

# Análisis de Errores de Cache

Luis Tong

Marzo 2017

## Introducción

En esta Práctica De Algoritmos Paralelos se hace un análisis de errores de cache. Se comparó dos algoritmos que hallan la multiplicación de matrices, donde el primer algoritmo es el clásico conocido formado por 3 bucles, mientras el segundo algoritmo es una mejora del primero usando bloques.

La comparación se realizó mediante el uso de herramientas tales como valgring y kcache-gring, quienes nos dan una evaluación más precisa en términos de L1, L2 entre otros para conocer los errores de cache.

## Comparación

Se midió mediante el tamaño de matriz en nuestro caso matrices cuadradas.

**Tamaño de matriz = 100**

Tipos	Llamantes	Todos los llamantes	Mapa de llamadas	Código fuente
Tipo de evento	Incl.	Auto	Corto	Fórmula
Obtención de instrucción	 35.46	 35.46	Ir	
Fallo al obtener instrucción de L1	0.43	0.43	I1mr	
Fallo al obtener instrucción de LL	0.45	0.45	ILmr	
Acceso de lectura de datos	 32.36	 32.36	Dr	
Fallo al leer datos de L1	 80.06	 80.06	D1mr	
Fallo al leer datos de LL	0.00	0.00	DLmr	
Acceso de escritura de datos	 27.22	 27.22	Dw	
Fallo al escribir datos en L1	0.00	0.00	D1mw	
Fallo al escribir datos en LL	0.00	0.00	DLmw	
Fallo de suma de L1	 71.24	 71.24	L1m =	I1mr + D1mr + D1mw
Fallo de suma de último nivel	0.05	0.05	LLm =	ILmr + DLmr + DLmw
Estimación de ciclo	 35.28	 35.28	CEst =	Ir + 10 L1m + 100 LLm

Figura 1: Primer Algoritmo con matriz de tamaño 100













Tipos	Llamantes	Todos los llamantes	Mapa de llamadas	Código fuente
Tipo de evento	Incl.	Auto	Corto	Fórmula
Obtención de instrucción	 36.05	 36.05	Ir	
Fallo al obtener instrucción de L1	0.48	0.48	I1mr	
Fallo al obtener instrucción de LL	0.51	0.51	ILmr	
Acceso de lectura de datos	 32.46	 32.46	Dr	
Fallo al leer datos de L1	 22.05	 22.05	D1mr	
Fallo al leer datos de LL	0.00	0.00	DLmr	
Acceso de escritura de datos	 27.25	 27.25	Dw	
Fallo al escribir datos en L1	0.00	0.00	D1mw	
Fallo al escribir datos en LL	0.00	0.00	DLmw	
Fallo de suma de L1	 14.99	 14.99	L1m =	I1mr + D1mr + D1mw
Fallo de suma de último nivel	0.06	0.06	LLm =	ILmr + DLmr + DLmw
Estimación de ciclo	 35.55	 35.55	CEst =	Ir + 10 L1m + 100 LLm

Figura 2: Segundo Algoritmo con matriz de tamaño 100

Como podemos observar en las diferentes imágenes, se marca de colores azul ,como acceso de lectura, y color verde como fallo al leer datos. Entonces hay diferencia en los eventos de Fallo al Leer el Dato de L1, donde para el Primer Algoritmo es casi 4 veces mayor que el Segundo, también notamos que en el evento de Fallo de Suma de L1 para el Primer Algoritmo es más de 3 veces mayor que el Segundo.

## Tamaño de matriz = 500













Tipo de evento	Incl.	Auto	Corto	Fórmula
Obtención de instrucción	 35.46	 35.46	Ir	
Fallo al obtener instrucción de L1	0.43	0.43	I1mr	
Fallo al obtener instrucción de LL	0.45	0.45	ILmr	
Acceso de lectura de datos	 32.36	 32.36	Dr	
Fallo al leer datos de L1	 80.06	 80.06	D1mr	
Fallo al leer datos de LL	0.00	0.00	DLmr	
Acceso de escritura de datos	 27.22	 27.22	Dw	
Fallo al escribir datos en L1	0.00	0.00	D1mw	
Fallo al escribir datos en LL	0.00	0.00	DLmw	
Fallo de suma de L1	 71.24	 71.24	L1m =	I1mr + D1mr + D1mw
Fallo de suma de último nivel	0.05	0.05	LLm =	ILmr + DLmr + DLmw
Estimación de ciclo	 35.28	 35.28	CEst =	Ir + 10 L1m + 100 LLm

Figura 3: Primer Algoritmo con matriz de tamaño 500

Tipo de evento	Incl.	Auto	Corto	Fórmula
Obtención de instrucción	36.05	36.05	Ir	
Fallo al obtener instrucción de L1	0.48	0.48	I1mr	
Fallo al obtener instrucción de LL	0.51	0.51	ILmr	
Acceso de lectura de datos	32.46	32.46	Dr	
Fallo al leer datos de L1	22.05	22.05	D1mr	
Fallo al leer datos de LL	0.00	0.00	DLmr	
Acceso de escritura de datos	27.25	27.25	Dw	
Fallo al escribir datos en L1	0.00	0.00	D1mw	
Fallo al escribir datos en LL	0.00	0.00	DLmw	
Fallo de suma de L1	14.99	14.99	L1m =	I1mr + D1mr + D1mw
Fallo de suma de último nivel	0.06	0.06	LLm =	ILmr + DLmr + DLmw
Estimación de ciclo	35.55	35.55	CEst =	Ir + 10 L1m + 100 LLm

Figura 4: Segundo Algoritmo con matriz de tamaño 500

Igualmente como el anterior caso, podemos ver la diferentes valores para estos tipos de datos, en este caso notamos que El Fallo al Leer datos de L1 para el Primer Algoritmo es también 4 veces mayor que el Segundo, y también notamos que en el evento de Fallo de Suma de L1 para el Primer Algoritmo es más de 3 veces mayor que el Segundo.

## Conclusión

En el caso del primer algoritmo el bucle más interno de su multiplicador de matriz lee filas o columnas enteras en secuencia, el caché se llena gradualmente de datos, pero el tamaño del caché es limitado, por lo que si las filas son realmente largas, el caché debe tirar lo que cargó inicialmente, para dar cabida a cosas nuevas, es decir que cuando llegue al final e inicie el siguiente y necesitará algunos datos que estuvieron recientemente en la cache, y de esta manera esperar a que regresen de la memoria otra vez.

En mi caso se muestra que el segundo algoritmo es cuatro veces más rápido, la manera que se aplica es la siguiente, igual que el primer algoritmo estará cargando los siguientes datos por bloques, pero hay menos, porque si los bloques encajan en cachés donde no pueden matrices enteras, obtendrá más del trabajo total realizado por cache, los valores de un bloque pasarán a más de un valor de resultado antes ellos son retirados, pero accederá a los siguientes datos, debido que ya esta cargado.

En conclusión el segundo algoritmo podrá tener mas bucles anidados, pero su manera de acceso a memoria es mas rápida que el primero, para hacer esto posible debemos ver la manera de cargar los datos proximos a usar en la memoria cache, y esto debemos aplicarlo desde nuestro programa, buscando la mejor manera de acomodar nuestros datos temporales y espaciales.