

Architektury systemów komputerowych

Lista zadań nr 10

Na zajęcia 14 i 15 maja 2018

UWAGA! W trakcie prezentacji rozwiązań należy być przygotowanym do wyjaśnienia pojęć, które zostały oznaczone **wytluszczoną** czcionką.

Zadanie 1. Odnieś się do poniższego cytatu w kontekście swojej praktyki programistycznej:

„Programmers waste enormous amounts of time thinking about, or worrying about, the speed of noncritical parts of their programs, and these attempts at efficiency actually have a strong negative impact when debugging and maintenance are considered. We should forget about small efficiencies, say about 97% of the time: **premature optimization is the root of all evil**. Yet we should not pass up our opportunities in that critical 3%.” — **Donald Knuth, Wikiquote**

Do czego służą programy profilujące? Czym różni się profil płaski od profilu grafu wywołań? Czemu profilowanie programu wymaga zbudowania go ze specjalną opcją kompilatora `-pg`? Odpowiadając na pytania posłuż się artykułem „**Speed your code with the GNU profiler¹**”.

Zadanie 2. Mamy system z pamięcią operacyjną adresowaną bajtowo. Szerokość szyny adresowej wynosi 12. Pamięć podręczna ma organizację sekcyjno-skojarzeniową o dwuelementowych zbiorach, a blok ma 4 bajty. Dla podanego niżej stanu pamięci podręcznej wyznacz, które bity adresu wyznaczają: offset, indeks, znacznik. Wszystkie wartości numeryczne podano w systemie szesnastkowym.

Indeks	Znacznik	Valid	B0	B1	B2	B3
0	00	1	40	41	42	43
	83	1	FE	97	CC	D0
1	00	1	44	45	46	47
	83	0	–	–	–	–
2	00	1	48	49	4A	4B
	40	0	–	–	–	–
3	FF	1	9A	C0	03	FF
	00	0	–	–	–	–

Określ, które z poniższych operacji odczytu wygenerują trafienie i ew. jakie wartości wczytają:

Adres	Trafienie?	Wartość
832
835
FFD

Zadanie 3. Rozważmy **pamięć podręczną z mapowaniem bezpośrednim adresowaną bajtowo**. Używamy adresów 32-bitowych o następującym formacie: $(\text{tag}, \text{index}, \text{offset}) = (\text{addr}_{31...10}, \text{addr}_{9...5}, \text{addr}_{4...0})$.

- Jaki jest rozmiar bloku w 32-bitowych słowach?
- Ile wierszy ma nasza pamięć podręczna?
- Jaki jest stosunek liczby bitów składających dane do liczby bitów składających metadane?

¹<http://www.ibm.com/developerworks/library/l-gnuprof.html>

Zadanie 4. Rozważmy pamięć podręczną z poprzedniego zadania. Mamy następującą sekwencję odwołań do słów pamięci:

0 4 16 132 232 160 1024 28 140 3100 180 2180

Założ, że na początku pamięć podręczna jest pusta. Jak wiele bloków zostało **zastąpionych**? Jaka jest efektywność pamięci podręcznej (liczba **trafień** procentowo)? Podaj zawartość pamięci podręcznej po wykonaniu powyższych odwołań – każdy ważny wpis wypisz jako krotkę (tag, index, ...). Dla każdego **chybienia** wskaż, czy jest ono przymusowe (ang. **compulsory miss**), wynika z ograniczonej pojemności pamięci (ang. **capacity miss**) czy też kolizji na danym adresie (ang. **conflict miss**).

Zadanie 5. Powtórz poprzednie zadanie dla następujących organizacji pamięci podręcznej:

- sekcyjno-skojarzeniowa 3-drożna, bloki długości dwóch słów, liczba bloków 24, polityka wymiany LRU,
- w pełni asocjacyjna, bloki długości słowa, liczba bloków 8, polityka wymiany LRU.

Zadanie 6. Rozważmy system z dwupoziomową pamięcią podręczną z polityką zapisu *write-back* z *write-allocate*. Dodatkowo zakładamy, że blok o określonym adresie może znajdować się tylko na jednym poziomie pamięci podręcznej (ang. *exclusive caches*). Przy pomocy drzewa decyzyjnego przedstaw jakie kroki należy wykonać, by obsłużyć chybiecie przy zapisie do L1? Nie zapomnij o bicie *dirty* i o tym, że pamięć podręczna może być całkowicie wypełniona! Zakładamy, że pamięć podręczna pierwszego poziomu nie może komunikować się bezpośrednio z pamięcią operacyjną.

Zadanie 7. Wiemy, że im większa pamięć podręczna tym dłuższy czas dostępu do niej. Załóżmy, że dostęp do pamięci głównej trwa 70ns, a dostępy do pamięci stanowią 36% wszystkich instrukcji. Rozważmy system z pamięcią podręczną o następującej strukturze: L1 – 2 KiB, współczynnik chybień 8.0%, czas dostępu 0.66ns (1 cykl procesora); L2 – 1 MiB, współczynnik chybień 0.5%, czas dostępu 5.62ns. Odpowiedz na pytania:

- Jaki jest średni czas dostępu do pamięci dla procesora tylko z cache L1, a jaki dla procesora z L1 i L2?
- Zakładając, że procesor charakteryzuje się współczynnikiem **CPI** (ang. *clocks per instruction*) równym 1.0 (bez wykonywania dostępu do pamięci), oblicz CPI dla procesora tylko z cache L1 i dla procesora z L1 i L2.

Zadanie 8. Dla czterodrożnej sekcyjno-skojarzeniowej pamięci podręcznej chcemy zaimplementować politykę zastępowania LRU. Masz do dyspozycji dodatkowe $\log_2(4!)$ bitów na zbiór. Nie można modyfikować zawartości linii w zbiorze, ani zamieniać elementów kolejnością. Jak wyznaczyć kandydata do usunięcia ze zbioru? Jak aktualizować informacje zawarte w dodatkowych bitach przy wykonywaniu dostępu do elementów zbioru?

Zadanie 9 (bonus). Wyjaśnij na czym polega problem współdzielenia pamięci w systemach wieloprocesorowych – tzw. problem spójności pamięci podręcznej (ang. *cache coherence*). Następnie wytłumacz na czym polega różnica między **protokołami kontroli spójności** – z podglądaniem (ang. *snooping*) i **katalogowym** directory. Narysuj diagram protokołu MESI i dla poszczególnych krawędzi podaj zdarzenia, które inicjują zmianę stanu.