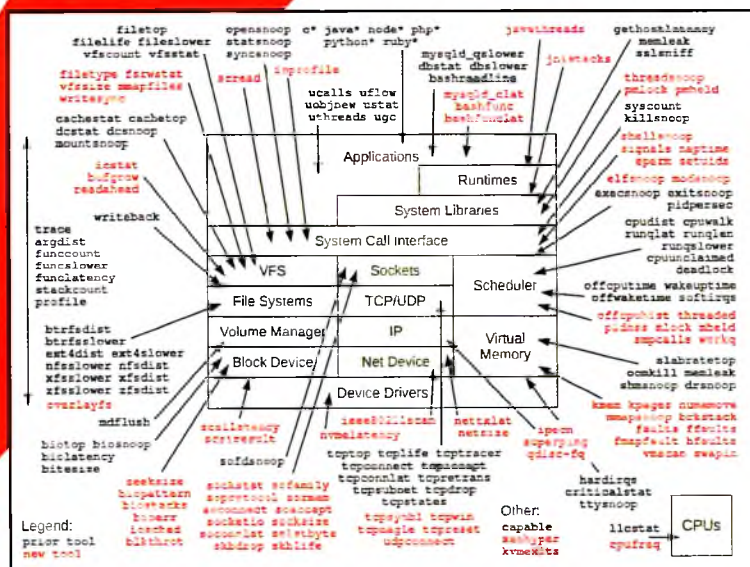


ВРФ

профессиональная
оценка
производительности

Брендан Грегг



Предисловие Алексея Старовойтова,
создателя нового ВРФ

КРАТКОЕ СОДЕРЖАНИЕ

Предисловие	29
Вступление	30
Глава 1. Введение	44
Глава 2. Основы технологии	60
Глава 3. Анализ производительности	117
Глава 4. ВСС	137
Глава 5. bpftrace	182
Глава 6. Процессоры	236
Глава 7. Память	301
Глава 8. Файловые системы	337
Глава 9. Дисковый ввод/вывод	406
Глава 10. Сети	455
Глава 11. Безопасность	556
Глава 12. Языки	584
Глава 13. Приложения	658
Глава 14. Ядро	702

Глава 15. Контейнеры.....	738
Глава 16. Гипервизоры.....	756
Глава 17. Другие инструменты BPF для анализа производительности	775
Глава 18. Советы, рекомендации и типичные проблемы	795
Приложение А. Однострочные сценарии для bpftrace	811
Приложение В. Шпаргалка по bpftrace	816
Приложение С. Разработка инструментов ВСС	819
Приложение D. С BPF.....	833
Приложение Е. Инструкции BPF.....	852
Глоссарий	858
Библиография	868

ОГЛАВЛЕНИЕ

Предисловие	29
Вступление	30
Где могут пригодиться инструменты оценки производительности BPF?.....	31
Об этой книге	31
Новые инструменты.....	32
О графическом пользовательском интерфейсе.....	32
О версиях Linux.....	33
О чем здесь не рассказывается.....	34
Структура	34
Для кого эта книга	35
Авторские права на исходный код.....	36
Дополнительные материалы и ссылки.....	37
Условные обозначения.....	37
Благодарности.....	38
Об авторе	42
От издательства.....	43
Глава 1. Введение	44
1.1. Что такое BPF и eBPF?.....	44
1.2. Что такое трассировка, прослушивание, выборка, профилирование и наблюдаемость?	45
1.3. Что такое BCC, bpftrace и IO Visor?.....	46
1.4. Первый взгляд на BCC: быстрый анализ	48
1.5. Область видимости механизма трассировки BPF.....	50
1.6. Динамическая инструментация: kprobes и uprobes.....	52
1.7. Статическая инструментация: точки трассировки и USDT	54

1.8. Первый взгляд на bpftrace: трассировка open()	55
1.9. Назад к ВСС: трассировка open()	57
1.10. Итоги	59

