A simple proof of optimality for the MIN cache replacement policy

Mun-Kyu Lee, Pierre Michaud, Jeong Seop Sim, Daehun Nyang

A simple proof of optimality for the MIN cache replacement policy

Mun-Kyu Lee*, Pierre Michaud†, Jeong Seop Sim*, DaeHun Nyang*

Abstract

The MIN cache replacement algorithm is an optimal off-line policy to decide which item to evict when a new item should be fetched into a cache. Recently, two short proofs were given by van Roy [B. van Roy, A short proof of optimality for the MIN cache replacement algorithm, Inform. Process. Lett. 102 (2007) 72-73] and Vogler [W. Vogler, Another short proof of optimality for the MIN cache replacement algorithm, Inform. Process. Lett. 106 (2008) 219–220]. We provide a simpler proof based on a novel invariant condition maintained through an incremental procedure.

1 Introduction

Let us consider a set Ω of items stored in slower memory and a replacement policy P. We know in advance the sequence of requests $\omega_1, \omega_2, \ldots, \omega_T$ from Ω over T time periods, where the size of each item ω_i is one data unit (a cache block, a memory page, etc.). The cache capacity is limited and fixed. We assume, without loss of generality, that the cache is initially full. Hence inserting a new item in the cache requires to evict another item. All the replacement policies start with the same initial cache content. Let $C_t^P \subset \Omega$ be the set of items stored in the cache just after ω_t is processed by P. If $\omega_t \in C_{t-1}^P$, a *hit* occurs, and $C_t^P = C_{t-1}^P$. Otherwise, a *miss* occurs, and an item (*victim*) in C_{t-1}^P should be replaced by ω_t. We denote the victim evicted by policy P at time t as v_t^P. Then, $C_t^P = (C_{t-1}^P - \{v_t^P\}) \cup \{\omega_t\}$. We define $v_t^P = NULL$ for a hit and $\{NULL\} = \varnothing$. We assume that items cannot be prefetched into the cache, i.e., we consider only *demand* policies that bring an item into the cache when that item is being requested [2].

The goal of a replacement policy is to minimize the number of misses. The MIN policy achieves this goal by replacing an item in the cache whose next request time is farthest in the future. If an item will not be requested by time T, its next request time is defined as ∞. If there are multiple items whose next request is at ∞, one of them is randomly selected as a victim. Thus, there may be more than one possibilities of sequence $v_1^{MIN}, \ldots, v_T^{MIN}$ for a given request sequence. Without loss of generality, we consider an arbitrary one among them and we call it MIN.

The MIN policy was proposed by Belady [1]. Mattson et al. [2] provided the first proof showing that MIN is an optimal *demand* policy. However, their proof is long and somewhat complicated. Recently, two short proofs were given by van Roy [3] and Vogler [4] using dynamic programming and amortized simulation techniques, respectively. We provide a more intuitive proof using an incremental procedure.

*Department of Computer and Information Engineering, Inha University, Incheon 402-751, Korea

†INRIA, Campus de Beaulieu, 35042 RENNES Cedex, France

2 Proof of Optimality for MIN

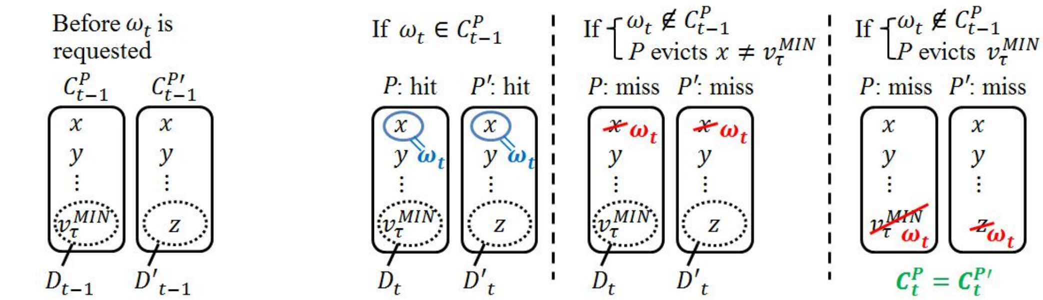
Consider two replacement policies P_1 and P_2. Given $\omega_1, \omega_2, \ldots, \omega_T$, if $v_t^{P_1} = v_t^{P_2}$ for $1 \le t \le \tau - 1$ and $v_\tau^{P_1} \neq v_\tau^{P_2}$, then $D(P_1, P_2)$, the deviation point of P_1 and P_2 is defined as τ. For convenience, we define $D(P_1, P_2)$ as $T+1$ if $v_t^{P_1} = v_t^{P_2}$ for $1 \le t \le T$. Let M_t^P be the total number of misses generated by P over $\omega_1, \omega_2, \ldots, \omega_t$.

Lemma 1 *Given any demand policy P with $D(P, MIN) = \tau$ $(1 \le \tau \le T)$, it is possible to derive a new demand policy P' with $D(P', MIN) > \tau$ which does not generate more misses than P.*

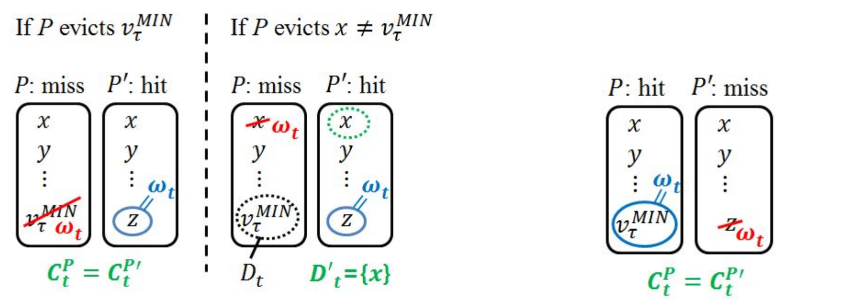
Proof. We design P' so that it imitates P, i.e., $v_t^{P'} = v_t^P$, for $1 \le t \le \tau - 1$, and $v_\tau^{P'} = v_\tau^{MIN}$. Then, $C_t^{P'} = C_t^P$ for $t \le \tau - 1$. After ω_τ is processed, $C_t^{P'} = (C_t^P - \{v_\tau^{MIN}\}) \cup \{v_\tau^P\}$ for $t = \tau$, and $M_\tau^{P'} = M_\tau^P$. In the case that $\tau = T$, this proves the lemma. If $\tau < T$, P' tries to imitate P again for $t > \tau$. To examine the possibility of this imitation, we define $D_t = C_t^P - C_t^{P'}$ and $D'_t = C_t^{P'} - C_t^P$, and show that if $C_t^P \neq C_t^{P'}$, the following invariants [I1] and [I2] always hold:

[I1] $D_t = \{v_\tau^{MIN}\}$ and $|D'_t| = 1$.

[I2] Either [I2A] $\left(D'_t = \{v_\tau^P\} \text{ and } M_t^{P'} = M_t^P\right)$ or



(a) **Cache at time $\mathbf{t-1}$** (b) **Case 1: $\omega_\mathbf{t} \notin \mathbf{D'_{t-1}}$ and $\omega_\mathbf{t} \neq \mathbf{v_\tau^{MIN}}$**



(c) **Case 2: $\omega_\mathbf{t} \in \mathbf{D'_{t-1}}$** (d) **Case 3: $\omega_\mathbf{t} = \mathbf{v_\tau^{MIN}} \in \mathbf{D_{t-1}}$**

Figure 1: Change of cache states at time t according to the request of ω_t.

[I2B] $\left(M_t^{P'} < M_t^P\right)$ holds.

Note that once the condition $C_t^P = C_t^{P'}$ is satisfied, then P' can follow exactly P thereafter and $C_u^P = C_u^{P'}$ for any $u > t$. We also see that [I1] and [I2A] (and thus [I2]) initially hold at $t = \tau$. Now we conduct a proof by induction on t. We consider the following three cases according to ω_t (see Figure 1):

Case 1 ($\omega_\mathbf{t} \notin \mathbf{D'_{t-1}}$ **and** $\omega_\mathbf{t} \neq \mathbf{v_\tau^{MIN}}$): If $\omega_t \in C_{t-1}^P$, then $\omega_t \in C_{t-1}^{P'}$. No replacement occurs both in C_{t-1}^P and $C_{t-1}^{P'}$, and the number of misses does not change for P and P', i.e., $M_t^P = M_{t-1}^P$ and $M_t^{P'} = M_{t-1}^{P'}$. Because $D_t = D_{t-1}$ and $D'_t = D'_{t-1}$, the invariants [I1] and [I2] hold. If $\omega_t \notin C_{t-1}^P$, then $\omega_t \notin C_{t-1}^{P'}$ because $\omega_t \notin D'_{t-1}$ and all the elements in $C_{t-1}^{P'} - D'_{t-1}$ are also in C_{t-1}^P (See Figure 1(a).). There are two possibilities according to the choice of P. If P evicts $x \neq v_\tau^{MIN}$, P' also evicts x. Then, because $D_t = D_{t-1}$, $D'_t = D'_{t-1}$, $M_t^P = M_{t-1}^P + 1$ and $M_t^{P'} = M_{t-1}^{P'} + 1$, the invariants [I1] and [I2] hold. If P evicts v_τ^{MIN}, P' evicts the element in D'_{t-1}, resulting in $C_t^{P'} = C_t^P$. Thereafter, P' follows P, guaranteeing that $M_u^{P'} \leq M_u^P$ for $u \geq t$.

Case 2 ($\omega_\mathbf{t} \in \mathbf{D'_{t-1}}$): This is a hit for P' and a miss for P. Therefore, invariant [I2B] holds for t. If P evicts v_τ^{MIN}, then $C_t^{P'} = C_t^P$. If P evicts $x \neq v_\tau^{MIN}$, invariant [I1] holds with $D'_t = \{x\}$.

Case 3 ($\omega_\mathbf{t} = \mathbf{v_\tau^{MIN}} \in \mathbf{D_{t-1}}$): This case causes a hit for P and a miss for P'. Then, P' replaces the item in D'_{t-1} with ω_t and follows P thereafter. If [I2B] held at time $t-1$, then $M_u^{P'} \leq M_u^P$ for $u \geq t$. On the other hand, if [I2A] held at time $t-1$, then $M_t^{P'}$ could be greater than M_t^P. However, we prove that the latter is not possible. To prove this by contradiction, let us assume that [I2A] holds at time $t-1$, which implies that only Cases 1 occurred up to time $t-1$. Because $C_u^P \neq C_u^{P'}$ for any u ($\tau \leq u \leq t-1$) by $C_{t-1}^P \neq C_{t-1}^{P'}$, we see that $D_{t-1} = D_{t-2} = \cdots = D_\tau = \{v_\tau^{MIN}\}$ and $D'_{t-1} = D'_{t-2} = \cdots = D'_\tau = \{v_\tau^P\}$. On the other hand, because at time τ, P' has selected a victim whose next request time was farthest in the future, v_τ^P must have been requested at some time u ($\tau < u \leq t-1$) before Case 3 happens. This implies that Case 2 happened at time u because $\omega_u \in D'_{u-1}$, which contradicts the assumption. □

Theorem 1 *For any demand policy P, $M_T^{MIN} \leq M_T^P$.*

Proof. By repeatedly applying the derivation procedure in Lemma 1 until $\tau = T$, we can incrementally transform P into MIN without increasing the number of misses. $\square$

Acknowledgement

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by MOE, Korea (grant number: 2014R1A1A2058514) and MSIP, Korea (grant number: 2014R1A2A1A11050337) and in part by Inha University Research Grant.

References

[1] L. A. Belady. A study of replacement algorithms for a virtual-storage computer. *IBM Systems Journal*, 5(2):78–101, 1966.

[2] R. L. Mattson, J. Gecsei, D. R. Slutz, and I. L. Traiger. Evaluation techniques for storage hierarchies. *IBM Systems Journal*, 9(2):78–117, 1970.

[3] B. van Roy. A short proof of optimality for the MIN cache replacement algorithm. *Information Processing Letters*, 102:72–73, 2007.

[4] W. Vogler. Another short proof of optimality for the MIN cache replacement algorithm. *Information Processing Letters*, 106:219–220, 2008.