Performance of a cache with Random Replacement and Zipf document popularity

Philippe Olivier
Orange Labs Networks
38-40 Rue du Général Leclerc
92794 Issy-les-Moulineaux, France
phil.olivier@orange.com

Alain Simonian
Orange Labs Networks
38-40 Rue du Général Leclerc
92794 Issy-les-Moulineaux, France
alain.simonian@orange.com

ABSTRACT

The performance of a cache with Random Replacement policy is addressed in the case of a population of objects having a Zipf popularity distribution with decay parameter s. The main purpose of the paper is to provide new theoretical results on this scheme, within the Independent Reference Model framework. When $s > 1$, we derive a closed-form expression for the miss probability which is exact when s is an even integer and provides good approximation for all real s. In the case $s \leq 1$, we consider two different regimes where cache size C and document population N jointly grow to infinity. When C grows sublinearly with N, the miss probability tends to 1 and an asymptotic expression for the hit probability is provided for $1/2 < s < 1$. When C is linear with N, the miss probability is proved to have a non-zero limit, whose analytic expression is given, if $s < 1$, and to be of order $1/\log N$ if $s = 1$. Besides, some numerical experiments are reported which assess the validity and potential usefulness of the obtained analytical results.

Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: Network Architecture and Design.

General Terms

Performance.

Keywords

Network Caching, Replacement Policy, Stochastic Model, Asymptotic Analysis.

1. INTRODUCTION

A main feature of the forthcoming data networks is to bring information from repository servers down to the nearest nodes to the user delivery points, in order to save bandwidth needs and to decrease response times [13]. Such distributed content networks rely, in particular, on the ability for nodes to perform caching whereby the most popular documents (e.g. WWW documents, video files, ...) can be locally stored, at least temporarily. The performance of caching function is therefore one of the key ingredients for ensuring a good service level in operational networks. In this paper, we investigate the performance of a cache using the Random replacement discipline (RAND). Due to its simplicity, the RAND policy is worth considering and it has been shown [9] to be almost as efficient in relevant cases as the Least Recently Used (LRU) policy, commonly known as being optimal in practical situations.

The cache model we consider pertains to the so-called Independent Reference Model (IRM) framework where consecutive requests are assumed to be mutually independent. The IRM assumption is known to be somewhat unrealistic when compared to the real traffic traces; it is, however, considered as a first step for the sake of mathematical tractability, and proves useful to gain insight into the behaviour of caches as well as to enable further analysis taking request correlations into account [7].

The miss probability that a request is not satisfied at the cache level defines an essential performance indicator of the cache behaviour. For the RAND policy, a general expression for it has been given in the early paper [10], which was also shown to hold for the First In First Out (FIFO) replacement discipline. The computation for large cache capacity C and catalog size N proves, however, to be quite a challenging problem due to its combinatorial complexity. Nevertheless, a quite efficient and numerically stable method has been derived by [7], which allows one to compute the exact miss probability with $O(NC)$ complexity.

For very large population and cache sizes, simple approximations have been investigated for the RAND policy. A well-known method is proposed by [6] which, based on a probability equilibrium argument, computes an approximate hit probability via quite a simple iterative algorithm. Another method has been recently published [8]. It is similar in its approach to the so-called *Che approximation* [5] for LRU, as it is based on the numerical evaluation of a critical parameter, the cache "characteristic time", as the unique root of an implicit equation. These two methods show very good and comparable accuracy. From another point of view, a fluid analysis has been performed in [15], where the derivation of the miss probability appears as a justification for the approximation given in [6]. Finally, some exact formulae and asymptotic results are provided in [9] for a Zipf popularity distribution with decay parameter larger than 1.

ValueTools'13, December 10 – 12 2013, Turin, Italy

VALUETOOLS 2013, December 10-12, Torino, Italy

DOI 10.4108/icst.valuetools.2013.254387

Note that most of the previously quoted works only deal with numerical evaluations of the miss probability. In the perspective of very large cache dimensioning with RAND policy, and in order to be able to predict some general features or trends on cache performance, the search for closed forms and/or asymptotic formulae still remains a desirable goal. This is the very purpose of the present work to provide several theoretical contributions in that direction. We here specifically assume that the object popularity follows a Zipf distribution. Such a flat-tailed popularity distribution has been widely reported in the literature [2, 3, 14]; moreover, the derivation of a miss probability in closed-form is far from simple in that case, contrary to, e.g., the case of a geometric distribution as shown in [9].

The rest of the paper is organized as follows. Section 2 gives an overview of the general results which hold for the RAND policy and any popularity distribution. The derivation of exact results for a Zipf parameter larger than 1 is provided in Section 3. This is our first main contribution which complements in various aspects, including the complete derivation of some technical proofs, the results obtained in [9]. Section 4 provides the second main achievement, i.e. the investigation of two asymptotic regimes, namely $C = O(N^{1-s})$ and $C = O(N)$, in the case of a finite population and a Zipf parameter s less than 1. In each case, some numerical investigations are reported in order to illustrate the analysis. Section 5 briefly summarizes the obtained results.

2. CACHING WITH RAND POLICY

2.1 Cache performance with RAND

Consider a cache memory with size C which is offered requests for objects; all objects are enumerated with decreasing popularity order, the probability of object j to be requested being denoted by q_j. The population of objects has size N, which may be finite or infinite. Time is supposed to be discrete: at time $t \in \mathbb{N}$, the t-th requested object at the cache is denoted by $R(t) \in \mathbb{N}^*$. All random variables $R(t)$ are assumed to be mutually independent and identically distributed according to $\mathbb{P}(R(t) = j) = q_j$, $\forall j \in \{1, 2, ..., N\}$, corresponding to the commonly referred IRM framework.

In the stationary regime, a requested object not present in the cache is retrieved from the catalog repository (at a higher level in the network) and then inserted in the cache at the expense of another document which is thus ejected from it. In the present study, the object replacement policy is assumed to follow the RAND discipline, where the object to be replaced is uniformly chosen among the C objects currently present in the cache. In this context, the stationary miss probability $M_N(C)$, i.e. the probability that an object of any rank requested at any time is not contained in the cache, has been expressed in [10] in the general form

$$M_N(C) = \frac{\displaystyle\sum_{1 \leq j_1 < ... < j_C \leq N} q_{j_1} ... q_{j_C} \sum_{j \notin \{j_1, ..., j_C\}} q_j}{\displaystyle\sum_{1 \leq j_1 < ... < j_C \leq N} q_{j_1} ... q_{j_C}}. \qquad (2.1)$$

Note this result is also shown to hold for a cache managed according to the FIFO replacement discipline; all forthcoming results derived for RAND therefore equally apply to FIFO.

Now, define coefficient $G_N(C)$ by

$$G_N(C) = \sum_{1 \leq j_1 < ... < j_C \leq N} q_{j_1} ... q_{j_C} \qquad (2.2)$$

(with the convention $G_N(0) = 1$); note that $G_N(C) = 0$ for $C > N$. Expression (2.1) can simply be written as (see [9])

$$M_N(C) = (C + 1) \frac{G_N(C + 1)}{G_N(C)}. \qquad (2.3)$$

Following (2.2), the evaluation of coefficient $G_N(C)$ requires the numerical computation of a number $\binom{N}{C}$ of $q_{j_1} ... q_{j_C}$ terms; for large values of both C and N, such a number is very large and ensuring the numerical accuracy of the corresponding sum $G_N(C)$ becomes numerically intractable, all the more in the practically considered cases where the popularity distribution has a flat tail, i.e., probability q_j decreases slowly for increasing rank j. Alternative methods for evaluating the miss probability must therefore be envisaged. In the following, the calculation of coefficients $G_N(C)$ will be performed through their associated generating function F_N defined by

$$F_N(z) = \sum_{C \geq 0} G_N(C)\, z^C, \quad z \in \mathbb{C}. \qquad (2.4)$$

It is also shown in [9] that $F_N(z)$ equals the product

$$F_N(z) = \prod_{1 \leq j \leq N} (1 + q_j z), \quad z \in \mathbb{C}. \qquad (2.5)$$

2.2 General recursion formulae

As a first result, a relation allowing us to recursively compute all quantities $G_N(C)$ can be stated as follows.

PROPOSITION 2.1. *For any popularity distribution* $(q_j)_{j \in \mathbb{N}^*}$, *coefficients* $G_N(C)$ *can be recursively computed by*

$$G_N(C) = \frac{1}{C} \sum_{k=1}^{C} (-1)^{k-1} G_N(C - k) A_k, \ \ \forall\, C \in \mathbb{N}^*, \qquad (2.6)$$

with initial value $G_N(0) = 1$ *and* $A_k = \sum_{j \geq 1} q_j^k$, $k \in \mathbb{N}^*$.

