $\lambda=\overline{\lambda}$. As a result, g defined on $\mathcal{P}^{(r)}$ by

$$g(w)=|\lambda(w)-\overline{\lambda}|=|\lambda(w)-\lambda(\overline{w})|=\left|\log\frac{1-r+w(1)}{1-r+\overline{w}(1)}\right|,$$

is a Lyapunov function for the dynamical system defined by F.

For $C>2$, we conjecture the existence of such a g.

Theorem 1 *Assume that a Lyapunov function exists for the dynamical system given by F on $\mathcal{P}^{(r)}$ then, as m tends to $+\infty$, the invariant measure of $(W_n^m)_{n\in\mathbb{N}}$ converges to $\delta_{\bar{w}}$ where $\bar{w}$ is the unique fixed point of F. Thus the following diagram commutes,*

$$\begin{tikzcd}
(W_n^m)_{n\in\mathbb{N}} \arrow[r, "(d)", "n\to+\infty"'] \arrow[d, "(d)", "m\to+\infty"'] & W_\infty^m \arrow[d, "(d)"] \\
(w_n)_{n\in\mathbb{N}} \arrow[r] & \bar{w}
\end{tikzcd}$$

Proof: We prove that $\delta_{\bar{w}}$ is the unique invariant measure π of P with support in $\mathcal{P}^{(r)}$. Let g be the Lyapunov function for F on $\mathcal{P}^{(r)}$. π is P-invariant, thus $\pi P=\pi$ and $\pi Pg=\pi g$ which can be rewritten $\int(g\circ F-g)d\pi=0$. This implies that $g=g\circ F$ holds π almost surely because $g-g\circ F\ge 0$. Equality being only true at $\bar{w}$, π has support in $\{\bar{w}\}$. □

2 A more general model

2.1 Mice with general size distribution

Let $(W_n^m)_{n\in\mathbb{N}}$ be the sequence of vectors giving the proportions of urns at $0,\ldots,C$ just before the nth shift time in a model where balls are thrown by batches. The balls in a batch are thrown together in a unique urn chosen at random among the m urns. The ith batch is composed with S_i balls and $(S_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables distributed as a random variable S on $\mathbb{N}^*$ with support containing 1. Let ϕ be the generating function of S. The quantity S_i is called the size of batch i. The dynamics is the same: If, before the nth shift time, the state is $w\in\mathcal{P}_m^{(r)}$, it first becomes $s(w)$ and then a number τ_n^m defined by (1) of successive batches are thrown in urns until $\lceil rm\rceil$ urns are non empty. The model generalizes the previous one obtained for $S=1$.

Let F be defined on $\mathcal{P}$ by

$$F(w) = T_c(s(w) * \mathcal{C}_{\lambda(w),S}) \tag{20}$$

where T_C, λ and S are already defined and $\mathcal{C}_{\lambda(w),S}$ is a compound Poisson distribution i.e. the distribution of the random variable

$$Y = \sum_{i=1}^{X} S_i \tag{21}$$

where X is independent of $(S_i)_{i\in\mathbb{N}}$ with Poisson distribution of parameter $\lambda(w)$.

We mimic the arguments in Section 1 to obtain the convergence of the stationary distribution of the ergodic Markov chain $(W_n^m)_{n\in\mathbb{N}}$ as m tends to $+\infty$ to a Dirac measure at the unique fixed point of F. Propositions 1 and 2 hold. The fixed points of F are described in the following proposition.

Proposition 4 *F defined for $w \in \mathcal{P}$ by*

$$F(w) = T_c(s(w) * \mathcal{C}_{\lambda(w),S})$$

has a unique fixed point on $\mathcal{P}^{(r)}$ which is exactly the invariant measure $\mu_{\bar\lambda}$ of the number of customers in a $M/G/1/C$ queue with batches of customers arriving according to a Poisson process with intensity $\bar\lambda$, batch sizes being i.i.d. distributed as S with generating function ϕ and service times 1, where $\bar\lambda$ is determined by the implicit equation

