

Spis treści

Przedmowa	15
Wprowadzenie	16
Podziękowania	21
O autorze	22
Pomiary	23
1. Mój program działa zbyt wolno	25
1.1. Kontekst centrum danych	25
1.2. Sprzęt w centrach danych	27
1.3. Oprogramowanie w centrum danych	28
1.4. Latencja z długiego ogona rozkładu	30
1.5. Model myślenia	32
1.6. Szacowanie rzędu wielkości	32
1.7. Dlaczego transakcje działają powoli?	33
1.8. Pięć podstawowych zasobów	35
1.9. Podsumowanie	35
2. Pomiary procesorów	37
2.1. Trochę historii	38
2.2. Obecna sytuacja	41
2.3. Pomiar latencji instrukcji add	42
2.4. Niepowodzenie z prostym, sekwencyjnym kodem	44
2.5. Niepowodzenia z prostą pętlą, kosztami wykonywania pętli i kompilatorem optymalizującym	44
2.6. Niepowodzenie z martwą zmienną	47
2.7. Lepsza pętla	48
2.8. Zmienne zależne	49
2.9. Faktyczna latencja wykonywania	49
2.10. Więcej niuansów	50
2.11. Podsumowanie	51

Ćwiczenia	51
3. Pomiar pamięci	53
3.1. Pomiar czasu dostępu do pamięci	53
3.2. Pamięć	54
3.3. Struktura pamięci podręcznej	56
3.4. Wyrównanie danych	59
3.5. Struktura bufora TLB	60
3.6. Pomiary	61
3.7. Pomiar wielkości wiersza pamięci podręcznej	62
3.8. Problem: wstępne wczytywanie wiersza $N + 1$	64
3.9. Odczyt uzależniony od poprzedniej operacji	65
3.10. Nielosowy dostęp do pamięci DRAM	66
3.11. Pomiar łącznej wielkości każdego poziomu pamięci podręcznej	68
3.12. Pomiar stopnia wielodrożności pamięci podręcznej na poszczególnych poziomach	70
3.13. Czas dostępu do bufora TLB	71
3.14. Niepełne wykorzystanie pamięci podręcznej	71
3.15. Podsumowanie	71
Ćwiczenia	72
4. Interakcje procesora i pamięci	74
4.1. Interakcje związane z pamięcią podręczną	75
4.2. Dynamika prostego mnożenia macierzy	76
4.3. Szacunki	77
4.4. Inicjowanie, kontrola wyników i obserwacja	78
4.5. Początkowe wyniki	78
4.6. Szybsze mnożenie macierzy metodą transpozycji	81
4.7. Szybsze mnożenie macierzy z wykorzystaniem podbloków	83
4.8. Obliczenia z uwzględnianiem pamięci podręcznej	84
4.9. Podsumowanie	85
Ćwiczenia	85
5. Pomiar dysków twardych i nośników SSD	86

5.1. Dyski twarde	86
5.2. Nośniki SSD	90
5.3. Dostęp do dysku w programie i buforowanie danych na dysku	92
5.4. Jak szybki jest odczyt danych z dysku?	94
5.5. Nieco prostych obliczeń	97
5.6. Jak szybki jest zapis danych na dysku?	99
5.7. Wyniki	99
5.8. Odczyt danych z dysku	100
5.9. Zapis danych na dysku	104
5.10. Odczyt danych z nośnika SSD	107
5.11. Zapis danych na nośniku SSD	109
5.12. Wiele transferów	109
5.13. Podsumowanie	110
Ćwiczenia	111
6. Pomiary dotyczące sieci	113
6.1. Ethernet	115
6.2. Koncentratory, przełączniki i routery	117
6.3. Protokół TCP/IP	118
6.4. Pakiety	119
6.5. Wywołania RPC	120
6.6. Niezidentyfikowany czas	122
6.7. Obserwowanie ruchu w sieci	123
6.8. Definicja przykładowego komunikatu RPC	126
6.9. Projekt rejestrowania zdarzeń	129
6.10. Przykładowy system klient-serwer oparty na wywołaniach RPC	130
6.11. Przykładowy program serwera	131
6.12. Blokady wirujące	132
6.13. Przykładowy program klienta	133
6.14. Pomiar jednego przykładowego wywołania RPC między klientem a serwerem	136
6.15. Przetwarzanie końcowe dzienników wywołań RPC	137

6.16. Obserwacje	137
6.17. Podsumowanie	139
Ćwiczenia	140
7. Dyskowa baza danych i jej interakcje z siecią	142
7.1. Wyrównywanie pomiarów czasu	142
7.2. Wiele klientów	149
7.3. Blokady wirujące	149
7.4. Pierwszy eksperyment	150
7.5. Baza danych na dysku	153
7.6. Drugi eksperyment	153
7.7. Trzeci eksperyment	158
7.8. Rejestrowanie informacji	160
7.9. Wyjaśnienie zmienności latencji transakcji	160
7.10. Podsumowanie	161
Ćwiczenia	162
Część II Obserwacja	163
8. Rejestrowanie zdarzeń	165
8.1. Narzędzia do obserwacji	165
8.2. Rejestrowanie zdarzeń	166
8.3. Podstawowe mechanizmy rejestrowania zdarzeń	166
8.4. Rozbudowane rejestrowanie zdarzeń	167
8.5. Znaczniki czasu	168
8.6. Identyfikatory wywołań RPC	169
8.7. Formaty plików dziennika	170
8.8. Zarządzanie plikami dziennika	171
8.9. Podsumowanie	172
9. Miary zagregowane	173
9.1. Zdarzenia występujące jednostajnie i seryjnie	174
9.2. Mierzone okresy	175
9.3. Oś czasu	175
9.4. Dalsze podsumowywanie osi czasu	177

9.5. Skale czasowe dla histogramów	179
9.6. Agregowanie pomiarów dotyczących zdarzeń	182
9.7. Zmiany wzorców wartości w czasie	183
9.8. Czas między aktualizacjami	184
9.9. Przykładowe transakcje	186
9.10. Podsumowanie	187
10. Panele kontrolne	189
10.1. Przykładowa usługa	189
10.2. Przykładowe panele kontrolne	191
10.3. Główny panel kontrolny	192
10.4. Panele kontrolne instancji	196
10.5. Panele kontrolne serwerów	197
10.6. Testy poprawności	197
10.7. Podsumowanie	198
Ćwiczenia	199
11. Inne dostępne narzędzia	200
11.1. Rodzaje narzędzi do obserwacji	200
11.2. Obserwowane dane	202
11.3. Polecenie top	204
11.4. Pseudopliki /proc i /sys	204
11.5. Polecenie time	205
11.6. Polecenie perf	205
11.7. Narzędzie oprofile, profiler procesora	207
11.8. Narzędzie strace, wywołania systemowe	211
11.9. Narzędzie ltrace, wywołania bibliotek języka C w procesorze	214
11.10. Narzędzie ftrace, śledzenie funkcji jądra w procesorze	214
11.11. Operacje malloc i free, narzędzie mtrace	217
11.12. Śledzenie operacji dyskowych, narzędzie blktrace	219
11.13. Śledzenie sieci, tcpdump i Wireshark	222
11.14. Blokady sekcji krytycznych, narzędzie locktrace	223
11.15. Oferowane obciążenie, wywołania wychodzące i latencja transakcji	224

11.16. Podsumowanie	225
Ćwiczenia	226
12. Ślady	227
12.1. Zalety śledzenia	227
12.2. Wady śledzenia	228
12.3. Trzy pytania na start	229
12.4. Przykład: jeden z pierwszych śladów licznika programu	232
12.5. Przykład: liczba instrukcji i czas na funkcję	234
12.6. Studium przypadku: ślady poszczególnych funkcji w serwisie Gmail	238
12.7. Podsumowanie	243
13. Zasady projektowania narzędzi do obserwacji	244
13.1. Co obserwować?	244
13.2. Jak często i jak długo?	245
13.3. Jakie koszty są dopuszczalne?	246
13.4. Konsekwencje projektowe	247
13.5. Studium przypadku: kubeczki w histogramie	247
13.6. Projektowanie sposobu wyświetlania danych	249
13.7. Podsumowanie	251
Część III Narzędzie KUtrace	253
14. KUtrace: cele, projekt, implementacja	255
14.1. Ogólne informacje	255
14.2. Cele	256
14.3. Projekt	257
14.4. Implementacja	259
14.5. Patche i moduł jądra	260
14.6. Program sterujący	261
14.7. Przetwarzanie końcowe	261
14.8. Uwagi na temat bezpieczeństwa	261
14.9. Podsumowanie	262
15. KUtrace: patche jądra Linuksa	263
15.1. Struktury danych bufora śladu	264