PROOF. By (2.5), the logarithmic derivative $F'_N(z)/F_N(z)$ expands in power series of z as

$$\frac{F'_N(z)}{F_N(z)} = \sum_r \frac{q_r}{1 + q_r z} = \sum_{C \geq 0} (-1)^C A_{C+1} z^C \qquad (2.7)$$

for $|z| < 1$, with A_{C+1} defined as in the Proposition. By (2.4), the series $F'_N(z) = \sum_{C \geq 0} (C + 1) G_N(C + 1) z^C$ must then coincide with the expression derived from (2.7), i.e.,

$$F'_N(z) = \sum_{C \geq 0} z^C \left(\sum_{n=0}^{C} (-1)^n A_{n+1} G_N(C - n) \right)$$

and the index change $k = n + 1$ gives relation (2.6). $\square$

This result may appear quite simple at first view since the computation of $G_N(C)$ "only" requires the numerical evaluation of C series, maybe infinite, A_k, $1 \leq k \leq C$. The right-hand side of (2.6), however, is an alternate sum of positive numbers with close values, which leads to numerical instabilities as checked in the case of Zipf popularity distributions considered below.

In terms of numerical stability, a more powerful recursion formula has been derived in [7]. By considering the ratio $J_N(C) = G_N(C)/G_N(C-1)$ which is only needed in the evaluation of (2.3), actually, the following equation is obtained:

$$J_N(C) = \frac{J_{N-1}(C) + q_N}{J_{N-1}(C-1) + q_N}\; J_{N-1}(C-1). \tag{2.8}$$

This double recursion allows one to evaluate the exact miss probability by means of a numerically stable algorithm with $O(NC)$ complexity. Such an algorithm, although a great deal more efficient than that based on (2.2) and (2.3), shows its limitations, however, when the popularity distribution has an infinite, or at least very large, support. This will be illustrated below in Section 3.2, where a Zipf distribution with decay parameter larger than 1 is considered.

3. RAND WITH ZIPF DISTRIBUTION

We here assume the population of objects is infinite, all independent variables $R(t)$ being identically distributed according to a Zipf distribution

$$\mathbb{P}(R = j) = q_j = \frac{A}{j^s}, \;\; j \in \mathbb{N}^*, \tag{3.1}$$

with parameter $s > 1$ and normalizing constant $A = 1/\zeta(s)$, where ζ denotes Riemann Zeta function. Throughout this section, N is infinite and subscript N is omitted to denote quantities $M, G, F, ...$ Estimates for $G(C)$ and $M(C)$ have been extensively studied in [9] for large C and $s > 1$; it has been shown, in particular, that

$$M(C) \sim \frac{\rho_s}{\zeta(s)\; C^{s-1}} \tag{3.2}$$

with prefactor $\rho_s = [(\pi/s)/\sin(\pi/s)]^s$, for any real $s > 1$.

3.1 Exact results

Beside asymptotic (3.2), exact formulae for $M(C)$ have also been briefly mentioned in [9] in the specific cases $s = 2$, $s = 4$ and $s = 6$, namely

$$\begin{cases} M(C) = \dfrac{3}{2C+3}, \quad M(C) = \dfrac{45}{(4C+3)(4C+5)(2C+3)}, \\[2ex] M(C) = \dfrac{840}{(3C+2)(6C+5)(6C+7)(3C+4)(2C+3)}, \end{cases} \tag{3.3}$$

respectively, for any $C \geq 0$. In this section, we extend such rational expressions of $M(C)$ to the case when parameter s is any even integer. The complete proof of the following proposition is given in Appendix A. It mainly relies on using the infinite product expression (2.5) of generating function $F_N(z)$, its linearization in terms of sinh and cosh functions, and the evaluation of dominant terms in that linearization.

Proposition 3.1. *For a Zipf popularity distribution with even integer parameter s, the miss probability is given by*

$$M(C) = \frac{s!}{2^{s-1}} \frac{\beta_s}{B_s} \frac{(C+1)}{\displaystyle\prod_{j=s/2+1}^{3s/2} (sC+j)}\; \Delta_s(C) \tag{3.4}$$

for all $C \geq 0$, with

$$\beta_s = (-1)^{s/2-1} \left[\sin(\pi/s) \right]^{-s}, \tag{3.5}$$

where B_s denotes the s-th Bernoulli number and

(i) $\Delta_s(C)$ tends exponentially fast to 1 when $C \uparrow +\infty$,

(ii) $\Delta_s(C) = 1$ for all C when $s = 2, 4, 6$ and is given an explicit formulation when $s = 8$ (see (A.8) in Appendix A).

Recall that Bernoulli numbers $(B_n)_{n \geq 0}$ are rational numbers related to $\zeta(s)$ by $\zeta(s) = (-1)^{s/2-1}(2\pi)^s B_s/(2\; s!)$ and which appear in various fields of Number Theory. They can be conveniently computed by means of the formula

$$B_n = \sum_{k=0}^{n} \frac{1}{k+1} \sum_{j=0}^{k} (-1)^j \binom{k}{j} j^n. \tag{3.6}$$

By Proposition 3.1, $M(C)$ is consequently rational in C only for values $s = 2$, $s = 4$ and $s = 6$, corresponding to formulae (3.3). Finally, we consistently verify that letting C tend to infinity in (3.4) again gives asymptotic (3.2), as expected.

3.2 Numerical results

We here report some numerical experiments illustrating the validity and usefulness of Proposition 3.1. First, comparing with miss probabilities directly computed from general formula (2.1), we verified that for even values of s, the rational approximation, i.e., (3.4) without the correcting factor $\Delta_s(C)$, proves to be excellent: with $s = 10$, for instance, (3.4) matches the exact miss probability when $C > 2$ within 10^{-6} relative precision (complete numerical results are not detailed here for the sake of conciseness).

Now considering approximations in the case of Zipf parameter $s > 1$ which is not necessarily an even integer, Figure 1 shows the miss probability (on logarithmic scale plots) as a function of cache size when the Zipf distribution parameter is $s = 1.5$ or $s = 2.5$. On both plots of this figure, 'exact' curves refer to miss probabilities computed from general recursion (2.6) and 'asymptotic' to that obtained from (3.2). Moreover, 'extrapolated' curves refer to that obtained from a direct generalization of (3.4) when expressing all factorials in terms of the Gamma function, that is,

$$M_{ext}(C) = (C+1)\; \frac{\Gamma(3s/2+1)}{\Gamma(s/2+1)}\; \frac{\Gamma(sC+s/2+1)}{\Gamma(sC+3s/2+1)}. \tag{3.7}$$

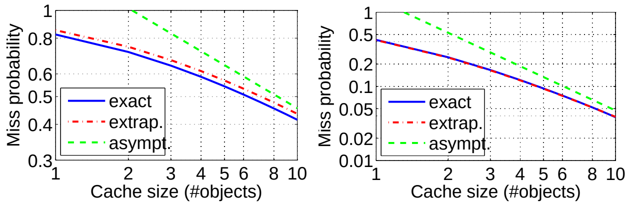

Figure 1: Miss probability against cache size, $s = 1.5$ (left) and $s = 2.5$ (right).

We note that the 'extrapolated' approximation provides excellent fit to exact results when $s = 2.5$, the corresponding curve being visually undistinguishable from the groundtruth 'exact' one. The approximation provided by (3.7) is less good for $s = 1.5$, but still much better than that provided by asymptotic (3.2), at least for small cache sizes.

The computation of miss probabilities by means of general formula (2.1) becomes more and more time consuming as s decreases towards 1, the heavy-tail nature of the Zipf popularity distribution becoming more and more acute. The recursion (2.6) shows numerical instabilities as soon as the cache capacity C exceeds quite a few tens of elements, while the use of recursion (2.8) does not help much in the present case where N is supposed to be infinite (of course the popularity distribution has to be truncated and then the computing time highly depends on the desired accuracy). Thus, using the closed-form extrapolation (3.7) may provide an alternative way of evaluating the miss probability for any cache size C when the population size N is arbitrarily large.

4. RAND WITH TRUNCATED ZIPF

We now assume a finite population of objects with size $N \geq 1$. Consider the truncated Zipf probability distribution

$$\mathbb{P}(R = j) = q_j = \frac{1}{\zeta_N(s)} \frac{1}{j^s}, \;\; 1 \leq j \leq N. \qquad (4.1)$$

with given decay exponent $s > 0$; the normalising constant for that distribution is $\zeta_N(s) = \Sigma_{1 \leq j \leq N} 1/j^s$. We intend to derive estimates of $M_N(C)$ for exponent $s \in \,]0, 1]$, with large C and finite $N > C$; to this end, two distinct asymptotic regimes are considered for which quite distinct quantitative and qualitative features are derived. In the first regime (Section 4.1), we assume that the capacity C and the catalogue size N scale as $C = \delta N^{1-s}$ with fixed constant δ and study the convergence speed of the hit rate $1 - M_N(C)$ to 0 as N increases. In the second regime (Section 4.2), the capacity C scales with the catalogue size as $C = \delta N$ with fixed δ; we then show that the miss rate $M_N(C)$ has either a non-zero limit for $0 < s < 1$ or has order $1/\log N$ for $s = 1$.

