

УДК 004.42
ББК 32.973
К31

Качин К., Гуерру Р., Родригес Л.

К31 Введение в надежное и безопасное распределенное программирование / пер. с англ. А. Н. Киселева. – М.: ДМК Пресс, 2016. – 512 с.: ил.

ISBN 978-5-97060-369-7

В современных вычислениях программы нередко объединяют несколько процессов. Основная проблема, возникающая при создании таких распределенных программ, состоит в том, чтобы заставить все процессы вместе работать над решением общей задачи, даже в случае отказов некоторых из них.

Данная книга содержит введение в абстракции распределенного программирования и знакомит с фундаментальными алгоритмами и их реализациями в нескольких распределенных окружениях. Перед читателем будут раскрыты важные проблемы распределенных вычислений и основные алгоритмические приемы их решения. На подробных примерах читатель сможет понять, как с помощью этих приемов конструировать распределенные приложения. Обсуждение каждой темы завершается множеством упражнений и их решений.

Издание предназначено для студентов высших учебных заведений, аспирантов, а также программистов, занимающихся разработкой высоконадежных распределенных приложений.

УДК 004.42
ББК 32.973

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Springer-Verlag Berlin Heidelberg. RUSSIAN language edition published by DMK PUBLISHERS. Copyright © 2016.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-3-642-15259-7 (анг.)

Copyright © Springer-Verlag Berlin Heidelberg
2011, 2006

ISBN 978-5-97060-369-7 (рус.)

© Оформление, перевод, ДМК Пресс, 2016

Содержание

Предисловие	13
Глава 1. Введение.....	19
1.1. Обоснование.....	19
1.2. Абстракции распределенного программирования	21
1.2.1. Естественно распределенные системы.....	22
1.2.2. Искусственно распределенные системы	25
1.3. Сквозной аргумент	26
1.4. Программные компоненты	27
1.4.1. Композиционная модель.....	27
1.4.2. Программный интерфейс	30
1.4.3. Модули	32
1.5. Классы алгоритмов.....	36
1.6. Практикум	37
1.6.1. Модуль Print	37
1.6.2. Модуль BoundedPrint	39
1.6.3. Объединение модулей.....	43
1.6.4. Эксперимент.....	44
1.7. Примечания к главе.....	45
Глава 2. Базовые абстракции.....	46
2.1. Распределенные вычисления	47
2.1.1. Процессы и сообщения.....	47
2.1.2. Автоматы и шаги.....	47
2.1.3. Безопасность и живучесть.....	49
2.2. Абстракции процессов.....	51
2.2.1. Отказы процессов.....	51
2.2.2. Аварии	51
2.2.3. Пропуск данных.....	53
2.2.4. Аварии с восстановлением	53
2.2.5. Перехват сообщений.....	56
2.2.6. Произвольное поведение	56
2.3. Криптографические абстракции.....	58
2.3.1. Хэш-функции.....	58
2.3.2. Коды аутентификации сообщений.....	58
2.3.3. Цифровые подписи	59
2.4. Абстракции связи.....	60
2.4.1. Отказы связи	61
2.4.2. Связи с приемлемыми потерями	63

2.4.3. Связи с повторами.....	63
2.4.4. Идеальные связи.....	65
2.4.5. Идеальные связи с журналированием	67
2.4.6. Идеальная связь с аутентификацией.....	70
2.4.7. Об абстракциях связей	72
2.5. Предположения о синхронности	74
2.5.1. Асинхронные системы	74
2.5.2. Синхронные системы.....	75
2.5.3. Частичная синхронность	77
2.6. Абстракция времени.....	78
2.6.1. Обнаружение отказов	78
2.6.2. Идеальное обнаружение отказа.....	79
2.6.3. Выбор лидера	82
2.6.4. Эвентуально идеальный датчик отказов.....	84
2.6.5. Эвентуальный выбор лидера.....	87
2.6.6. Византийский выбор лидера	91
2.7. Модели распределенных систем.....	94
2.7.1. Комбинации абстракций	95
2.7.2. Подготовка.....	96
2.7.3. Кворумы.....	97
2.7.4. Измерение стоимости алгоритмов	98
2.8. Упражнения.....	99
2.9. Решения.....	100
2.10. Практикум	103
2.10.1. Передаваемое событие.....	103
2.10.2. Сообщение	104
2.10.3. Связь «точка-точка» с приемлемыми потерями	105
2.10.4. Идеальная связь «точка-точка»	105
2.10.5. Идеальный датчик отказов	106
2.11. Примечания к главе.....	108

Глава 3. Надежная рассылка..... 111

3.1. Обоснование.....	111
3.1.1. Клиент-серверные вычисления	111
3.1.2. Системы со множеством участников.....	112
3.2. Негарантированная рассылка.....	113
3.2.1. Спецификация.....	113
3.2.2. Алгоритм простой рассылки.....	114
3.3. Обычная надежная рассылка	115
3.3.1. Спецификация.....	115
3.3.2. Алгоритм для модели с остановкой после отказов: ленивая надежная рассылка.....	116
3.3.3. Алгоритм для модели без уведомления об отказах: жадная надежная рассылка.....	118

3.4. Унифицированная надежная рассылка	120
3.4.1. Спецификация.....	120
3.4.2. Алгоритм для модели с остановкой после отказов: унифицированная надежная рассылка с подтверждением всеми	121
3.4.3. Алгоритм для модели без уведомления об отказах: унифицированная надежная рассылка с подтверждением большинством	124
3.5. Рассылка с повторами.....	125
3.5.1. Спецификация.....	125
3.5.2. Алгоритм для модели с восстановлением после отказов: простая рассылка с повторами	126
3.6. Негарантированная рассылка с журналированием.....	127
3.6.1. Обзор.....	127
3.6.2. Спецификация.....	128
3.6.3. Алгоритм для модели с восстановлением после отказов: простая рассылка с журналированием	129
3.7. Унифицированная надежная рассылка с журналированием.....	130
3.7.1. Спецификация.....	130
3.7.2. Алгоритм для модели с восстановлением после отказов: Унифицированная надежная рассылка с подтверждением большинством и журналированием.....	131
3.8. Вероятностная рассылка.....	132
3.8.1. Масштабируемость надежной рассылки.....	133
3.8.2. Распространение эпидемии	134
3.8.3. Спецификация.....	134
3.8.4. Рандомизированный алгоритм: жадная вероятностная рассылка.....	135
3.8.5. Рандомизированный алгоритм: ленивая вероятностная рассылка	138
3.9. Рассылка в порядке FIFO и причинная рассылка.....	142
3.9.1. Обзор.....	142
3.9.2. Спецификация порядка FIFO	143
3.9.3. Алгоритм для модели без уведомления об отказах: рассылка с использованием последовательных номеров	143
3.9.4. Спецификация причинно-следственного порядка.....	144
3.9.5. Алгоритм для модели без уведомления об отказах: рассылка в причинно-следственном порядке без ожидания	146
3.9.6. Алгоритм для модели с остановкой после отказов: сборка мусора причинно-следственного прошлого.....	148
3.9.7. Алгоритм для модели без уведомления об отказах: причинно-следственная рассылка с ожиданием	150
3.10. Византийская согласованная рассылка	152
3.10.1. Обоснование	153
3.10.2. Спецификация	154
3.10.3. Алгоритм для модели с произвольным поведением: эхо-рассылка с аутентификацией	155

3.10.4. Алгоритм для модели с произвольным поведением: эхо-рассылка с подписанными сообщениями	158
3.11. Византийская надежная рассылка	160
3.11.1. Спецификация	160
3.11.2. Алгоритм для модели с произвольным поведением: двойная эхо-рассылка с аутентификацией	161
3.12. Византийские широкоэмитательные каналы	164
3.12.1. Спецификации	164
3.12.2. Алгоритм для модели с произвольным поведением: Византийский согласованный канал	166
3.12.3. Алгоритм для модели с произвольным поведением: Византийский надежный канал	167
3.13. Упражнения	167
3.14. Решения	169
3.15. Практикум	178
3.15.1. Простая рассылка	178
3.15.2. Ленивая надежная рассылка	181
3.15.3. Унифицированная надежная рассылка с подтверждением всеми	184
3.15.4. Унифицированная надежная рассылка с подтверждением большинством	187
3.15.5. Вероятностная надежная рассылка	188
3.15.6. Рассылка в причинно-следственном порядке без ожидания	191
3.15.7. Рассылка в причинно-следственном порядке без ожидания и с уборкой мусора	197
3.15.8. Рассылка в причинно-следственном порядке с ожиданием	204
3.16. Примечания к главе	208