15.2. Format surowych bloków śladu	264
15.3. Rekordy śladu	267
15.4. Rekordy z liczbą instrukcji na cykl (I/C)	268
15.5. Znaczniki czasu	269
15.6. Numery zdarzeń	270
15.7. Zagnieżdżone rekordy śladu	270
15.8. Kod	270
15.9. Śledzenie pakietów	271
15.10. Patche dla procesorów x86-64 firm AMD i Intel	273
15.11. Podsumowanie	274
Ćwiczenia	275
16. KUTrace: wczytywany moduł dla systemu Linux	276
16.1. Struktury danych interfejsu jądra	276
16.2. Wczytywanie i zwalnianie modułu	277
16.3. Inicjowanie śledzenia i sterowanie nim	278
16.4. Implementacja wywołań do generowania śladu	278
16.5. Insert1	278
16.6. InsertN	281
16.7. Przełączanie się do nowego bloku	281
16.8. Podsumowanie	282
17. KUTrace: sterowanie w trybie użytkownika	283
17.1. Sterowanie procesem śledzenia	283
17.2. Samodzielny program kutrace_control	284
17.3. Podstawowa biblioteka kutrace_lib	285
17.4. Interfejs do sterowania wczytywanym modułem	285
17.5. Podsumowanie	286
18. Przetwarzanie końcowe w narzędziu KUTrace	287
18.1. Szczegółowe omówienie przetwarzania końcowego	287
18.2. Program rawtoevent	288
18.3. Program eventtospan	290
18.4. Program spantotrim	292

18.5. Program spantospan	292
18.6. Programy samptoname_k i samptoname_u	293
18.7. Program makeself	293
18.8. Format plików JSON w narzędziu KUtrace	293
18.9. Podsumowanie	296
19. KUtrace: wyświetlanie dynamiki działania oprogramowania	297
19.1. Wprowadzenie	297
19.2. Obszar 1 — kontrolki	298
19.3. Obszar 2 — oś y	300
19.4. Obszar 3 — osie czasu	301
19.5. Obszar 4 — legenda dotycząca liczby instrukcji na cykl	307
19.6. Obszar 5 — oś x	307
19.7. Obszar 6 — zapisywanie i wczytywanie	307
19.8. Kontrolki pomocnicze	308
19.9. Podsumowanie	309
Część IV Wnioskowanie	311
20. Na co zwracać uwagę?	313
20.1. Wprowadzenie	313
21. Wykonywanie za dużej ilości kodu	316
21.1. Wprowadzenie	316
21.2. Program	317
21.3. Zagadka	317
21.4. Pomiary i wnioski	318
21.5. Rozwiązanie zagadki	322
21.6. Podsumowanie	323
22. Powolne wykonywanie kodu	324
22.1. Wprowadzenie	324
22.2. Program	325
22.3. Zagadka	325
22.4. Konkurujący program z operacjami zmiennoprzecinkowymi	328
22.5. Konkurujący program korzystający z pamięci	330

22.6. Rozwiązanie zagadki	331
22.7. Podsumowanie	332
23. Oczekiwanie na procesor	334
23.1. Program	334
23.2. Zagadka	335
23.3. Pomiary i wnioski	335
23.4. Zagadka numer 2	336
23.5. Rozwiązanie zagadki numer 2	338
23.6. Dodatkowa zagadka	341
23.7. Podsumowanie	343
Ćwiczenia	343
24. Oczekiwanie na pamięć	344
24.1. Program	344
24.2. Zagadka	345
24.3. Pomiary i wnioski	345
24.4. Zagadka numer 2 — dostęp do tablicy stron	349
24.5. Wyjaśnienie zagadki numer 2	350
24.6. Podsumowanie	351
Ćwiczenia	351
25. Oczekiwanie na dysk	352
25.1. Program	352
25.2. Zagadka	353
25.3. Pomiary i wnioski	353
25.4. Odczyt 40 MB	356
25.5. Odczyt sekwencyjnych bloków po 4 KB	357
25.6. Odczyt losowych bloków po 4 KB	359
25.7. Zapis i synchronizacja 40 MB na nośniku SSD	361
25.8. Odczyt 40 MB z nośnika SSD	361
25.9. Dwa programy jednocześnie używające dwóch plików	362
25.10. Wyjaśnienie zagadek	364
25.11. Podsumowanie	364