4.1 Asymptotic hit probability for $C = O(N^{1-s})$

First consider the asymptotic regime where $C = \delta N^{1-s}$ with fixed $\delta > 0$. Let $U_N(C) = C! \; G_N(C)$ so that miss rate $M_N(C)$ is equivalently expressed by the ratio

$$M_N(C) = \frac{U_N(C+1)}{U_N(C)}$$

for $1 \leq C \leq N$; as a preliminary, we show that $U_N(C)$ can be written in terms of the truncated Zeta function ζ_N.

Lemma 4.1. ***(See proof in Appendix B)***
(i) *For all $s > 0$, coefficient $G_N(C)$ can be written as*

$$G_N(C) = \sum_{m=1}^{C} \frac{(-1)^{C-m}}{m!} \, T_N^m(C) \qquad (4.2)$$

where

$$T_N^m(C) = \frac{1}{\zeta_N(s)^C} \sum_{k_1, \ldots, k_m \geq 1 | k_1 + \ldots + k_m = C} \; \prod_{\ell=1}^{m} \frac{\zeta_N(k_\ell s)}{k_\ell}. \qquad (4.3)$$

(ii) *Furthermore, coefficient $U_N(C)$ reads*

$$U_N(C) = 1 + \sum_{q=2}^{C} \frac{C!}{(C-q)!} \frac{1}{\zeta_N(s)^q} \sum_{n=0}^{\lfloor q/2 \rfloor} \frac{(-1)^{q-n}}{n!} V_N^{q-n,n} \qquad (4.4)$$

where we set

$$V_N^{m,r} = \sum_{k_1 > 1, \ldots, k_r > 1 | k_1 + \ldots + k_r = m + r} \; \prod_{\ell=1}^{r} \frac{\zeta_N(k_\ell s)}{k_\ell} \qquad (4.5)$$

for $m \geq r \geq 1$, together with $V_N^{0,0} = 1$, $V_N^{m,r} = 0$ for $m < r$, and $V_N^{m,0} = 0$ for $m \geq 1$.

The limiting behaviour of coefficient $U_N(C)$ can now be stated as follows.

Lemma 4.2. *For $s \in \,]1/2, 1[$, define*

$$U(x) = \prod_{j \geq 1} \left[1 + \frac{x}{j^s} \right] \exp \left(\frac{-x}{j^s} \right), \quad x > 0. \qquad (4.6)$$

If $C = \delta N^{1-s}$ for some finite constant $\delta > 0$ such that $0 < \delta(1-s) < 1$, then coefficient $U_N(C)$ tends to the limit $U(\delta(1-s))$ when $N \uparrow +\infty$.

Proof. Start from expression (4.4) for $U_N(C)$. We use the notations $\mathbf{k}_n$, $\mathrm{Tr}(.)$ and $\mathbb{L}^n$ introduced in Appendix B*(ii)*. For given $q > 1$, each positive sum $V_N^{q-n,n}$, $0 \leq n \leq \lfloor q/2 \rfloor$, defined in (4.5), increases to its limit

$$V^{q-n,n} = \sum_{\mathbf{k}_n \in \mathbb{L}^n | \mathrm{Tr}(\mathbf{k}_n) = q} \; \prod_{\ell=1}^{n} \frac{\zeta(k_\ell s)}{k_\ell} \qquad (4.7)$$

when $N \uparrow +\infty$, since for given $k_\ell > 1$ and $s > 1/2$, we have $k_\ell s > 1$ and the limit $\lim_{N \uparrow +\infty} \zeta_N(k_\ell s) = \zeta(k_\ell s)$ is therefore well-defined. Besides, we have $\zeta_N(s) \sim N^{1-s}/(1-s)$ when $0 < s < 1$; for any $\varepsilon > 0$, there is thus some N_0 such that $\zeta_N(s) \geq (1-\varepsilon) N^{1-s}/(1-s)$ for all $N \geq N_0$, hence

$$\frac{C!}{(C-q)!} \, \frac{1}{\zeta_N(s)^q} \leq \frac{\delta^q \; (1-s)^q}{(1-\varepsilon)^q}$$

for any given $q > 1$, large enough N and $C = \delta N^{1-s}$. For large enough N (depending on ε), we thus deduce that

$$| \; U_N(C) \; | \leq 1 + \sum_{q \geq 2} \frac{\delta^q (1-s)^q}{(1-\varepsilon)^q} \sum_{n=0}^{\lfloor q/2 \rfloor} \frac{V^{q-n,n}}{n!}.$$

If $\delta(1-s) < 1$, we can choose ε such that $\delta(1-s)/(1-\varepsilon) < 1$, and the r.h.s. of the above inequality is therefore finite (see below). By the dominated convergence theorem, we conclude that $\lim_{N \uparrow +\infty} U_N(C) = U(\delta(1-s))$, where function U is defined by $U(x) = 1 + V(x)$ with

$$V(x) = \sum_{q \geq 2} x^q \sum_{n=0}^{\lfloor q/2 \rfloor} \frac{(-1)^{q-n}}{n!} V^{q-n,n} = \sum_{q \geq 2} (-1)^q x^q V^{q,0} - \sum_{q \geq 2} (-1)^q x^q V^{q-1,1} + \sum_{n \geq 2} \frac{(-1)^n}{n!} \sum_{q \geq n} (-1)^q x^q V^{q-n,n} = V_0(x) - V_1(x) + V_2(x).$$

Let us now calculate $V(x)$. Using (4.7), we have $V^{q,0} = 0$ for $q \geq 2$ so that $V_0(x) = 0$; further

$$V_1(x) = \sum_{k > 1} (-1)^k x^k \frac{\zeta(ks)}{k}$$

and

$$V_2(x) = \sum_{n \geq 2} \frac{(-1)^n}{n!} \sum_{q \geq n} (-1)^q x^q \sum_{\mathbf{k}_r \in \mathbb{L}^r | \mathrm{Tr}(\mathbf{k}_r) = q} \; \prod_{\ell=1}^{n} \frac{\zeta(k_\ell s)}{k_\ell}$$
$$= \sum_{n \geq 2} \frac{(-1)^n}{n!} \left(\sum_{k > 1} (-1)^k x^k \frac{\zeta(ks)}{k} \right)^n = e^{-V_1(x)} + V_1(x) - 1$$

hence $V(x) = e^{-V_1(x)} - 1$; finally, calculate

$$V_1(x) = \sum_{k>1} (-1)^k \frac{x^k}{k} \sum_{j \geq 1} \frac{1}{j^{ks}} = \sum_{j \geq 1} \sum_{k>1} \frac{(-1)^k}{k} \frac{x^k}{j^{ks}}$$

$$= \sum_{j \geq 1} \left[\frac{x}{j^s} - \log\left(1 + \frac{x}{j^s}\right) \right] = \log \prod_{j \geq 1} \left[1 + \frac{x}{j^s}\right]^{-1} \exp\left(\frac{x}{j^s}\right)$$

with $0 < x < 1$ and where the infinite product is well-defined for $s > 1/2$. Expression (4.6) of $U(x) = 1 + V(x) = e^{-V_1(x)}$ eventually follows. $\square$

We can now assert the estimate for the hit probability $1 - M_N(C)$, $1/2 < s < 1$, in proper asymptotic conditions.

Proposition 4.1. *For $s \in]1/2, 1[$, let*

$$\Phi(x) = \sum_{j \geq 1} \frac{1}{j^{2s} + x j^s}, \quad x \geq 0.$$

If $C = \delta N^{1-s}$ for some finite $\delta > 0$ with $0 < \delta(1-s) < 1$, the hit probability can be evaluated by

$$1 - M_N(C) \sim (1-s)^2 \, \Phi(\delta(1-s)) \cdot \frac{C}{N^{2(1-s)}} \tag{4.8}$$

when $N \uparrow +\infty$.