Глава 4. Разделяемая память210

4.1. Введение	211
4.1.1. Разделяемое хранилище в распределенной системе	211
4.1.2. Регистр	211
4.1.3. Завершенность и предшествование	214
4.2. Обычный регистр (1, N)	216
4.2.1. Спецификация	216
4.2.2. Алгоритм для модели с остановкой после отказов: обычный регистр «Чтение в одном – запись во всех»	217
4.2.3. Алгоритм для модели без уведомления об отказах: обычный регистр с голосованием большинством	220
4.3. Атомарный регистр (1, N)	223
4.3.1. Спецификация	223
4.3.2. Преобразование обычного регистра (1, N) в атомарный регистр (1, N)	226

4.3.3. Алгоритм атомарного регистра для модели с остановкой после отказов: чтение с записью – запись во всех	231
4.3.4. Алгоритм атомарного регистра для модели без уведомления об отказах: чтение с записью – запись в большинстве	233
4.4. Атомарный регистр (N, N)	235
4.4.1. Несколько пишущих процессов	235
4.4.2. Спецификация	236
4.4.3. Преобразование атомарного регистра $(1, N)$ в атомарный регистр (N, N)	237
4.4.4. Алгоритм для модели с остановкой после отказов: чтение с записью – запись с проверкой во всех	241
4.4.5. Алгоритм для модели без уведомления об отказах: чтение с записью – запись с проверкой в большинстве	244
4.5. Обычный регистр $(1, N)$ с журналированием	247
4.5.1. Предшествование в модели с восстановлением после отказов	247
4.5.2. Спецификация	248
4.5.3. Алгоритм для модели с восстановлением после отказов: голосование большинством с журналированием	250
4.6. Византийский безопасный регистр $(1, N)$	253
4.6.1. Спецификация	254
4.6.2. Алгоритм для модели с произвольным поведением после отказов: Византийский маскирующий кворум	255
4.7. Византийский обычный регистр $(1, N)$	257
4.7.1. Спецификация	258
4.7.2. Алгоритм для модели с произвольным поведением: Аутентификация данных Византийским кворумом	258
4.7.3. Алгоритм для модели с произвольным поведением после отказов: Византийский кворум с двухфазной записью	261
4.8. Византийский атомарный регистр $(1, N)$	268
4.8.1. Спецификация	268
4.8.2. Алгоритм для модели с произвольным поведением после отказов: Византийский кворум со слушателями	269
4.9. Упражнения	273
4.10. Решения	274
4.11. Практикум	280
4.11.1. Обычный регистр $(1, N)$	280
4.11.2. Атомарный регистр $(1, N)$	284
4.11.3. Атомарный регистр (N, N)	288
4.12. Примечания к главе	292

Глава 5. Консенсусы.....296

5.1. Обычный консенсус	297
5.1.1. Спецификация	297

5.1.2. Алгоритм для модели с остановкой после отказов: консенсус заполнением	298
5.1.3. Алгоритм для модели с остановкой после отказов: иерархический консенсус	302
5.2. Унифицированный консенсус	305
5.2.1. Спецификация.....	305
5.2.2. Алгоритм для модели с остановкой после отказов: унифицированный консенсус с заполнением	306
5.2.3. Алгоритм для модели с остановкой после отказов: унифицированный иерархический консенсус	308
5.3. Унифицированный консенсус для модели с уведомлением об отказах.....	311
5.3.1. Обзор.....	311
5.3.2. Смена эпох	312
5.3.3. Консенсус эпохи.....	316
5.3.4. Алгоритм для модели с уведомлением об отказах: консенсус с лидером	322
5.4. Консенсус с журналированием	325
5.4.1. Спецификация.....	325
5.4.2. Смена эпох с журналированием.....	327
5.4.3. Консенсус эпохи с журналированием.....	328
5.4.4. Алгоритм для модели с восстановлением после отказов: консенсус с лидером и журналированием	332
5.5. Рандомизированный консенсус	333
5.5.1. Спецификация.....	334
5.5.2. Общая монета	335
5.5.3. Рандомизированный алгоритм для модели без уведомления об отказах: рандомизированный двоичный консенсус	336
5.5.4. Рандомизированный алгоритм для модели без уведомления об отказах: рандомизированный консенсус с большой областью выбора.....	341
5.6. Византийский консенсус	343
5.6.1. Спецификация.....	343
5.6.2. Византийская смена эпох	346
5.6.3. Византийский консенсус эпохи.....	348
5.6.4. Алгоритм для модели с уведомлением об отказах и произвольным поведением после отказов: Византийский консенсус с лидером	360
5.7. Византийский рандомизированный консенсус.....	362
5.7.1. Спецификация.....	362
5.7.2. Рандомизированный алгоритм для модели с произвольным поведением после отказов: Византийский рандомизированный двоичный консенсус	362
5.8. Упражнения.....	367
5.9. Решения.....	369
5.10. Практикум	380