Ćwiczenia	365
26. Oczekiwanie na sieć	366
26.1. Wprowadzenie	367
26.2. Programy	368
26.3. Eksperyment numer 1	368
26.4. Zagadka z eksperymentu numer 1	369
26.5. Pomiary i wnioski z eksperymentu numer 1	371
26.6. Eksperyment numer 1. A co z czasem między wywołaniami RPC?	375
26.7. Eksperyment numer 2	377
26.8. Eksperyment numer 3	377
26.9. Eksperyment numer 4	378
26.10. Wyjaśnienie zagadek	381
26.11. Dodatkowa anomalia	382
26.12. Podsumowanie	384
27. Oczekiwanie na blokady	385
27.1. Wprowadzenie	385
27.2. Program	390
27.3. Eksperyment numer 1 — długi czas utrzymywania blokady	393
27.3.1. Proste zajmowanie blokady	394
27.3.2. Nasycenie blokady	394
27.4. Zagadki w eksperymencie numer 1	395
27.5. Pomiary i wnioski w eksperymencie numer 1	395
27.5.1. Zawłaszczenie blokady	397
27.5.2. Zagłodzenie w oczekiwaniu na blokadę	397
27.6. Eksperyment numer 2 — rozwiązanie problemu zawłaszczania blokady	398
27.7. Eksperyment numer 3 — rozwiązanie problemu rywalizacji przez zastosowanie wielu blokad	399
27.8. Eksperyment numer 4 — rozwiązanie problemu rywalizacji o blokadę dzięki mniejszej ilości pracy przy zajętej blokadzie	401
27.9. Eksperyment numer 5 — eliminowanie rywalizacji o blokadę	

dzięki zastosowaniu techniki RCU dla panelu kontrolnego	402
27.10. Podsumowanie	404
28. Oczekiwanie na podstawie czasu	406
28.1. Okresowe wykonywanie pracy	406
28.2. Limity czasu	407
28.3. Podział czasu	408
28.4. Wewnętrzne opóźnienia w wykonywaniu	408
28.5. Podsumowanie	409
29. Oczekiwanie na kolejki	410
29.1. Wprowadzenie	410
29.2. Rozkład żądań	412
29.3. Struktura kolejki	413
29.4. Zadania robocze	414
29.5. Zadanie główne	414
29.6. Operacje Dequeue	415
29.7. Operacja Enqueue	415
29.8. Klasa blokady wirującej	415
29.9. Procedura odpowiedzialna za „pracę”	416
29.10. Proste przykłady	416
29.11. Co mogło pójść nie tak?	418
29.12. Częstotliwość procesora	418
29.13. Złożone przykłady	420
29.14. Oczekiwanie na procesory — dziennik wywołań RPC	420
29.15. Analiza oczekiwania na procesor za pomocą narzędzia KUtrace	421
29.16. Błąd w klasie PlainSpinLock	424
29.17. Źródłowa przyczyna	426
29.18. Poprawiona klasa PlainSpinLock zapewniająca obserwowalność	427
29.19. Równoważenie obciążenia	427
29.20. Zapewnianie obserwowalności długości kolejki	428
29.21. Aktywne oczekiwanie na końcu	429
29.22. Jeszcze jedna usterka	430

29.23. Dokładne sprawdzanie	430
29.24. Podsumowanie	431
Ćwiczenia	431
30. Podsumowanie	433
30.1. Czego udało Ci się nauczyć?	433
30.2. Czego nie omówiłem?	435
30.3. Dalsze kroki	436
30.4. Podsumowanie (całej książki)	436
A. Przykładowe serwery	439
A.1. Sprzęt z przykładowych serwerów	439
A.2. Łącza serwerów	441
B. Rekordy śladu	442
B.1. Rekordy śladu o stałej długości	443
B.2. Rekordy o zmiennej długości	443
B.3. Numery zdarzeń	444
B.3.1. Zdarzenia wstawiane przez patche narzędzia KUtrace dla jądra	445
B.3.2. Zdarzenia wstawiane przez kod trybu użytkownika	446
B.3.3. Zdarzenia wstawiane przez kod przetwarzania końcowego	447
Literatura	448
Słowniczek	456