$$\mu_{\overline{\lambda}}(0) = 1 - r$$

which is equivalent to

$$\bar\lambda = \log\left(1 + \frac{\mu_{\bar\lambda}(1)}{1-r}\right)$$

where for $i \in \{0, \dots, C\}$,

$$\mu_\lambda(i) = \frac{\nu_\lambda(i)}{\sum_{l=0}^{C} \nu_\lambda(l)}$$

and ν_λ is given by

$$\sum_{n=0}^{+\infty} \nu_\lambda(n) u^n = \nu_\lambda(0) \frac{g \circ \phi(u)(u-1)}{u - g \circ \phi(u)}, \quad |u| < 1$$

where $g_\lambda(u) = e^{-\lambda(1-u)}$ and $\nu_\lambda(0) = 1 - \lambda\mathbb{E}(S)$ when $\lambda < 1$.

Recall that the first terms of ν_λ are given by

$$\nu_\lambda(1) = \nu_\lambda(0)(e^\lambda - 1)$$
$$\text{and } \nu_\lambda(2) = \nu_\lambda(0) e^\lambda (e^\lambda - 1 - \lambda\mathbb{P}(S = 1))$$

where $\nu_\lambda(0) = 1 - \lambda\mathbb{E}(S)$ when $\lambda < 1$, which generalizes the previous expressions. For $C = 2$, the Lyapunov function defined when $S = 1$ still works. Furthermore, for $C > 2$, we assume the existence of a Lyapunov function for F. Theorem 1 still holds.

2.2 A multi-stage filter

The filter is previously supposed to have only one stage. Let now assume that the filter has k stages of m counters each. The natural model then consists of k sets of m urns where, when a ball is thrown, k copies of this ball are sent simultaneously and independently into the k stages, each falling at random in one of the m urns of its set. If some ball hits an urn with C balls, then it is rejected. Moreover, when the proportion of non-empty urns in the *whole* filter reaches r, then one ball is removed from each non-empty urn: This is called a shift.

The previous analysis extends with one main difference: When the system is initialized at some state $(w_j(i), 1 \leq j \leq k, 0 \leq i \leq C)$, where $w_j(i)$ is the proportion of urns with i balls in stage j, the number τ_1^m of balls thrown in each stage before the next shift is now asymptotically equivalent to $\lambda(w)m$, where $w = (w(i), 0 \leq i \leq C)$ here gives the *global* proportions of urns in each possible state in the whole filter, that is $w(i) = \dfrac{1}{k}\sum_{j=1}^{k} w_j(i)$ for $0 \leq i \leq C$. Thus λ is the same function as for the one-filter case, here evaluated at the global proportions:

$$\lambda(w) = \log\Big(1 + \frac{\sum_{j=1}^{k} w_j(1)}{k(1-r)}\Big).$$

The one-stage proof of (3) is however not reproducible here, due to the lack of a representation of τ_1^m analogous to (1) of Section 1.

Another proof can be written. The alternative argument is provided by noticing that when αm balls per stage are thrown, with $\alpha \notin [\lambda(w)-\varepsilon, \lambda(w)+\varepsilon]$ (using the same arguments (i) and (ii) as in Section 1), the empirical distributions of the urns at each stage are precisely known (for large m) and do not correspond to the global proportion r of non-empty urns.

Once the (uniform) convergence of τ_1^m/m is established, the proof then proceeds along the same lines as for $k = 1$ (the same reasoning holding for each stage).

Notice however that the Markov property does not hold for the process of global proportions at shift times, so that convergence in distribution is proved for the process of proportions detailed by stage, then inducing convergence.

3 Discussions

3.1 Synthesis: false positives and false negatives

From a practical point of view, the main results are Propositions 3 and 4. Given some size distribution for the flows (the generating function ϕ of Section 2), these propositions show how the values of the counters can be computed from the different parameters of the algorithm, since these values are encoded by the fixed point $\overline{w}$ of F: according to Theorem 1, $\overline{w}$ is the state reached in the stationary regime when there is one stage and also when there are several stages (see Subsection 2.2: one has $\sum_{j=1}^{k} \overline{w}_j(C) = k\overline{w}(C)$). Moreover, the convergence is experimentally really fast (see the remark below), which ensures that in practice the algorithm lives in the stationary phase. The component $\overline{w}(i)$ of $\overline{w}$ gives the approximate proportion of counters having value i in the whole Bloom filter. $\overline{\lambda}$ is the number of packets that arrive between two shifts. $\overline{w}$ and $\overline{\lambda}$ are respectively related to the number of false positives and to the number of false negatives:

The probability that a packet is a false positive is less than $(\overline{w}(C-1)+\overline{w}(C))^k$ since, in stage j, the probability to hit a counter at height C is at most $\overline{w}_j(C-1)+\overline{w}_j(C)$ and

$$\prod_{j=1}^{k}(\overline{w}_j(C-1)+\overline{w}_j(C)) \leq (\overline{w}(C-1)+\overline{w}(C))^k.$$

The quantity $m\overline{\lambda}$ is the time (number of packets) between two shifts, which is connected to the number of false negatives according to the discussion “False positive and false negative” of the Introduction.

3.2 Implementation and tests

The algorithm has been implemented with an improvement called the *min-rule* already proposed in [7]. Instead of increasing the k counters, an arriving packet is incrementing only the counters among k having the minimum value. Analytically more difficult to study, the algorithm should perform better: Heuristically, more flows are needed to reach high values of the counters inducing fewer false positives; moreover, the time between two shifts is longer and hence the number of false negatives is decreased. It has been tested against on two ADSL traffic traces from France Telecom, involving millions of flows. The performance of the algorithm is evaluated comparing the real number of elephants with the value estimated by the algorithm. Even under the min-rule, the algorithm performs well only if r stays under a critical value r_c, closely dependent on the mice distribution.

Simulations have been processed with a one-stage filter with flows of size 1 to evaluate the transient phase duration. It appears that the number of shifts to reach the stationary phase is not much greater than C. Such a result on the speed of convergence seems however theoretically out of reach.

References

[1] Nelson Antunes, Christine Fricker, Philippe Robert, and Danielle Tibi. Stochastic networks with multiple stable points. *Annals of Probability*, 36(1):255–278, 2008.

[2] Youssef Azzana. *Mesures de la topologie et du trafic Internet*. PhD thesis, Université Pierre et Marie Curie, july 2006.

[3] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[4] E. D. Demaine, A. López-Ortiz, and J. I. Munro. Frequency estimation, of internet packet streams with limited space. In *Proceedings of the 10th Annual European Symposium on Algorithms (ESA 2002)*, pages 348–360, Rome, Italy, September 2002.

[5] Vincent Dumas, Fabrice Guillemin, and Philippe Robert. A Markovian analysis of additive-increase multiplicative-decrease algorithms. *Adv. in Appl. Probab.*, 34(1):85–111, 2002.

[6] M. Durand and P. Flajolet. Loglog counting of large cardinalities. In G. Di Battista and U. Zwick, editors, *Proceedings of the annual european symposium on algorithms, (ESA03)*, pages 605–617, 2003.

[7] C. Estan and G. Varghese. New directions in traffic measurement and accounting. In John Wroclawski, editor, *Proceedings of the ACM SIGCOMM 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM-02)*, volume 32, 4 of *Computer Communication Review*, pages 323–338, New York, August 19–23 2002. ACM Press.

[8] P. Flajolet. On adaptative sampling. *Computing*, pages 391–400, 1990.

[9] P. Flajolet, E. Fusy, O. Gandouët, and F. Meunier. Hyperloglog: the analysis of a near-optimal cardinality estimation algorithm. In *Proceedings of the 13th conference on analysis of algorithm (AofA 07)*, pages 127–146, Juan-les-Pins, France, 2007.

[10] O. Gandouet and A. Jean-Marie. Loglog counting for the estimation of ip traffic. In *Proceedings of the 4th Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, Nancy, France, 2006.

[11] F. Giroire. Order statistics and estimating cardinalities of massive data sets. *Discrete Applied Mathematics*, to appear.

[12] G. S. Manku and R. Motwani. Approximate frequency counts aver data streams. In *Proceedings of the 28th VLDB Conference*, pages 346–357, Hong Kong, China, 2002.

[13] Philippe Robert. *Stochastic networks and queues*, volume 52 of *Applications of Mathematics*. Springer-Verlag, Berlin, 2003. Stochastic Modelling and Applied Probability.