Proof. Write by definition

$$1 - M_N(C) = \frac{U_N(C) - U_N(C+1)}{U_N(C)} \tag{4.9}$$

where by Lemma 4.2, $U_N(C) \to U(\delta(1-s))$ when $N \uparrow +\infty$ and $C = \delta N^{1-s}$. By (4.4), we readily calculate the difference

$$U_N(C+1) - U_N(C) = \frac{D_N(C)}{C+1} + \frac{(C+1)!}{\zeta_N(s)^{C+1}} E_N(C) \tag{4.10}$$

where

$$D_N(C) = \sum_{q=2}^{C} \frac{(C+1)\, q}{C+1-q} \, \frac{C!}{(C-q)!} \, \frac{1}{\zeta_N(s)^q} \sum_{n=0}^{\lfloor q/2 \rfloor} \frac{(-1)^{q-n}}{n!} V_N^{q-n,n}$$

and

$$E_N(C) = \sum_{n=0}^{\lfloor (C+1)/2 \rfloor} \frac{(-1)^{C+1-n}}{n!} V_N^{C+1-n,n}.$$

The MacLaurin-Cauchy theorem applied to decreasing function $x \in [1, +\infty[\mapsto 1/x^s$ asserts that there exists $\gamma_s > 0$ such that $\lim_{N \uparrow +\infty} [\, \sum_{j=1}^N j^{-s} - \int_1^N x^{-s} \mathrm{d}x \,] = \gamma_s$. For large N, we thus obtain

$$\zeta_N(s) = \sum_{j=1}^{N} \frac{1}{j^s} = \frac{N^{1-s}}{1-s} + \left(\gamma_s - \frac{1}{1-s}\right) + o(1)$$

hence $\zeta_N(s)^C \sim N^{(1-s)\delta N^{1-s}} (1-s)^{-\delta N^{1-s}} e^{\,\delta\, [(1-s)\gamma_s - 1]}$ for $C = \delta N^{1-s}$. Stirling's formula for the factorial and the latter estimate then provide

$$\frac{(C+1)!}{\zeta_N(s)^{C+1}} \sim (\delta(1-s))^{\delta N^{1-s}+1} \frac{e^{-\delta N^{1-s}}}{e^{\delta\, [(1-s)\gamma_s - 1]}} \sqrt{2\pi\delta N^{1-s}}$$

which tends to 0 when $N \uparrow +\infty$, provided that $\delta(1-s)) < e$ (which condition is fulfilled since we assume the stronger constraint $\delta(1-s)) < 1$). Thus, noting that the sum $E_N(C)$ is bounded, we conclude that the second term in the r.h.s. of (4.10) tends to 0 exponentially fast when $N \uparrow +\infty$. Considering now the sum $D_N(C)$ in the r.h.s. of (4.10), recall that $C!/[(C-q)!\ \zeta_N(s)^q] \sim \delta^q (1-s)^q = x^q$ and apply the dominated convergence theorem (as in the proof of Lemma 4.2 above) to derive that $D_N(C)$ has the limit

$$\sum_{q \geq 2} q x^q \sum_{n=0}^{\lfloor q/2 \rfloor} \frac{(-1)^{q-n}}{n!} V^{q-n,n} = x\, U'(x),$$

with function U defined in (4.6). We deduce from (4.9), (4.10) and the above evaluations that

$$1 - M_N(C) \sim -\frac{1}{C+1} \cdot \frac{x\, U'(x)}{U(x)} \sim -\frac{x\, (\log U)'(x)}{C}.$$

To compute the latter, use definition (4.6) to obtain

$$(\log U)'(x) = \sum_{j \geq 1} \left(\frac{1}{j^s + x} - \frac{1}{j^s} \right) = -x \Phi(x)$$

so that $1 - M_N(C) \sim x^2 \Phi(x)/C$ with $x = \delta(1-s)$, and result (4.8) follows. $\square$

Note that hit rate estimate is here provided for an exponent $s \in]1/2, 1[$; a similar approach might provide corresponding estimates for smaller exponents $s \leq 1/2$. We have limited the discussion for practical reasons, as actual values of s measured to date for popularity distributions in content delivery networks are essentially larger than 1/2 [3].

4.2 Asymptotic miss probability for $C = O(N)$

We now consider the case when the cache size C and the object population N grow together to infinity as $C = O(N)$. To derive asymptotic estimates for miss probability $M_N(C)$, we use a Large Deviations approach to derive an estimate of associated coefficient $G_N(C)$.

As a preliminary step, we state that, for any given $C < N$, there is a unique real positive solution $z = \theta_{N,C}$ to equation

$$z \, \frac{F_N'(z)}{F_N(z)} = C. \tag{4.11}$$

In fact, differentiate expression (2.5) for generating function F_N, so that (4.11) reads $g_N(z) = C$ where

$$g_N(z) = \sum_{j=1}^{N} \frac{z}{\zeta_N(s)\, j^s + z};$$

function g_N is defined in $\mathbb{C} \setminus\ ]-\infty, -\zeta_N(s)]$, is strictly increasing on the real interval $[0, +\infty[$, with $g_N(0) = 0$ and $\lim_{z \uparrow +\infty} g_N(z) = N$; the existence and unicity of unique positive root $z = \theta_{N,C}$ follows.

For $0 < s \leq 1$, define function L by

$$L(x) = \frac{1}{x} \int_0^x \frac{\mathrm{d}t}{1 + t^s}, \quad x > 0. \tag{4.12}$$

It is easily shown that equation

$$L(\xi) = \delta \tag{4.13}$$

with given $\delta \in]0, 1[$ has a unique unique solution $\xi > 0$. Now, the following lemma will help for further calculations.

Lemma 4.3. *For $0 < s \leq 1$ and large $z = \zeta_N(s)\ (N/x)^s$, $x = O(1)$, $\log F_N(z)$ expands as*

$$\log F_N(z) = H(x)N - \frac{s}{2}\log N + \frac{1}{2}\log\left(1 + x^s\right) - s + \frac{s}{12} + T_N(z) + O\left(\frac{1}{N^s}\right), \qquad (4.14)$$

with some remainder term $T_N(z)$ and

$$H(x) = \log\left(1 + x^s\right) - s\log x + s\ L(x), \quad x > 0.$$

Proof. For $z > 0$ and $t \geq 1$, define function h by

$$h(z;t) = \log\left(1 + \frac{z}{\zeta_N(s)\ t^s}\right) \qquad (4.15)$$

so that $\log F_N(z) = \sum_{1 \leq j \leq N} h(z;j)$. The Euler-MacLaurin formula ([11], Chap.VI, Sect.2, formula (16.4)) applied to the first order then yields

$$\log F_N(z) = \int_1^N h(z;t)\ \mathrm{d}t + \frac{1}{2}\left[h(z;N) + h(z;1)\right] + \frac{1}{12}\left[h'(z;N) - h'(z;1)\right] + T_N(z) \qquad (4.16)$$

with remainder term

$$T_N(z) = \int_1^N \frac{P_3(t)}{3!}\, h^{(3)}(z;t)\ \mathrm{d}t,$$

where P_3 is the periodic Bernoulli function of third order. Simple calculations then allow us to evaluate each term of (4.16) for $z = \zeta_N(s)\ (N/x)^s$, $x = O(1)$, from which (4.14) follows. $\square$

The estimation of coefficients $G_N(C)$ can now be derived by invoking the Large Deviations setting provided in ([4], Theorem 4.1). According to that theorem,

1. if there exists $a > 0$ such that function $u \mapsto \log F_N(e^u)$ is analytic and bounded in $\{u \in \mathbb{C},\ \mid \Re(u) \mid \leq a\}$;

2. if there exists some constant $\sigma > 0$ such that

$$\lim_{C \uparrow +\infty} e^{u\sqrt{C}}\ \frac{F_N(\theta_{N,C}e^{-u/\sqrt{C}})}{F_N(\theta_{N,C})} = e^{\sigma^2 u^2/2}$$

for any given $u \in \mathbb{C}$ with $\Re(u) = 0$;

3. and if given any $\varepsilon > 0$, there exists $\eta \in\ ]0,1[$ and an integer C_ε such that

$$C \geq C_\varepsilon \Longrightarrow \sup_{\varepsilon \leq |y| \leq \pi}\left|\frac{F_N(\theta_{N,C}\ e^{iy})}{F_N(\theta_{N,C})}\right|^{1/C} \leq \eta,$$

then coefficient $G_N(C)$ can be estimated by

$$G_N(C) \sim \frac{1}{\sigma\sqrt{2\pi C}} \cdot \frac{F_N(\theta_{N,C})}{\theta_{N,C}^C} \qquad (4.17)$$

as C tends to infinity. In the present case, the verification of technical conditions *1.*, *2.* and *3.* for generating function F_N defined in (2.5) proceeds as in ([9], Appendix D) (details are here omitted for necessary conciseness). Lemma 4.3 and general estimate (4.17) then lead to the following result for the miss probability.

Proposition 4.2. ***(See proof in Appendix C)*** *Let the cache size C grow to infinity as $C = \delta N$ with fixed $\delta \in\ ]0,1[$. Given the unique solution ξ to (4.13), which depends on s and δ, then*

a) *if $0 < s < 1$, miss probability $M_N(C)$ tends to the limit*

$$\lim_{C \uparrow +\infty} M_N(C) = \delta\ (1 - s)\ \xi^s; \qquad (4.18)$$

b) *if $s = 1$, miss probability $M_N(C)$ is estimated by*

$$M_N(C) \sim \frac{\delta\ \xi}{\log N}. \qquad (4.19)$$

These results are formally similar to those obtained with the LRU replacement policy investigated in [12] for Zipf distributions with parameter $s \leq 1$; recall (**ibid.**, Theorems 2 and 3 with $k = 1$) that for $C = \delta N$ and large N, the miss probability for LRU is estimated by

$$\lim_{C \uparrow +\infty} M_N(C) = \frac{1 - s}{s}\eta^{\frac{1}{s}-1}\Gamma\left(1 - \frac{1}{s};\eta\right),\ 0 < s < 1, \qquad (4.20)$$

$$M_N(C) \sim \frac{\Gamma(0;\eta)}{\log N},\ s = 1. \qquad (4.21)$$

Here $\eta = \eta_\delta(s)$ denotes the unique solution to equation

$$1/s\ \Gamma\left(-1/s;\eta\right)\eta^{1/s} = 1 - \delta$$

where $(s;x) \mapsto \Gamma(s;x)$ is the incomplete Gamma function. A brief comparison of RAND and LRU performance is given in the next section in such an asymptotic regime.

4.3 Numerical results

We now provide some numerical results illustrating the asymptotic results derived above. First, the cache and object population sizes being related by $C = \delta N^{1-s}$, Figure 2 plots the miss probability (4.8) as a function of cache size for parameter $s = 0.6$ and a coefficient $\delta = 0.5$ or $\delta = 5$. On both plots, 'exact' curves refer to miss probabilities computed by means of recursion formula (2.6) (they might be computed from (2.8) as well), and 'asymptotic' curves refer to that obtained from (4.8).

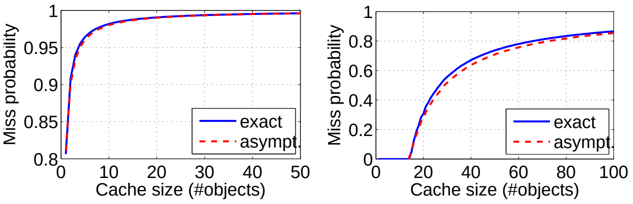

Figure 2: ***Miss probability against cache size,*** $s = 0.6$ ***and coefficient*** $\delta = 0.5$ ***(left) and*** $\delta = 5$ ***(right).***

In Figure 2, both 'exact' and 'asymptotic' curves perfectly match in the case $\delta = 0.5$ where the condition $\delta(1 - s) < 1$ of Proposition 4.1 holds. In the case $\delta = 5$ where the latter condition is not fulfilled, the 'asymptotic' curve provides a slightly optimistic estimate of the miss rate. On the whole, our asymptotic expression provides quite a good approximation for the cache performance, even for low values of the cache size.

Table 1: Comparison of asymptotic miss probabilities from RAND and LRU with $\delta = 0.1$

Replac. Policy	$s = 0.6$	$s = 0.8$	$s = 1$
RAND	$M = 0.765$	$M = 0.521$	$Num = 3.61$
LRU	$M = 0.742$	$M = 0.484$	$Num = 1.11$

Turn now to the evaluation of asymptotics given in Proposition 4.2 when $C = \delta N$. Figure 3 shows the miss probability as a function of cache size when $s = 0.6$, $s = 0.8$, or $s = 1$, and when coefficient $\delta = 0.1$ in each case. On all plots, the 'exact' curve provided by recursion (2.8) is taken as the ground-truth. We observe that the miss probability rather slowly converges towards the limit given by (4.18) when $s < 1$. The approximation provided by this limit is not so good as the one provided by asymptotic (4.8) (when $C = \delta N^{1-s}$). Moreover, the estimate provided is slightly optimistic, i.e. the miss probability is under-estimated. On the contrary, when $s = 1$, asymptotic (4.19) is both pessimistic, thus providing a worst case for dimensioning purposes, and also reasonably accurate even for small cache sizes.

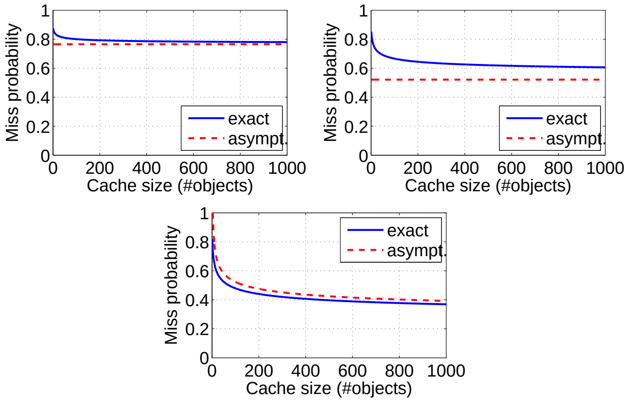

Figure 3: Miss probability against cache size, $s = 0.6$ (top left), $s = 0.8$ (top right) and $s = 1$ (bottom), and $\delta = 0.1$.

A comparison of asymptotic miss probabilities achieved by the RAND and LRU policies when $C = \delta N$, according to formulae (4.18) to (4.21), is provided in Table 1. The respective limits for miss probability are denoted by M when $s = 0.6$ or $s = 0.8$, while the numerator of asymptotic miss probability in (4.19) and (4.21) is denoted by Num when $s = 1$. What we mainly observe from these results is that RAND discipline is only slightly less efficient than LRU when the Zipf distribution parameter s is significantly less than 1. As expected, the performance gain provided by LRU increases with s, e.g., when $s = 1$ the difference is quite noticeable.

To end this section, we turn our attention back to [7] where some numerical results illustrate the efficiency of recursion (2.8) for computing the exact miss probability. In the linear asymptotic scaling $C = \delta N$ with $s = 0.5$ and $\delta = 0.3$, the authors note that the miss probability seems to tend to a finite limiting value for growing N, see Table 1 in [7]. Applying (4.18), we here deduce that this limit equals 0.5965, a value consistent with the results for N varying from 10 to 10,000 given in [7].

5. CONCLUSION

New theoretical results have been provided for the performance of a cache with Random Replacement discipline, when assuming a Zipf distribution of object popularity. In the case when the decay parameter s is larger than 1, a rational expression of the miss probability has been shown to be exact for some integer values and to provide good approximations for real s. When s is smaller than 1, asymptotic expressions for the miss rate are provided when the cache capacity C and the catalog size N scale together either as $C = O(N^{1-s})$ or $C = O(N)$. We believe that such results can be used as basic tools for the efficient computation of miss probabilities, useful for large cache dimensioning, and in more complex situations such as the design of a network of caches.

6. ACKNOWLEDGMENTS

The authors acknowledge the anonymous reviewers for their comments and constructive suggestions which helped to improve the quality of this paper.

7. REFERENCES

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, 1972.

[2] L. Breslau, P. Cue, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and zipf-like distributions: Evidence and implications. In *Proc. of IEEE INFOCOM*, pages 126–134, 1999.

[3] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. *Proceedings of ACM IMC*, 2007.

[4] N. Chaganty and J. Sethuraman. Multi-dimensional Large Deviations Local Limit Theorems. *Journal of Maultivariate Analysis*, 20:190–204, 1986.

[5] H. Che, Y. Tung, and Z. Wang. Hierarchical Web Caching Systems: Modeling, design and experimental results. *IEEE JSAC*, 20(7):1305–1314, Sept. 2002.

[6] A. Dan and D. Towsley. An approximate analysis of the lru and fifo buffer replacement schemes. *SIGMETRICS Perform. Eval. Rev.*, 18(1):143–152, Apr. 1990.

[7] R. Fagin and T. G. Price. Efficient calculation of expected miss ratios in the independent reference model. *SIAM J. Comput.*, 7(3):288–297, 1978.

[8] C. Fricker, P. Robert, and J. Roberts. A versatile and accurate approximation for LRU cache Performance. In *Proc. of ITC 24*, 2012.

[9] M. Gallo, B. Kauffmann, L. Muscariello, A. Simonian, and C. Tanguy. Performance evaluation of the random replacement policy for network of caches. *To appear in Performance Evaluation*, 2013.

[10] E. Gelenbe. A unified approach to the evaluation of a class of replacement algorithms. *IEEE Trans. on Computers*, C-22(6):611–618, 1973.

[11] R. Godement. *Analyse Mathématique II*. Springer Verlag, 2003.

[12] P. Jelenkovic, X. Kang, and A. Radovanovic. The persistent-access-caching algorithm. *Random Structures and Algorithms*, 33(2):219–251, Sept. 2008.

[13] T. Koponen, M. Chawla, B. Chu, A. Ermolinsk, K. Kim, S. Shenker, and I. Stoica. A data-oriented (and beyond) network architecture. *Proceedings of ACM SIGCOMM*, 2007.

[14] S. Mitra, M. Agrawal, A. Yadav, N. Carlsson, D. Eager, and A. Mahanti. Characterizing web-based video sharing workloads. *ACM Trans. Web*, 5(2):8:1–8:27, May 2011.