5.10.1. Протокол обычного консенсуса заполнением.....	380
5.10.2. Протокол обычного иерархического консенсуса.....	385
5.10.3. Унифицированный консенсус заполнением.....	389
5.10.4. Унифицированный иерархический консенсус	392
5.11. Примечания к главе.....	396

Глава 6. Варианты консенсуса400

6.1. Рассылка в едином порядке.....	400
6.1.1. Обзор.....	400
6.1.2. Спецификации.....	402
6.1.3. Алгоритм для модели без уведомления об отказах: рассылка в едином порядке на основе консенсуса	404
6.2. Византийская рассылка в едином порядке	407
6.2.1. Обзор.....	407
6.2.2. Спецификация.....	408
6.2.3. Алгоритм для модели с уведомлением об отказах и произвольным поведением после отказов: Византийская рассылка с ротацией отправителя	408
6.3. Обрывающаяся надежная рассылка	413
6.3.1. Обзор.....	413
6.3.2. Спецификация.....	414
6.3.3. Алгоритм для модели с остановкой после отказов: унифицированная обрывающаяся надежная рассылка на основе консенсуса	414
6.4. Быстрый консенсус.....	417
6.4.1. Обзор.....	417
6.4.2. Спецификация.....	418
6.4.3. Алгоритм для модели без уведомления об отказах: от унифицированного консенсуса к унифицированному быстрому консенсусу	419
6.5. Быстрый Византийский консенсус.....	422
6.5.1. Обзор.....	422
6.5.2. Спецификация.....	422
6.5.3. Алгоритм для модели с произвольным поведением: от Византийского консенсуса к быстрому Византийскому консенсусу	423
6.6. Неблокирующее атомарное подтверждение.....	425
6.6.1. Обзор.....	425
6.6.2. Спецификация.....	427
6.6.3. Алгоритм для модели с остановкой после отказов: неблокирующее атомарное подтверждение на основе консенсуса	427
6.7. Членство в группах.....	430
6.7.1. Обзор.....	430
6.7.2. Спецификация.....	431

6.7.3. Алгоритм для модели с остановкой после отказов: членство в группе на основе консенсуса	432
6.8. Синхронизация взаимодействий с представлением	435
6.8.1. Обзор.....	435
6.8.2. Спецификация.....	436
6.8.3. Алгоритм для модели с остановкой после отказов: синхронизация взаимодействий с представлением на основе TRB.....	438
6.8.4. Алгоритм для модели с остановкой после отказов: унифицированная синхронизация взаимодействий с представлением.....	443
6.9. Упражнения.....	447
6.10. Решения.....	449
6.11. Практикум	464
6.11.1. Унифицированная рассылка в едином порядке	464
6.11.2. Неблокирующее атомарное подтверждение на основе консенсуса	471
6.11.3. Членство в группе на основе консенсуса	474
6.11.4. Синхронизация с представлением на основе TRB.....	478
6.12. Примечания к главе.....	485
Глава 7. Заключительные замечания	488
7.1. Реализация в Arria	488
7.2. Дополнительные реализации.....	489
7.3. Для дальнейшего чтения	491
Использованная литература	493
Список модулей	501
Список алгоритмов	503
Предметный указатель	506