Глава 2. Основы технологии..... 60

2.1. BPF в иллюстрациях.....	61
2.2. BPF.....	61
2.3. Расширенный BPF (eBPF).....	63
2.3.1. Зачем инструментам оценки производительности нужен BPF.....	65
2.3.2. BPF и модули ядра.....	68
2.3.3. Разработка программ для BPF.....	68
2.3.4. Обзор инструкций BPF: bpftool.....	69
2.3.5. Обзор инструкций BPF: bpftrace.....	76
2.3.6. BPF API	77
2.3.7. Управление конкурентностью в BPF	82
2.3.8. Интерфейс sysfs для BPF	83
2.3.9. BPF Type Format (BTF)	83
2.3.10. BPF CO-RE	84
2.3.11. Ограничения BPF	85
2.3.12. Дополнительные источники о BPF	86
2.4. Обход трассировки стека.....	86
2.4.1. Стек на основе указателя на список фреймов.....	86
2.4.2. Использование отладочной информации	88
2.4.3. Last Branch Record (LBR).....	88
2.4.4. ORC	88
2.4.5. Символы	89
2.4.6. Для дополнительного чтения	89
2.5. Флейм-графики.....	89
2.5.1. Трассировка стека.....	90
2.5.2. Профилирование трассировки стека.....	90
2.5.3. Флейм-график.....	91
2.5.4. Особенности флейм-графика.....	92

2.5.5. Разновидности	94
2.6. Источники событий	94
2.7. kprobes.....	94
2.7.1. Как работает kprobes.....	95
2.7.2. Интерфейсы kprobes.....	97
2.7.3. BPF и kprobes.....	98
2.7.4. Дополнительные источники информации о kprobes	99
2.8. uprobes.....	99
2.8.1. Как работает uprobes.....	100
2.8.2. Интерфейсы uprobes.....	101
2.8.3. BPF и uprobes.....	101
2.8.4. Оверхед uprobes и будущие улучшения	102
2.8.5. Дополнительные источники информации о uprobes	103
2.9. Точки трассировки.....	103
2.9.1. Инструментация точек трассировки	104
2.9.2. Как работают точки трассировки.....	105
2.9.3. Интерфейсы точек трассировки.....	106
2.9.4. BPF и точки трассировки	106
2.9.5. Неструктурированные точки трассировки в BPF	107
2.9.6. Дополнительные источники информации	108
2.10. USDT.....	108
2.10.1. Добавление поддержки USDT.....	109
2.10.2. Как работает USDT	110
2.10.3. BPF и USDT.....	111
2.10.4. Дополнительные источники информации о USDT	112
2.11. Динамический USDT	112
2.12. PMC.....	113
2.12.1. Режимы PMC.....	114
2.12.2. PEBS	115
2.12.3. Облачные вычисления	115
2.13. perf_events	115
2.14. Итоги.....	116

Глава 3. Анализ производительности	117
3.1. Обзор	117
3.1.1. Цели	118
3.1.2. Действия	119
3.1.3. Многочисленные проблемы производительности.....	119
3.2. Методологии оценки производительности.....	120
3.2.1. Определение характера рабочей нагрузки.....	120
3.2.2. Анализ с последовательным увеличением детализации	121
3.2.3. Метод USE	122
3.2.4. Чек-листы.....	123
3.3. Чек-лист инструментов Linux для анализа за 60 секунд	124
3.3.1. uptime	124
3.3.2. dmesg tail	125
3.3.3. vmstat 1	125
3.3.4. mpstat -P ALL 1.....	126
3.3.5. pidstat 1.....	127
3.3.6. iostat -xz 1.....	127
3.3.7. free -m	128
3.3.8. sar -n DEV 1	129
3.3.9. sar -n TCP,ETCP 1.....	129
3.3.10. top.....	130
3.4. Чек-лист инструментов BCC.....	130
3.4.1. execsnoop.....	131
3.4.2. opensnoop	131
3.4.3. ext4slower.....	132
3.4.4. biolatency.....	132
3.4.5. biosnoop	133
3.4.6. cachestat.....	133
3.4.7. tcpconnect	134
3.4.8. tcpaccept	134
3.4.9. tcpretrans.....	134
3.4.10. runqlat.....	135

3.4.11. profile.....	135
3.5. Итоги	136
Глава 4. ВСС.....	137
4.1. Компоненты ВСС.....	138
4.2. Возможности ВСС.....	139
4.2.1. Возможности в пространстве ядра	139
4.2.2. Возможности в пространстве пользователя.....	140
4.3. Установка ВСС.....	140
4.3.1. Требования к конфигурации ядра	140
4.3.2. Ubuntu	141
4.3.3. RHEL.....	141
4.3.4. Другие дистрибутивы	142
4.4. Инструменты ВСС	142
4.4.1. Инструменты, рассматриваемые в книге	142
4.4.2. Характеристики инструментов.....	143
4.4.3. Специализированные инструменты	144
4.4.4. Многоцелевые инструменты	146
4.5. funccount.....	147
4.5.1. Примеры funccount	147
4.5.2. Синтаксис funccount	149
4.5.3. Однострочные сценарии funccount	150
4.5.4. Порядок использования funccount.....	150
4.6. stackcount	151
4.6.1. Пример stackcount.....	152
4.6.2. Создание флейм-графиков с помощью stackcount	153
4.6.3. Искаженные трассировки.....	154
4.6.4. Синтаксис stackcount	154
4.6.5. Однострочные сценарии stackcount.....	155
4.6.6. Порядок использования stackcount	155
4.7. trace	156
4.7.1. Пример trace	157
4.7.2. Синтаксис trace	157

4.7.3. Однострочные сценарии trace	159
4.7.4. trace и структуры	159
4.7.5. Использование trace для отладки утечек дескрипторов файлов.....	160
4.7.6. Порядок использования trace	162
4.8. argdist.....	163
4.8.1. Синтаксис argdist.....	164
4.8.2. Однострочные сценарии argdist	165
4.8.3. Порядок использования argdist.....	166
4.9. Документация инструментов.....	167
4.9.1. Страница справочного руководства: opensnoop	167
4.9.2. Файл с примерами: opensnoop	170
4.10. Разработка инструментов ВСС.....	172
4.11. Внутреннее устройство ВСС	172
4.12. Отладка ВСС.....	174
4.12.1. Отладка с помощью printf().....	174
4.12.2. Отладочный вывод ВСС	177
4.12.3. Флаги отладки ВСС	178
4.12.4. bpfflist	179
4.12.5. bpftool	179
4.12.6. dmesg	179
4.12.7. Сброс событий.....	180
4.13. Итоги.....	181
Глава 5. bpfftrace	182
5.1. Компоненты bpfftrace	183
5.2. Возможности bpfftrace.....	184
5.2.1. Источники событий bpfftrace	184
5.2.2. Действия bpfftrace.....	184
5.2.3. Общие возможности bpfftrace	185
5.2.4. Сравнение bpfftrace с другими инструментами мониторинга	185
5.3. Установка bpfftrace.....	187
5.3.1. Требования к конфигурации ядра	187
5.3.2. Ubuntu	187

5.3.3. Fedora.....	188
5.3.4. Действия после сборки	188
5.3.5. Другие дистрибутивы	188
5.4. Инструменты bpftrace.....	188
5.4.1. Инструменты, рассматриваемые в книге	188
5.4.2. Характеристики инструментов.....	190
5.4.3. Использование инструментов.....	190
5.5. Однострочные сценарии bpftrace	191
5.6. Документация bpftrace	192
5.7. Программирование на bpftrace.....	192
5.7.1. Порядок использования.....	193
5.7.2. Структура программы.....	193
5.7.3. Комментарии.....	194
5.7.4. Формат определения зондов	194
5.7.5. Подстановочные символы в определениях зондов.....	195
5.7.6. Фильтры.....	196
5.7.7. Действия.....	196
5.7.8. Hello, World!	196
5.7.9. Функции	197
5.7.10. Переменные	197
5.7.11. Функции карт	198
5.7.12. Определение продолжительности выполнения <code>vfs_read()</code>	199
5.8. Порядок использования bpftrace.....	201
5.9. Типы зондов в bpftrace	202
5.9.1. <code>tracpoint</code>	203
5.9.2. <code>usdt</code>	204
5.9.3. <code>kprobe</code> и <code>kretprobe</code>	205
5.9.4. <code>uprobe</code> и <code>uretprobe</code>	206
5.9.5. <code>software</code> и <code>hardware</code>	206
5.9.6. <code>profile</code> и <code>interval</code>	208
5.10. Управление потоком выполнения в bpftrace.....	209
5.10.1. Фильтры	209
5.10.2. Тернарные операторы	209

5.10.3. Инструкция if.....	210
5.10.4. Развернутые циклы	210
5.11. Операторы bpftrace.....	210
5.12. Переменные bpftrace.....	211
5.12.1. Встроенные переменные.....	211
5.12.2. Встроенные переменные pid, comm и uid.....	212
5.12.3. Встроенные переменные kstack и ustack.....	212
5.12.4. Встроенные переменные: позиционные параметры.....	214
5.12.5. Временные переменные	215
5.12.6. Карты.....	215
5.13. Функции bpftrace.....	216
5.13.1. printf()	217
5.13.2. join().....	218
5.13.3. str().....	219
5.13.4. kstack() и ustack()	219
5.13.5. ksym() и usym().....	220
5.13.6. kaddr() и uaddr().....	221
5.13.7. system()	221
5.13.8. exit().....	222
5.14. Функции-карты в bpftrace	222
5.14.1. count()	223
5.14.2. sum(), avg(), min() и max().....	224
5.14.3. hist()	224
5.14.4. lhist().....	225
5.14.5. delete().....	226
5.14.6. clear() и zero().....	226
5.14.7. print().....	227
5.15. Направления развития bpftrace в будущем	228
5.15.1. Режимы явной адресации.....	228
5.15.2. Другие расширения	229
5.15.3. ply.....	230
5.16. Внутреннее устройство bpftrace	230

5.17. Отладка bpftrace	231
5.17.1. Отладка с помощью printf().....	232
5.17.2. Режим отладки	232
5.17.3. Режим подробного вывода.....	234
5.18. Итоги	235
Глава 6. Процессоры	236
6.1. Основы.....	237
6.1.1. Основы работы процессоров	237
6.1.2. Возможности BPF	240
6.1.3. Стратегия.....	242
6.2. Традиционные инструменты.....	243
6.2.1. Статистика ядра	244
6.2.2. Статистика оборудования	247
6.2.3. Выборка характеристик работы оборудования.....	249
6.2.4. Выборка по времени	250
6.2.5. Получение статистик и трассировка событий.....	254
6.3. Инструменты BPF	257
6.3.1. execsnoop.....	258
6.3.2. exitsnoop.....	261
6.3.3. runqlat	262
6.3.4. runqlen.....	266
6.3.5. runqslower	269
6.3.6. cpudist	270
6.3.7. cpufreq	272
6.3.8. profile	274
6.3.9. offcputime	278
6.3.10. syscount	283
6.3.11. argdist и trace.....	285
6.3.12. funccount.....	288
6.3.13. softirqs.....	290
6.3.14. hardirqs	291
6.3.15. smpcalls.....	292

6.3.16. llcstat	296
6.3.17. Другие инструменты	296
6.4. Однострочные сценарии для BPF	297
6.4.1. BCC	297
6.4.2. bpftrace	298
6.5. Дополнительные упражнения	299
6.6. Итоги	300
Глава 7. Память	301
7.1. Основы	302
7.1.1. Основы управления памятью	302
7.1.2. Возможности BPF	307
7.1.3. Стратегия	309
7.2. Традиционные инструменты	310
7.2.1. Журнал ядра	311
7.2.2. Статистики ядра	312
7.2.3. Аппаратные статистики и выборки	315
7.3. Инструменты BPF	317
7.3.1. oomkill	318
7.3.2. memleak	319
7.3.3. mmapnoop	321
7.3.4. brkstack	323
7.3.5. shmsnoop	324
7.3.6. faults	325
7.3.7. ffaults	327
7.3.8. vmscan	328
7.3.9. drsnoop	331
7.3.10. swapin	332
7.3.11. hfaults	333
7.3.12. Другие инструменты	334
7.4. Однострочные сценарии для BPF	334
7.4.1. BCC	334
7.4.2. bpftrace	335

7.5. Дополнительные упражнения	335
7.6. Итоги	336
Глава 8. Файловые системы.....	337
8.1. Основы	338
8.1.1. Основы файловых систем.....	338
8.1.2. Возможности BPF	341
8.1.3. Стратегия.....	342
8.2. Традиционные инструменты.....	344
8.2.1. df	344
8.2.2. mount.....	344
8.2.3. strace	345
8.2.4. perf.....	346
8.2.5. fatrace	348
8.3. Инструменты BPF	349
8.3.1. opensnoop.....	351
8.3.2. statsnoop.....	353
8.3.3. syncsnoop.....	355
8.3.4. mmapfiles	357
8.3.5. scread	358
8.3.6. fmapfault	360
8.3.7. filelife	361
8.3.8. vfststat.....	362
8.3.9. vfscount	364
8.3.10. vfssize.....	365
8.3.11. fsrwstat.....	367
8.3.12. fileslower.....	369
8.3.13. filetop.....	371
8.3.14. writesync	374
8.3.15. filetype.....	375
8.3.16. cachestat	378
8.3.17. writeback	380
8.3.18. dctestat	382

8.3.19. dcsnoop	384
8.3.20. mountsnoop	386
8.3.21. xfsslower	386
8.3.22. xfsdist	388
8.3.23. ext4dist.....	390
8.3.24. icstat.....	393
8.3.25. bufgrow	395
8.3.26. readahead.....	396
8.3.27. Другие инструменты	397
8.4. Однострочные сценарии для BPF.....	398
8.4.1. BCC.....	398
8.4.2. bpftrace.....	399
8.4.3. Примеры использования однострочных сценариев BPF	401
8.5. Дополнительные упражнения	404
8.6. Итоги	405
Глава 9. Дисковый ввод/вывод.....	406
9.1. Основы.....	407
9.1.1. Основы дисков.....	407
9.1.2. Возможности BPF	410
9.1.3. Стратегия.....	412
9.2. Традиционные инструменты.....	412
9.2.1. iostat.....	413
9.2.2. perf.....	415
9.2.3. blktrace.....	416
9.2.4. Логирование SCSI.....	417
9.3. Инструменты BPF	418
9.3.1. biolatency.....	419
9.3.2. biosnoop	425
9.3.3. biotop	429
9.3.4. bitesize	430
9.3.5. seeksize	432
9.3.6. biopattern	434

9.3.7. biostacks.....	435
9.3.8. bioerr.....	438
9.3.9. mdflush.....	441
9.3.10. iosched.....	442
9.3.11. scsilatency.....	444
9.3.12. scsiresult.....	446
9.3.13. nvmlatency.....	448
9.4. Однострочные сценарии для BPF.....	450
9.4.1. BCC.....	451
9.4.2. bpftrace.....	451
9.4.3. Примеры использования однострочных сценариев BPF.....	452
9.5. Дополнительные упражнения.....	453
9.6. Итоги.....	454
Глава 10. Сети.....	455
10.1. Основы.....	456
10.1.1. Основы организации сетей.....	456
10.1.2. Возможности BPF.....	464
10.1.3. Стратегия.....	466
10.1.4. Типичные ошибки трассировки.....	467
10.2. Традиционные инструменты.....	468
10.2.1. ss.....	469
10.2.2. ip.....	470
10.2.3. nstat.....	471
10.2.4. netstat.....	472
10.2.5. sar.....	474
10.2.6. nicstat.....	475
10.2.7. ethtool.....	475
10.2.8. tcpdump.....	477
10.2.9. /proc.....	478
10.3. Инструменты BPF.....	479
10.3.1. sockstat.....	481
10.3.2. sofamilly.....	483

10.3.3. soprocol.....	486
10.3.4. soconnect.....	488
10.3.5. soaccept.....	490
10.3.6. socketio	493
10.3.7. socksize.....	495
10.3.8. sormem	497
10.3.9. soconnlst.....	500
10.3.10. so1stbyte.....	503
10.3.11. tcpconnect.....	505
10.3.12. tcpaccept	507
10.3.13. tcplife.....	511
10.3.14. tcptop	515
10.3.15. tcpsnoop	516
10.3.16. tcpretrans	517
10.3.17. tcpsynbl	520
10.3.18. tcpwin.....	522
10.3.19. tcpnagle.....	524
10.3.20. udpconnect	526
10.3.21. gethostlatency.....	527
10.3.22. ipecn	529
10.3.23. superping.....	530
10.3.24. qdisc-fq.....	533
10.3.25. qdisc-cbq, qdisc-cbs, qdisc-codel, qdisc-fq_codel, qdisc-red и qdisc-tbf	535
10.3.26. netsize.....	536
10.3.27. nettxlat.....	539
10.3.28. skbdrop	541
10.3.29. skblife	543
10.3.30. ieee80211scan.....	545
10.3.31. Другие инструменты.....	547
10.4. Однострочные сценарии для BPF	547
10.4.1. BCC	548

10.4.2. bpftrace	549
10.4.3. Примеры использования однострочных сценариев BPF	551
10.5. Дополнительные упражнения	554
10.6. Итоги	555
Глава 11. Безопасность	556
11.1. Основы	556
11.1.1. Возможности BPF	557
11.1.2. Непривилегированные пользователи BPF	561
11.1.3. Настройка безопасности BPF	562
11.1.4. Стратегия	563
11.2. Инструменты BPF	563
11.2.1. execsnoop	565
11.2.2. elfsnoop	565
11.2.3. modsnoop	567
11.2.4. bashreadline	568
11.2.5. shellsnoop	569
11.2.6. ttysnoop	570
11.2.7. opensnoop	572
11.2.8. eperm	573
11.2.9. tcpconnect и tcpaccept	574
11.2.10. tcpreset	575
11.2.11. capable	576
11.2.12. setuids	579
11.3. Однострочные сценарии для BPF	581
11.3.1. BCC	581
11.3.2. bpftrace	582
11.3.3. Примеры использования однострочных сценариев BPF	582
11.4. Итоги	583
Глава 12. Языки	584
12.1. Основы	584
12.1.1. Компилируемые языки	585

12.1.2. Языки с динамической компиляцией	586
12.1.3. Интерпретируемые языки.....	588
12.1.4. Возможности BPF.....	589
12.1.5. Стратегия	590
12.1.6. Инструменты BPF	591
12.2. C	591
12.2.1. Символы функций на C	592
12.2.2. Трассировка стека программного кода на C.....	596
12.2.3. Трассировка функций на C	597
12.2.4. Трассировка смещений в функциях на C.....	598
12.2.5. Трассировка программного кода на C с помощью зондов USDT.....	598
12.2.6. Трассировка программного кода на C с помощью однострочных сценариев.....	599
12.3. Java	601
12.3.1. Трассировка libjvm	602
12.3.2. jnistacks	603
12.3.3. Имена потоков выполнения в Java.....	605
12.3.4. Символы методов Java.....	607
12.3.5. Приемы трассировки стека Java.....	609
12.3.6. Зонды USDT в Java	613
12.3.7. profile.....	619
12.3.8. offcputime.....	623
12.3.9. stackcount	629
12.3.10. javastat.....	632
12.3.11. javathreads.....	633
12.3.12. javacalls.....	635
12.3.13. javaflow.....	636
12.3.14. javagc.....	637
12.3.15. javaobjnew	638
12.3.16. Однострочные сценарии для трассировки кода на Java.....	639
12.4. Командная оболочка bash	640
12.4.1. Подсчет вызовов функций	642

12.4.2. Трассировка аргументов функции (bashfunc.bt)	643
12.4.3. Задержки в функциях (bashfunclat.bt)	645
12.4.4. /bin/bash.....	647
12.4.5. Зонды USDT в /bin/bash.....	650
12.4.6. Однострочные сценарии для трассировки bash	651
12.5. Другие языки.....	651
12.5.1. JavaScript (Node.js)	652
12.5.2. C++	654
12.5.3. Golang	654
12.6. Итоги.....	657
Глава 13. Приложения.....	658
13.1. Основы.....	659
13.1.1. Основы приложений.....	659
13.1.2. Пример приложения: сервер MySQL.....	660
13.1.3. Возможности BPF.....	661
13.1.4. Стратегия	662
13.2. Инструменты BPF.....	663
13.2.1. execsnoop.....	665
13.2.2. threadsnoop	666
13.2.3. profile.....	668
13.2.4. threaded	671
13.2.5. offcputime.....	672
13.2.6. offcputhist.....	676
13.2.7. syscount	679
13.2.8. ioprofile.....	681
13.2.9. Указатели фреймов в libc.....	682
13.2.10. mysql_qslower.....	683
13.2.11. mysql_clat	686
13.2.12. signals.....	689
13.2.13. killsnoop	691
13.2.14. pmlock и pmheld	692

12.4.2. Трассировка аргументов функции (bashfunc.bt).....	643
12.4.3. Задержки в функциях (bashfunclat.bt).....	645
12.4.4. /bin/bash.....	647
12.4.5. Зонды USDT в /bin/bash.....	650
12.4.6. Однострочные сценарии для трассировки bash	651
12.5. Другие языки.....	651
12.5.1. JavaScript (Node.js)	652
12.5.2. C++	654
12.5.3. Golang.....	654
12.6. Итоги.....	657
Глава 13. Приложения.....	658
13.1. Основы.....	659
13.1.1. Основы приложений.....	659
13.1.2. Пример приложения: сервер MySQL.....	660
13.1.3. Возможности BPF.....	661
13.1.4. Стратегия	662
13.2. Инструменты BPF.....	663
13.2.1. execsnoop.....	665
13.2.2. threadsnoop	666
13.2.3. profile.....	668
13.2.4. threaded	671
13.2.5. offcputime.....	672
13.2.6. offcpuhist.....	676
13.2.7. syscount	679
13.2.8. ioprofile	681
13.2.9. Указатели фреймов в libc.....	682
13.2.10. mysqld_qlower.....	683
13.2.11. mysqld_clat	686
13.2.12. signals.....	689
13.2.13. killsnoop	691
13.2.14. pmlock и pmheld	692

13.2.15. napttime	697
13.2.16. Другие инструменты.....	698
13.3. Однострочные сценарии для BPF	698
13.3.1. BCC	698
13.3.2. bpftrace	699
13.4. Примеры использования однострочных сценариев BPF	700
13.4.1. Подсчет количества вызовов в секунду функций из библиотеки libpthread для доступа к условным переменным	700
13.5. Итоги.....	701
Глава 14. Ядро.....	702
14.1. Основы.....	703
14.1.1. Основы ядра.....	703
14.1.2. Возможности BPF.....	705
14.2. Стратегия.....	707
14.3. Традиционные инструменты.....	708
14.3.1. Ftrace.....	708
14.3.2. perf sched.....	711
14.3.3. slabtop	712
14.3.4. Другие инструменты	712
14.4. Инструменты BPF.....	713
14.4.1. loads.....	714
14.4.2. offcputime.....	715
14.4.3. wakeuptime.....	717
14.4.4. offwaketime.....	719
14.4.5. mlock и mheld	720
14.4.6. Спин-блокировки.....	724
14.4.7. kmem	725
14.4.8. kpages.....	726
14.4.9. memleak	727
14.4.10. slabratetop	728
14.4.11. numamove.....	729
14.4.12. workq.....	730

14.4.13. Тасклеты	732
14.4.14. Другие инструменты	732
14.5. Однострочные сценарии для BPF	733
14.5.1. BCC	733
14.5.2. bpftrace	734
14.6. Примеры использования однострочных сценариев BPF	735
14.6.1. Подсчет обращений к системным вызовам по именам функций системных вызовов	735
14.6.2. Подсчет запусков hrtimer функциями ядра	736
14.7. Сложности	736
14.8. Итоги	737
Глава 15. Контейнеры	738
15.1. Основы	738
15.1.1. Возможности BPF	740
15.1.2. Сложности	741
15.1.3. Стратегия	743
15.2. Традиционные инструменты	744
15.2.1. Анализ на уровне хоста	744
15.2.2. Анализ на уровне контейнера	745
15.2.3. systemd-cgtop	745
15.2.4. kubectl top	746
15.2.5. docker stats	746
15.2.6. /sys/fs/cgroups	747
15.2.7. perf	747
15.3. Инструменты BPF	748
15.3.1. runqlat	748
15.3.2. pidnss	749
15.3.3. blkthrot	751
15.3.4. overlayfs	752
15.4. Однострочные сценарии для BPF	754
15.5. Дополнительные упражнения	755
15.6. Итоги	755

Глава 16. Гипервизоры	756
16.1. Основы	756
16.1.1. Возможности BPF	758
16.1.2. Возможные стратегии	759
16.2. Традиционные инструменты	760
16.3. Инструменты BPF для анализа на уровне гостевой ОС	760
16.3.1. Гипервызовы Xen	761
16.3.2. xenhyper	765
16.3.3. Обратные вызовы Xen	766
16.3.4. cpustolen	768
16.3.5. Трассировка выходов в HVM	769
16.4. Инструменты BPF анализа на уровне хоста	769
16.4.1. kvmexits	770
16.4.2. Возможные направления развития в будущем	773
16.5. Итоги	774
Глава 17. Другие инструменты BPF для анализа производительности	775
17.1. Vector и Performance Co-Pilot (PCP)	776
17.1.1. Визуализация	777
17.1.2. Визуализация: тепловые карты	777
17.1.3. Визуализация: табличное представление данных	779
17.1.4. Метрики BCC	780
17.1.5. Внутреннее устройство	781
17.1.6. Установка PCP и Vector	782
17.1.7. Подключение и просмотр данных	782
17.1.8. Настройка BCC PMDA	784
17.1.9. Возможные направления развития в будущем	785
17.1.10. Для дополнительного чтения	785
17.2. Grafana и Performance Co-Pilot (PCP)	785
17.2.1. Установка и настройка	786
17.2.2. Подключение и просмотр данных	786
17.2.3. Возможные направления развития в будущем	788
17.2.4. Для дополнительного чтения	788

17.3. Экспортер Cloudflare eBPF Prometheus (с Grafana).....	788
17.3.1. Сборка и запуск экспортера ebf 789	789
17.3.2. Настройка Prometheus для мониторинга экземпляра ebpf_exporter.....	789
17.3.3. Создание запроса в Grafana	790
17.3.4. Для дополнительного чтения.....	790
17.4. kubectl-trace	790
17.4.1. Трассировка узлов.....	791
17.4.2. Трассировка подов и контейнеров.....	791
17.4.3. Для дополнительного чтения.....	793
17.5. Другие инструменты.....	793
17.6. Итоги.....	794
Глава 18. Советы, рекомендации и типичные проблемы.....	795
18.1. Типичная частота событий и оверхед	795
18.1.1. Частота.....	796
18.1.2. Выполняемые действия.....	798
18.1.3. Проверяй себя.....	800
18.2. Выборка с частотой 49 или 99 Гц.....	800
18.3. Желтые свиньи и серые крысы	801
18.4. Пишите целевое ПО	802
18.5. Изучайте системные вызовы	803
18.6. Не усложняйте.....	804
18.7. Отсутствие событий.....	805
18.8. Отсутствие трассировок стека	806
18.8.1. Как исправить проблему отсутствия трассировок стека.....	807
18.9. Отсутствие символов (имен функций) в выводе	808
18.9.1. Как исправить проблему отсутствия символов: среда выполнения с JIT (Java, Node.js, ...)	808
18.9.2. Как исправить проблему отсутствия символов: двоичные файлы ELF (C, C++, ...)	809
18.10. Отсутствие функций в трассировке.....	809
18.11. Циклы обратной связи.....	810
18.12. Сброс событий.....	810

Приложение А. Однострочные сценарии для bpftrace	811
Приложение В. Шпаргалка по bpftrace	816
Приложение С. Разработка инструментов BCC	819
Ресурсы.....	819
Пять советов	819
Примеры инструментов.....	820
Инструмент 1: hello_world.py	820
Инструмент 2: sleepsnoop.py	821
Инструмент 3: bitehist.py	823
Инструмент 4: biolatency.....	827
Дополнительная информация.....	832
Приложение D. С BPF	833
Почему на С?.....	833
Пять советов	835
Программы на С	836
ВНИМАНИЕ: изменения в API	837
Компиляция	837
Инструмент 1: Hello, World!.....	838
Инструмент 2: bigreads.....	841
Инструмент 3: bitehist.....	846
perf C.....	849
Инструмент 1: bigreads.....	849
Дополнительная информация.....	851
Приложение Е. Инструкции BPF	852
Вспомогательные макросы.....	852
Инструкции	855
Кодирование.....	856
Ссылки.....	857
Глоссарий	858
Библиография	868