



Fluid limit analysis of FIFO and RR caching for independent reference models

Naoki Tsukada^a, Ryo Hirade^{a,b}, Naoto Miyoshi^{a,*}

^a Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Japan

^b IBM Research–Tokyo, Japan

ARTICLE INFO

Article history:

Received 12 July 2011

Received in revised form 19 December 2011

Accepted 29 May 2012

Available online 15 June 2012

Keywords:

Caching algorithms

First-in first-out

Random replacement

Fluid limit analysis

Mean field interaction models

ABSTRACT

We study the fluid limit analysis of random replacement (RR) caching for the independent reference model. Applying the limit theorem for the mean field interaction model, we derive the fluid limit of fault probability in the transient state as well as in the steady state. Since the stationary fault probability for the RR cache is identical to that for the first-in first-out (FIFO) cache, our results on the stationary fault probability are available for the FIFO caching. We see that the fluid limit of stationary fault probability that we obtain is coincident with the existing heuristic approximation of stationary fault probability. That is, our fluid limit analysis gives a rigorous theoretical foundation to the heuristic approximation.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In computer systems and communication networks, a caching technique is used for high-speed access to a finite subset out of a large number of items by storing the subset in a quickly accessible memory, called a cache. For this technique to work well, a replacement rule called a caching algorithm is crucial; that is, which items should be stored in the cache and how they should be updated. The performance of a caching algorithm is often evaluated in terms of the fault probability, which is the probability with which the requested item is not found in the cache.

The least-recently-used (LRU) and first-in first-out (FIFO) caching algorithms are well-known simple algorithms and have been studied in the literature. To keep frequently requested items in the cache, the LRU algorithm works as follows: when the requested item is not in the cache, the least recently requested item in the cache is replaced with the requested one. In the FIFO algorithm, on the other hand, when the requested item is not in the cache, the oldest item in the cache is replaced with the requested one. While the LRU algorithm performs relatively well and has been applied in many systems, the FIFO is shown to have higher fault probability than the LRU for the independent reference model, where requests for items are independent and identically distributed (i.i.d.) (see, e.g., [1]). Nevertheless, since some complex caching algorithms combining the LRU and FIFO have been proposed recently, such as the Full2Q by Johnson and Shasha [2] and the Multi-Queue by Zhou et al. [3], it is still meaningful to study the FIFO algorithm to evaluate the performance of such complex algorithms.

The existing works concerning the analysis of the FIFO algorithm have almost all been done for the independent reference model. King [4] considered a homogeneous Markov chain representing the evolution of cache contents and derived the stationary fault probability for the FIFO cache, as well as for the LRU. The evaluation of fault probability based on King's analysis, however, suffers from the computational complexity when the number of items and/or the capacity of cache

* Correspondence to: 2-12-1-W8-52 Ookayama, Tokyo 152-8552, Japan.

E-mail address: miyoshi@is.titech.ac.jp (N. Miyoshi).

URL: <http://www.is.titech.ac.jp/~miyoshi> (N. Miyoshi).

are large. Dan and Towsley [5] then presented a computationally efficient approximation evaluating the stationary fault probability for both the FIFO and LRU caches. While their approximation reportedly agrees well with simulation results, the argument for derivation is rather informal. Also, Gelenbe [6] showed the identity of stationary fault probability for the FIFO cache with that for the random replacement (RR) cache, where an item is chosen uniformly at random in the cache and is replaced with the requested one if it is not in the cache.

Since the stationary fault probabilities for the FIFO and RR are identical [6], we, in this paper, investigate the RR caching algorithm for the independent reference model, instead of the FIFO. We here study the fluid limit analysis of RR caching. As for the LRU, several works have considered the fluid limit analysis and presented some simple expressions evaluating the fault probability (see, e.g., Jelenković [7], and also Hirade and Osogami [8] for some complex caching algorithms, and Hattori and Hattori [9] for the related move-to-front list). It is, however, difficult to apply the techniques therein to the RR caching as well as to the FIFO, since these previous works built on such a special feature of LRU that an item is pushed out from the cache when the number of other distinct items requested at least once reaches the cache capacity before the same item is requested again. We thus associate the stochastic model of RR caching with a mean field interaction model studied by Benaïm and Le Boudec [10] (see also [11]) and apply the limit theorem for it. While the limit theorem for the mean field interaction model of [10] captures the transient behavior of the model in average, it does not yield satisfying results on the stationary behavior. For the stationary analysis, we then apply the recent results obtained by Le Boudec [12] exploiting the reversibility of the stochastic model under consideration. We can see that the fluid limit of stationary fault probability, which we obtain, is coincident with the approximated fault probability provided by [5], so that we can say that our fluid limit analysis gives a rigorous theoretical foundation to Dan–Towsley’s approximation.