[15] N. Tsukada, R. Hirade, and N. Miyoshi. Fluid limit analysis of FIFO and RR caching for independent reference models. *Performance Evaluation*, 2012.

APPENDIX

A. PROOF OF PROPOSITION 3.1

Let $s = 2p$ be the even parameter of the Zipf popularity distribution. From (2.5) and (3.1), the generating function F of coefficients $G_N(C)$ reads in this case

$$F(z) = \prod_{j \geq 1} (1 + q_j z) = \prod_{j \geq 1} \left(1 + \frac{z}{\zeta(2p) j^{2p}} \right). \qquad \text{(A.1)}$$

The derivation of (3.3) for $s = 2$ (see [9], Corollary 3.5) makes a central use of the infinite product formula ([1], p. 85, formula 4.5.68)

$$\frac{\sinh u}{u} = \prod_{j \geq 1} \left(1 + \frac{u^2}{\pi^2 j^2} \right), \quad u \in \mathbb{C}, \qquad \text{(A.2)}$$

by expanding its right-hand side into power series of u. To generalize this approach for any even parameter s, we now express the generating function F as a product of hyperbolic sine functions. First, performing the variable change

$$z = (-1)^{p+1} \frac{\zeta(2p)}{\pi^{2p}} u^{2p} \qquad \text{(A.3)}$$

and making use of identity $\prod_{k=1}^{p} \left(1 - \lambda_p^{2k} x\right) = 1 - x^p$ where $\lambda_p = e^{i\pi/p}$ is the elementary $(2p)^{th}$ root of unity, equality (A.1) first provides

$$H(u) = F(z) = \prod_{j \geq 1} \left(1 - \left(\frac{-u^2}{\pi^2 j^2} \right)^p \right) = \prod_{k=1}^{p} \prod_{j \geq 1} \left(1 + \frac{(\lambda_p^k u)^2}{\pi^2 j^2} \right)$$

for all $u \in \mathbb{C}$. Using (A.2), we then obtain

$$H(u) = \prod_{k=1}^{p} \frac{\sinh(\lambda_p^k u)}{\lambda_p^k u} = \frac{g(u)}{i^{p+1} u^p} \qquad \text{(A.4)}$$

where $g(u) = \prod_{k=1}^{p} \sinh(\lambda_p^k \ u)$. The next step is to 'linearize' the product $g(u)$ by introducing a generalization of the identity $\sinh a \sinh b = \frac{1}{2}[\cosh(a+b) - \cosh(a-b)]$.

Lemma A.1. *For all $p \geq 1$ and complex p-tuple $(x_k)_{1 \leq k \leq p}$, the product of hyperbolic sine functions can be 'linearized' as*

$$\prod_{k=1}^{p} \sinh x_k = \frac{1}{2^{p-1}} \sum_{\phi} \left[\prod_{k=1}^{p} \phi(k) \right] \sinh^{(p-1)} \left(\sum_{k=1}^{p} \phi(k) x_k \right) \qquad \text{(A.5)}$$

where function $\sinh^{(p-1)}$ denotes either $\sinh$ if $p-1$ is even or $\cosh$ if $p-1$ is odd, and ϕ stands for any of the 2^{p-1} mappings from $\{1, 2, ..., p\}$ to $\{-1, 1\}$ such that $\phi(k) = 1$.

The proof proceeds by simple recursion on p. Thus, the product $g(u)$, as an application of Lemma A.1, linearizes as

$$g(u) = \frac{1}{2^{p-1}} \sum_{\phi} \mathbf{R}_\phi \sinh^{(p-1)}(\mathbf{S}_\phi \ u), \quad u \in \mathbb{C},$$

where $\mathbf{R}_\phi = \prod_{k=1}^{p} \phi_{(}k)$ and $\mathbf{S}_\phi = \sum_{k=1}^{p} \phi(k) \lambda_p^k$. Expanding then the hyperbolic sine and cosine functions in power series, inverting the summation order in the obtained expansion of g and recalling by (A.4) that

$$g(u) = i^{p+1} u^p H(u) = i^{p+1} \sum_{n \geq 0} G(n) \left(\frac{(-1)^{p+1} \zeta(2p)}{\pi^{2p}} \right)^n u^{(2n+1)p},$$

we can identify like powers of u to get, for all n:

$$i^{p+1} G(n) \left(\frac{(-1)^{p+1} \zeta(2p)}{\pi^{2p}} \right)^n = \frac{1}{2^{p-1} (2np+p)!} \sum_{\phi} \mathbf{R}_\phi \mathbf{S}_\phi^{2np+p}.$$

Defining $\Psi(p,n) = \sum_\phi \mathbf{R}_\phi \ \mathbf{S}_\phi^{2np+p}$, we eventually obtain

$$G(n) = \frac{1}{2^{p-1} \ i^{p+1}} \left(\frac{(-1)^{p+1} \pi^{2p}}{\zeta(2p)} \right)^n \frac{\Psi(p,n)}{(2np+p)!} \qquad \text{(A.6)}$$

so that (2.3) provides

$$M(C) = \frac{(2p)!}{2^{2p-1} \ B_{2p}} \frac{(C+1)}{\prod_{j=p+1}^{3p} (2pC + j)} \frac{\Psi(p, C+1)}{\Psi(p,C)}. \qquad \text{(A.7)}$$

For illustration, let us explicit the residual factor $\Psi(p, C+1)/\Psi(p,C)$ according to the value of integer p:

- **Case** $\mathbf{p = 1}$ ($\mathbf{s = 2}$) $\lambda_1 = e^{i\pi} = -1$ and $\Psi(1,n) = -1$ for all n, so that $\Psi(1, C+1)/\Psi(1,C) = 1 = \beta_2$;
- **Case** $\mathbf{p = 2}$ ($\mathbf{s = 4}$) $\lambda_2 = e^{i\pi/2} = i$, $\phi_1(1) = -1$ or 1 and $\Psi(2,n) = (i-1)^{4n+2} - (-i-1)^{4n+2} = i \ (-1)^{n+1} \ 2^{2n+2}$ for all n, so that $\Psi(2, C+1)/\Psi(2,C) = -4 = \beta_4$;
- **Case** $\mathbf{p = 3}$ ($\mathbf{s = 6}$) here $\lambda_3 = e^{i\pi/3}$ and $\mathbf{S}_\phi$ takes only 4 values: $\lambda_3 + \lambda_3^2 - 1 = 2\lambda_3^2$, $\lambda_3 - \lambda_3^2 - 1 = 0$, $-\lambda_3 + \lambda_3^2 - 1 = -2$ and $-\lambda_3 - \lambda_3^2 - 1 = -2\lambda_3$. We thus have $\Psi(3,n) = (\lambda_3 + \lambda_3^2 - 1)^{6n+3} - (\lambda_3 - \lambda_3^2 - 1)^{6n+3} - (-\lambda_3 + \lambda_3^2 - 1)^{6n+3} + (-\lambda_3 - \lambda_3^2 - 1)^{6n+3} = 3 \cdot 2^{6n+3}$, so that $\Psi(3, C+1)/\Psi(3,C) = 2^6 = \beta_6$;
- **Case** $\mathbf{p = 4}$ ($\mathbf{s = 8}$) $\lambda_4 = e^{i\pi/4}$ and $\mathbf{S}_\phi$ takes only 8 values; after some lengthy but elementary calculations, we obtain

$$\begin{aligned} \frac{\Psi(4, C+1)}{\Psi(4,C)} &= -64 \ (17 + 12\sqrt{2}) \ \frac{1 + (17 - 12\sqrt{2})^{2C+3}}{1 + (17 - 12\sqrt{2})^{2C+1}} \\ &= \beta_8 \ \Delta_8(C) \end{aligned} \qquad \text{(A.8)}$$

where $\beta_8 = -\left[\sin(\pi/8)\right]^{-8} = -64 \ (17 + 12\sqrt{2})$. Besides, $\Delta_8(C)$ tends exponentially fast to 1 when C grows to infinity.

At this stage, Proposition 3.1 is proved for $s \in \{2, 4, 6, 8\}$. It remains to prove the general result for any even $s = 2p$; thanks to the intermediate result (A.7), one has only to show that $\Psi(p, C+1)/\Psi(p,C)$ can be written as $\beta_{2p} \ \Delta_{2p}(C)$ with $\Delta_{2p}(C)$ having the desired behaviour as $C \uparrow +\infty$.

- **General case** $\mathbf{s = 2p}$ Since $\mathbf{R}_\phi$ is by definition a product of some "1" and "-1", we have $\mathbf{R}_\phi^{2n+1} = \mathbf{R}_\phi$ and we can write

$$\Psi(p,n) = \sum_{\phi} \left(\mathbf{R}_\phi \ \mathbf{S}_\phi^p \right)^{2n+1}, \quad n \in \mathbb{N}. \qquad \text{(A.9)}$$

In the latter expression of $\Psi(p,n)$, we look for the prominant terms of the sum. For each mapping ϕ, define the integer k_1 (depending on ϕ) as the minimum index $k \in \{1, 2, ..., p\}$ such that $\phi(k) = 1$. $\mathbf{S}_\phi$ may thus be rewritten as

$$\mathbf{S}_\phi = \lambda_p^{k_1} + \sum_{k=k_1+1}^{p-1} \phi(k)\lambda_p^k + \sum_{k=p}^{p+k_1-1} \lambda_p^k. \tag{A.10}$$

For any specific mapping $\phi = \phi_{k_1}$ for which $\phi_{k_1}(k) = 1$ for all k in $\{k_1+1, ..., p-1\}$, expression (A.10) can be easily written as a sum of consecutive terms of a geometric series. This gives

$$\mathbf{S}_{\phi_{k_1}} = \sum_{k=k_1}^{p+k_1-1} \lambda_p^k = 2\frac{\lambda_p^{k_1}}{1-\lambda_p}$$

which, by equality $\mathbf{R}_{\phi_{k_1}} = (-1)^{k_1-1}$, leads to

$$\mathbf{R}_{\phi_{k_1}} \mathbf{S}^p_{\phi_{k_1}} = -\left(\frac{2}{1-\lambda_p}\right)^p = \frac{i^{p+1}}{(\sin \pi/2p)^p} \tag{A.11}$$

which is a constant ω_p depending on p only. Consequently, there are in (A.9) exactly p terms equal to ω_p^{2n+1}, i.e. those corresponding to each possible k_1 in $\{1, 2, ..., p\}$. We now claim that $|\omega_p|$ is a maximum of the module $|\mathbf{R}_\phi \mathbf{S}_\phi^p|$ among all mappings ϕ, that is, $\mathbf{R}_\phi \mathbf{S}_\phi^p = K_\phi\, \omega_p$ with $|K_\phi| < 1$ for all ϕ different from any of the $(\phi_{k_1})_{k_1 \in \{1,2,...,p\}}$. For the sake of conciseness, we here omit the details of that proof which mainly relies on a careful evaluation of the difference $|\mathbf{S}_{\phi_{k_1}}|^2 - |\mathbf{S}_\phi|^2$, showing that it is strictly positive as soon as the set $U^-(\phi) = \{k \in \{k_1+1, ..., p-1\};\ \phi(k) = -1\}$ is not empty for the considered mapping ϕ.

Once this key result is established, we can complete the proof: equation (A.9) now becomes

$$\Psi(p, n) = \omega_p^{2n+1} \left(p + \sum_{\phi \neq \phi_{k_1}} K_\phi^{2n+1} \right)$$

for all $n \in \mathbb{N}$ and the residual factor in expression (A.7) of the miss rate $M(C)$ now reads

$$\frac{\Psi(p, C+1)}{\Psi(p, C)} = \omega_p^2 \, \frac{1 + \frac{1}{p}\sum_{\phi \neq \phi_{k_1}} K_\phi^{2C+3}}{1 + \frac{1}{p}\sum_{\phi \neq \phi_{k_1}} K_\phi^{2C+1}} = \beta_{2p}\ \Delta_{2p}(C)$$

with function $\Delta_{2p}(C)$ tending exponentially fast to 1 as C infinitely grows.

What makes cases $p = 1, 2, 3$ particularly simple is that if $p = 1$ or $p = 2$, then $U^-(\phi) = \varnothing$ for any mapping ϕ, and if $p = 3$, the only mapping ϕ for which $U^-(\phi) \neq \varnothing$ is $\{\phi(1) = 1,\ \phi(2) = -1\}$, and verifies $\mathbf{S}_\phi = 0$ as shown in the case $p = 3$ above. In all three situations, we thus have $\Delta_{2p}(C) = 1$ for all C, and a rational expression for the miss rate $M(C)$ holds in those three cases only.

B. PROOF OF LEMMA 4.1

(i) From the truncated Zipf distribution (4.1), the power series expansion in z of the logarithm of (2.5) gives

$$\log F_N(z) = \sum_{1 \leq j \leq N} \sum_{k \geq 1} \frac{(-1)^{k-1}}{k} \frac{z^k}{\zeta_N(s)^k\, j^{ks}} = \sum_{k \geq 1} (-1)^{k-1} \frac{z^k}{k} \frac{\zeta_N(ks)}{\zeta_N(s)^k}$$

for $|z| < \zeta_N(s)$, and the exponentiation then provides

$$F_N(z) = \sum_{m \geq 0} \frac{1}{m!} \left(\sum_{k \geq 1} \frac{(-1)^{k-1}}{k} \frac{\zeta_N(ks)}{\zeta_N(s)^k} z^k \right)^m = 1 + \sum_{m \geq 1} \frac{1}{m!} \sum_{C \geq m} \left(\sum_{k_1 + ... + k_m = C} \prod_{\ell=1}^m \frac{(-1)^{k_\ell - 1}}{k_\ell} \frac{\zeta_N(k_\ell s)}{\zeta_N(s)^{k_\ell}} \right) z^C$$

from which expression (4.2) for coefficient $G_N(C)$ follows.

(ii) For $m \in \{1, ..., C\}$, we evaluate the sum $T_N^m(C)$. Write $\mathbf{k}_m$ for vector $(k_1, ..., k_m) \in \mathbb{N}^m$ with $\mathrm{Tr}(\mathbf{k}_m) = k_1 + ... + k_m$, and introduce $\mathbb{L}^m = \{\mathbf{k}_m \in \mathbb{N}^m, | k_1 > 1, ..., k_m > 1\}$. Expanding expression (4.3) for $T_N^m(C)$ according to the number r of indexes $j \in \{1, ..., m\}$ for which $k_j = 1$ (there are $\binom{m}{r}$ ways to choose such r-tuples), then applying the definition (4.5) of $V_N^{m,r}$ coefficients, we obtain

$$T_N^m(C) = \frac{1}{\zeta_N(s)^C} \sum_{r=0}^m \binom{m}{r} \zeta_N(s)^r \sum_{\mathbf{k}_{m-r} \in \mathbb{L}^{m-r}\ |\ \mathrm{Tr}(\mathbf{k}_{m-r}) = C-r} \prod_{\ell=1}^{m-r} \frac{\zeta_N(k_\ell s)}{k_\ell} = \sum_{r=0}^m \binom{m}{r} \zeta_N(s)^{r-C}\ V_N^{C-m, m-r}.$$

Identity (4.2) then yields, after interverting summations,

$$G_N(C) = \sum_{r=0}^C \frac{\zeta_N(s)^{r-C}}{r!} \sum_{m=\max(r,1)}^C \frac{(-1)^{C-m}}{(m-r)!} V_N^{C-m, m-r}.$$

Coefficient $U_N(C)$ consequently reads

$$U_N(C) = \sum_{q=0}^C \frac{C!}{(C-q)!} \frac{1}{\zeta_N(s)^q} \sum_{n=\max(0, 1-C+q)}^q \frac{(-1)^{q-n}}{n!}\ V_N^{q-n, n}$$

after using index changes $q = C - r$ and $n = m + q - C$, successively. As the term for $q = 0$ reduces to $V_N^{0,0} = 1$ and $V_N^{q-n,n} = 0$ for $q - n < n$, that is, $n > q/2$, formula (4.4) finally follows.

C. PROOF OF PROPOSITION 4.2

We now apply large deviations estimate (4.17) for $G_N(C)$. From expansion (2.3), we derive

$$M_N(C) \sim \delta N\ \frac{\theta_{N,C}^C}{\theta_{N,C+1}^{C+1}}\ \frac{F_N(\theta_{N,C+1})}{F_N(\theta_{N,C})} \tag{C.1}$$

for large $C = \delta N$ and given $s \in\ ]0, 1]$. In the following, ratio $\theta_{N,C}^C / \theta_{N,C+1}^{C+1}$ together with $F_N(\theta_{N,C+1})/F_N(\theta_{N,C})$ are estimated for large N to make asymptotic (C.1) explicit.

a) To proceed with the estimation of $\theta_{N,C}$, consider the variable change $z = \zeta_N(s)\ (N/y)^s$. Accordingly, the positive real number $\xi_{N,C}$ such that

$$\theta_{N,C} = \zeta_N(s) \left(\frac{N}{\xi_{N,C}} \right)^s \tag{C.2}$$

is the unique solution to $f_N(y) = C$, where

$$f_N(y) = \sum_{j=1}^N \frac{N^s}{N^s + j^s y^s}, \quad y > 0. \tag{C.3}$$

The estimation of $\theta_{N,C}$ is thus equivalent to that of $\xi_{N,C}$. Now define

$$f(y; u) = \frac{N^s}{N^s + u^s y^s}, \quad u \geq 1,\ y \geq 0, \tag{C.4}$$

so that $f_N(y) = \sum_{j=1}^{N} f(y;j)$. The Euler-MacLaurin summation formula ([11], Chap.VI, Sect.2, formula (16.4)) applied to first order to function f defined in (C.4) yields

$$f_N(y) = \int_1^N f(y;v)\,\mathrm{d}v + \frac{1}{2}\,[f(y;N) + f(y;1)]$$
$$+ \frac{1}{12}[f'(y;N) - f'(y;1)] + \int_1^N \frac{P_3(v)}{3!} f^{(3)}(y;v)\mathrm{d}v. \quad \text{(C.5)}$$

To evaluate this expansion, the derivatives of f with respect to the second variable v are required up to the third order, for given $y \geq 0$. We easily verify, in particular, that $f^{(3)}(y;v)$ is negative for all y and v; we thus obtain

$$\left| \int_1^N P_3(v)\; f^{(3)}(y;v)\;\mathrm{d}v \right| \leq M\left(f''(y;1) - f''(y;N)\right)$$

where $M = \sup_{v \in [1,N]} |P_3(v)|$. Calculating each term of (C.5), we get

$$f_N(y) = NL(y) - L\left(\frac{y}{N}\right) + \frac{1}{2}\left(\frac{1}{1+y^s/N^s} + \frac{1}{1+y^s}\right) + O(N^{-s}). \quad \text{(C.6)}$$

Taking $y = \xi_{N,C}$ in (C.6) and since $f_N(\xi_{N,C}) = C = \delta N$ by definition, we obtain

$$L(\xi_{N,C}) = \delta + \frac{1}{2N}\,\frac{\xi^s}{1+\xi^s} + O\left(\frac{1}{N^{s+1}}\right). \quad \text{(C.7)}$$

Function L is continuously differentiable on $]0,+\infty[$, with non vanishing derivative; applying its inverse L^{-1} to both sides of (C.7) gives

$$\xi_{N,C} = \xi + \frac{1}{2N}\,\frac{\xi^s}{1+\xi^s}\,(L^{-1})'(\delta) + O\left(\frac{1}{N^{s+1}}\right). \quad \text{(C.8)}$$

Since $f_N(\xi_{N,C+1}) = C + 1 = \delta N + 1$, we similarly have

$$L(\xi_{N,C+1}) = \delta + \frac{1}{2N}\,\frac{2+3\xi^s}{1+\xi^s} + O\left(\frac{1}{N^{s+1}}\right) \quad \text{(C.9)}$$

and obtain the expansion

$$\xi_{N,C+1} = \xi + \frac{1}{2N}\,\frac{2+3\xi^s}{1+\xi^s}\,(L^{-1})'(\delta) + O\left(\frac{1}{N^{s+1}}\right). \quad \text{(C.10)}$$

Let us now evaluate the ratio $\theta_{N,C}^{C}/\theta_{N,C+1}^{C+1}$ in the r.h.s. of (C.1). Setting $\alpha = (L^{-1})'(\delta)$, we first note that

$$\alpha = \frac{1}{L'(\xi)} = \frac{\xi(1+\xi^s)}{1 - \delta(1+\xi^s)} \quad \text{(C.11)}$$

is negative; by formula (C.2) together with (C.8) and (C.10), we then have

$$\frac{\theta_{N,C}^{C}}{\theta_{N,C+1}^{C+1}} = \frac{1}{\zeta_N(s)}\,\frac{\xi^s}{N^s}\left[1 + \frac{s}{2N}\,\frac{2+3\xi^s}{1-\delta(1+\xi^s)} + O\left(\frac{1}{N^{s+1}}\right)\right]$$
$$\times\left[1 + \frac{1}{N}\,\frac{1+\xi^s}{1-\delta(1+\xi^s)} + O\left(\frac{1}{N^{s+1}}\right)\right]^{s\delta N}$$

which eventually provides

$$\frac{\theta_{N,C}^{C}}{\theta_{N,C+1}^{C+1}} \sim \frac{1}{\zeta_N(s)}\,\frac{\xi^s}{N^s}\exp\left(\frac{s\,\delta(1+\xi^s)}{1-\delta(1+\xi^s)}\right). \quad \text{(C.12)}$$

b) To evaluate $F_N(\theta_{N,C+1})/F_N(\theta_{N,C})$ in the r.h.s. of (C.1), apply expansion (4.14) to $z = \theta_{N,C} = \zeta_N(s)\;(N/\xi_{N,C})^s$ and $z = \theta_{N,C+1} = \zeta_N(s)\;(N/\xi_{N,C+1})^s$, successively. From (C.8) and (C.10), we first deduce

$$\theta_{N,C+1} - \theta_{N,C} = \zeta_N(s)\left[\frac{-s\alpha}{\xi^{s+1}N^{1-s}} + O\left(\frac{1}{N}\right)\right] \quad \text{(C.13)}$$

with $\alpha = (L^{-1})'(\delta)$. We now evaluate each term of expansion (4.14); from expressions (C.7) and (C.9), respectively, we express the difference in the $H(.)$ terms as

$$[H(\xi_{N,C+1}) - H(\xi_{N,C})]\;N = s\left(1 - \frac{\alpha}{\xi(1+\xi^s)}\right) + O\left(\frac{1}{N^s}\right); \quad \text{(C.14)}$$

besides, a straightforward calculation gives

$$\frac{1}{2}\left[\log\left(1+\xi_{N,C+1}^s\right) - \log\left(1+\xi_{N,C}^s\right)\right] = O\left(\frac{1}{N}\right). \quad \text{(C.15)}$$

The final step is to evaluate $T_N(\theta_{N,C+1}) - T_N(\theta_{N,C})$; we have

$$T_N(\theta_{N,C+1}) - T_N(\theta_{N,C})$$
$$= \int_1^N \frac{P_3(t)}{6}\left(h^{(3)}(\theta_{N,C+1};t) - h^{(3)}(\theta_{N,C};t)\right)\mathrm{d}t \quad \text{(C.16)}$$

where P_3 is bounded by constant M. For each $t \in [1,N]$, the Mean Value Theorem implies that

$$h^{(3)}(\theta_{N,C+1};t) - h^{(3)}(\theta_{N,C};t) = (\theta_{N,C+1} - \theta_{N,C})\,\frac{\partial h^{(3)}}{\partial z}(\theta_{t,N,C};t)$$

for some $\theta_{t,N,C} \in [\theta_{N,C}, \theta_{N,C+1}]$, which provides the upper bound

$$|T_N(\theta_{N,C+1}) - T_N(\theta_{N,C})|$$
$$\leq \frac{M}{6}\,|\theta_{N,C+1} - \theta_{N,C}|\int_1^N \left|\frac{\partial h^{(3)}}{\partial z}(\theta_{t,N,C};t)\right|\mathrm{d}t. \quad \text{(C.17)}$$

From the explicit expressions of derivatives $h''(z;t)$, $h^{(3)}(z;t)$ and $\partial_z h^{(3)}(z;t)$, and the fact that $\theta_{t,N,C} = \zeta_N(s)\;(N/\xi_{t,N,C})^s$ where $\xi_{t,N,C}$ has a positive lower bound and an upper bound, both independent of $t \in [1,N]$, we then deduce that there exist some positive constants M_1 and M_2 such that

$$\left|\frac{\partial h^{(3)}}{\partial z}(\theta_{t,N,C};t)\right| \leq \frac{M_1\zeta_N(s)^3N^{3s}}{t^3\theta_{t,N,C}^4} \leq \frac{M_2}{\zeta_N(s)N^st^3} \quad \text{(C.18)}$$

for all $t \in [1,N]$. Gathering (C.13), (C.17) and (C.18), we conclude that

$$|T_N(\theta_{N,C+1}) - T_N(\theta_{N,C})| \leq \frac{M_3}{N}\int_1^N \frac{\mathrm{d}t}{t^3} = O\left(\frac{1}{N}\right) \quad \text{(C.19)}$$

for some positive constant M_3. Recalling expression (C.11) of $\alpha = (L^{-1})'(\delta)$ and gathering expansions (C.14), (C.15) and (C.19), expansion (4.14) finally provides

$$\log\frac{F_N(\theta_{N,C+1})}{F_N(\theta_{N,C})} = \frac{s\;\delta(1+\xi^s)}{\delta(1+\xi^s) - 1} + O\left(\frac{1}{N^s}\right).$$

c) It follows from (C.1) and the latter evaluations that

$$M_N(C) \sim \frac{\delta\;\xi^sN^{1-s}}{\zeta_N(s)} \quad \text{(C.20)}$$

when $C = \delta N$ and $N \uparrow +\infty$. As $\zeta_N(s) \sim N^{1-s}/(1-s)$ for $0 < s < 1$ and $\zeta_N(1) \sim \log N$, the results claimed in (4.18) and (4.19) eventually follow from (C.20).