

Д. В. Биматов, С. П. Сущенко (Томск, ТГУ). Об эффективности многоуровневой памяти вычислительной системы.

Рассмотрим многоуровневую подсистему памяти с количеством уровней в иерархии равным $U > 1$. На каждом уровне кроме последнего, находится кэш. Последний уровень занимает оперативное запоминающее устройство (ОЗУ). Количество блоков ОЗУ, занятых выполняемыми приложениями, полагаем равным V . Кэш каждого уровня характеризуется объемом кэша, выраженном в количестве блоков V_u , числом групп G_u , коэффициентом ассоциативности A_u , количеством блоков ОЗУ, отображаемых на группу кэша u -го уровня $M_u = V/G_u$, механизмом идеального вытеснения [1]. Важнейшей операционной характеристикой кэша заданного уровня u является вероятность попадания в кэш, определяемая соотношением [1]:

$$\Pi_u = \sum_{g=0}^{G_u-1} \sum_{m=0}^{M_u-1} \Pi_u(g + G_u m) p(g + G_u m),$$

где $\Pi_u(g + G_u m)$ — вероятность того, что m -й блок ОЗУ, отображаемый на g -ю группу кэша, находится в кэше, $p(g + G_u m)$ — вероятности востребованности блоков памяти вычислителем, упорядоченные с ростом m по убыванию. Одна из основных проблем получения вероятностей попадания в кэши различных уровней состоит в построении по известному распределению востребованности блоков ОЗУ вычислителем отображений элементов ОЗУ в группы кэша каждого уровня. Для выполнения сравнительного анализа вероятностей попадания в различные уровни иерархической памяти в работе по заданному отображению распределения востребованности блоков ОЗУ центральным процессором на кэш первого уровня получены отображения на кэши других уровней для случая, когда количество групп в кэше u -го уровня кратно количеству групп в кэше первого уровня. Поскольку механизм идеального вытеснения обладает свойством концентрации в $A_u - 1$ блоках каждой группы кэша самых востребованных вычислителем блоков памяти, то без ограничения общности можно считать, что вероятности востребованности блоков памяти, отображаемых на каждую группу кэша первого уровня, упорядочены по убыванию. Вероятность попадания в кэш уровня $u = 1, \dots, U - 1$ при этом составляет:

$$\Pi_u = \sum_{g=0}^{G_u-1} \sum_{n=0}^{e_u-1} \left[\sum_{m=0}^{A_u-2} p(g + G_1 n + G_u m) + \frac{\sum_{m=A_u-1}^{M_u-1} [p(g + G_1 n + G_u m)]^2}{\sum_{m=A_u-1}^{M_u-1} p(g + G_1 n + G_u m)} \right].$$

Здесь $e_u = G_u/G_1$. В предположении, что вероятности востребованности блоков основной памяти, отображаемых на кэш первого уровня, одинаково распределены для всех групп по усеченному геометрическому закону с параметром $q \in [0, 1]$

$$p(g + G_1 m) = \frac{(1-q)q^m}{1-q^{M_1}}, \quad g = 0, \dots, G_1 - 1, \quad m = 0, \dots, M_1 - 1,$$

получены вероятности попадания в кэши иерархической памяти:

$$\Pi_u = \frac{1 + q^{e_u} - 2q^{e_u A_u} + q^{\alpha A_1} (1 - q^{e_u})}{(1 + q^{e_u})(1 - q^{\alpha A_1})}, \quad u = 1, \dots, U - 1,$$

где α — доля ОЗУ, занятая востребованными вычислителем приложениями, выраженная в количестве объемов кэша первого уровня. Отсюда при $q = 0$ получаем $\Pi_u = 1$, а при $q = 1$ (равномерное распределение) имеем $\Pi_u = e_u A_u / \alpha A_1 = V_u / V$. Для кэшей прямого отображения, реализованных на всех уровнях иерархической памяти ($A_u = 1$), приходим к

$$\Pi_u = \frac{(1 - q^{e_u})(1 + q^\alpha)}{(1 + q^{e_u})(1 - q^\alpha)}.$$

При равенстве числа групп в кэшах первого и u -го уровней ($e_u = 1$) вероятность попадания на u -м уровне определится зависимостью:

$$\Pi_u = \frac{[1 + q - 2q^{A_u} + q^{\alpha A_1}(1 - q)]}{(1 + q)(1 - q^{\alpha A_1})}.$$

В случае, если на первом уровне иерархической памяти реализован полностью ассоциативный кэш ($G_1 = 1$), а на u -м уровне — кэш прямого отображения ($A_u = 1$), то вероятность попадания в кэш u -го уровня составит:

$$\Pi_u = \frac{(1 - q^{V_u})(1 + q^V)}{(1 + q^{V_u})(1 - q^V)}.$$

На основе полученных соотношений проведен сравнительный анализ индексов эффективности функционирования подсистем памяти с различным числом уровней иерархии. Исследованы условия целесообразности изменения числа уровней подсистемы памяти. Найдены области значений параметров быстродействия, объема и ассоциативности кэшей различных уровней, обеспечивающие снижение среднего времени доступа к адресуемым объектам заданных приложений при реинжиниринге архитектуры подсистемы памяти.

СПИСОК ЛИТЕРАТУРЫ

1. *Сущенко М. С.* Анализ производительности множественного ассоциативного кэша. — Вестник Томского гос. ун-та, 2002, № 275, с. 218–223.