The rest of this paper is organized as follows. In the next section, we describe our stochastic model of a caching system and briefly review some related existing results in the works of [4,6,5]. The main results are given in Section 3, where we first derive the fluid limit of empirical measure for the cache contents in the transient state and then obtain the corresponding fluid limit in the steady state. The fluid limits of fault probabilities are derived from those of the empirical measure. The proofs are provided in Section 4, where we apply the limit theorem for the mean field interaction model of [10] for the proof of transient result and apply the results obtained by [12] for the steady state. Finally, we make some concluding remarks in Section 5.

2. Model and related existing results

The model consists of the set of items and a buffer with finite capacity, called a cache. The number of items is N ($\in \mathbb{N} = \{1, 2, \dots\}$), and the set of items is denoted by $\mathcal{N} = \{1, 2, \dots, N\}$. We assume that all items have the same size and that the cache has the capacity of K ($< N$) items and is always full of items. An item is requested randomly among \mathcal{N} at each time slot. If the requested item is in the cache, then no change occurs. When the requested item is not in the cache, we call it a cache fault. What happens when a cache fault occurs depends on the caching algorithm. In this paper, we consider the following two algorithms.

First-in first-out (FIFO): When the requested item is not in the cache, the item that stays in the cache for the longest time is replaced with the requested one.

Random replacement (RR): When the requested item is not in the cache, an item is chosen uniformly at random among ones in the cache and replaced with the requested one.

Throughout the paper, we consider the independent reference model, where requested items are independent and identically distributed for all time slots. The probability with which item i ($\in \mathcal{N}$) is requested is denoted by p_i , satisfying $p_i \geq 0$ for all $i \in \mathcal{N}$ and $\sum_{i=1}^N p_i = 1$.

In the remainder of this section, we make a brief review on related existing results for the independent reference models of FIFO and RR caches: (i) King’s exact stationary analysis of FIFO caching [4], (ii) Gelenbe’s identity of stationary fault probabilities for the FIFO and RR caches [6] and (iii) Dan–Towsley’s approximation of stationary fault probability for the FIFO cache [5]. Though they also analyzed the LRU caching, we here omit its results. For the moment, we assume that $p_i > 0$ for all $i \in \mathcal{N}$ (while we relax this assumption in the next section).

King’s exact stationary analysis of FIFO caching [4]. For the FIFO caching algorithm, King [4] considered the evolution of cache contents as a homogeneous Markov chain and derived the fault probability in the steady state. Consider the list of K items corresponding to the contents of FIFO cache as follows. When the requested item is found in the list, the list remains unchanged. On the other hand, if the requested item is not in the list, then it is placed at the first position of the list, other items are shifted one position down, and the item at the last (K th) position is pushed out. The item at the first position is then the newest in the cache and the item at the last position is the oldest. This list forms a homogeneous Markov chain within the state space $\Lambda_{N,K}$, which denotes the set of K -permutations (arrangements of K elements) taken from $\mathcal{N} = \{1, 2, \dots, N\}$; that is, $\Lambda_{N,K} = \{(i_1, i_2, \dots, i_K) \in \mathcal{N}^K : i_k \neq i_\ell \text{ for } k \neq \ell\}$. Since this Markov chain is irreducible in the finite state space, the unique stationary distribution exists and is given by

$$\pi_{\text{FIFO}}(i_1, i_2, \dots, i_K) = \frac{p_{i_1} p_{i_2} \cdots p_{i_K}}{\sum_{(j_1, \dots, j_K) \in \Lambda_{N,K}} p_{j_1} p_{j_2} \cdots p_{j_K}}, \quad (i_1, i_2, \dots, i_K) \in \Lambda_{N,K}. \quad (1)$$

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات