



低次キャッシュとプリフェッチ

入江 英嗣¹ 本城 剛毅² 平木 敬²

¹電気通信大学大学院情報システム学研究科 ²東京大学大学院情報理工学系研究科

[受賞論文]

動的推定によるプリフェッチ量最適化

入江英嗣(電気通信大学大学院情報システム学研究科), 本城剛毅, 平木敬(東京大学大学院情報理工学系研究科)
情報処理学会論文誌 コンピューティングシステム, Vol.3, No.3, pp.56-66 (2010)

「プリフェッチ」は、過去のアクセス履歴などから次のアドレスを予測し、事前にキャッシュを準備する技術である。プリフェッチの面白さは、この技術が、長大なメモレイテンシを隠蔽し得る数少ない手段である点にある。他の方法では向上が難しかったアプリケーションの性能が、積極的なプリフェッチによって数倍にも改善する。もちろん、この効果を得るためには、「次のアドレス」だけではなく、「数回先のアドレス」や「必要となるタイミング」など、より困難な予測を安定させなくてはならない。標記の論文では、プリフェッチ量を動的に最適化する技術を提案し、簡単なハードウェアによってプリフェッチが安定した性能を発揮できることを示した。このたび本会論文賞をいただくという名誉にあずかり、大変光栄に感じている。

本研究のきっかけは、低次キャッシュのプリフェッチでは新しい視点からのバランス調整が必要だと感じたことである。高次キャッシュと異なり、低次キャッシュは大容量かつミス時のペナルティが大きい。このため、確信度が低くても積極的にプリフェッチする方が良い結果につながり、予測精度を重視する従来の戦略では慎重になりすぎてしまう。しかし一方で、度を過ぎたプリフェッチは必要ラインの追出しやメモリバンド幅の浪費を引き起こす。適切なプリフェッチ量は、そのときそのときのアプリケーションやキャッシュ容量によって大きく変化する。関連研究¹⁾では予測精度だけでなく予測適用率を併せたフィードバック制御を提案しているが、筆者らは、予測の成績ではなく、キャッシュの余裕に着目してフィードバックする方が、多様な状況に対してロバストになるのではないかと考えた。

健康状態のメタファから、排出されるものすなわち追い出されるキャッシュラインを観察する、とい

うアイデアは早い段階で決まったが、機構は試行錯誤を繰り返した。当初ウィクティムキャッシュの利用を考えたが安定せず、その後思考実験から Cache Metabolism 値という指標を思いつき、研究会で発表して意見をいただいた²⁾。この議論の過程で、キャッシュ内の対流に注目した方が明快、とご指摘をいただき、より説得力を増した Cache Convection 値という本論文の提案に至った。これらの議論や査読の過程でのご指摘がなければ本論文の今の形はなく、深く感謝している。

プリフェッチは現在のプロセッサにすぐに実装可能な即効薬であると同時に、階層記憶の動的最適化という、本質的テーマでもある。本研究を通して、ワーキングセットとキャッシュの同期、という視点を得たが、より一般化して、リソース配分や、他の記憶階層の最適化など、広く答えを出す理論に結び付かないか、チャレンジしていきたいと考えている。

参考文献

- 1) Ebrahimi, E., Mutlu, O. and Patt, Y.: Techniques for Bandwidth-Efficient Prefetching of Linked Data Structures in Hybrid Prefetching Systems, Int. Symp. on High-Performance Computer Architecture, pp.7-17 (2009).
- 2) 入江英嗣, 本城剛毅, 平木 敬: 追出しラインに着目したプリフェッチスロットリング手法, 情報処理学会研究報告計算器アーキテクチャ, 2009-ARC-184(19), 1-9 (2009).

(2011年5月16日受付)

入江 英嗣 (正会員) irie@is.uec.ac.jp

2004年東京大学大学院情報理工学系研究科博士課程修了, 博士(情報理工学)。CREST 研究員, 東京大学助教を経て, 現在電気通信大学大学院情報システム学研究科准教授。コンピュータシステムの研究に従事。

本城 剛毅 honjo@is.s.u-tokyo.ac.jp

2011年東京大学大学院情報理工学系研究科修士課程修了。同年同大学院博士課程進学。プロセッサアーキテクチャの研究に従事。

平木 敬 (正会員) hiraki@is.s.u-tokyo.ac.jp

東京大学理学系研究科物理学専門課程博士課程退学, 理学博士。電子技術総合研究所, T.J. Watson 研究センターを経て現在東京大学大学院情報理工学系研究科教授。コンピュータシステムの研究開発に従事。