本书结合GCC4.4.0源代码，围绕GCC编译过程，详细介绍了GCC的设计框架和实现过程，从源代码到AST、从AST到GIMPLE、从GIMPLE到RTL，以及从RTL到最终的目标机器代码的详细过程，涉及各个阶段中间表示的详细分析、生成过程，使读者在了解编译原理的基础上进一步掌握其实现的总体流程和实现细节，让更多的读者对编译技术不再只停留在理论层面，而是能看到一个活生生编译系统实例的实现过程。

本书共有12章，第1章是GCC概述，第2章介绍GCC源代码分析工具，第3章介绍GCC总体结构，第4章介绍从源代码到AST/GENERIC，第5章介绍从AST/GENERIC到GIMPLE，第6章介绍GIMPLE处理及其优化，第7章介绍RTL，第8章介绍机器描述文件${target}.md，第9章介绍机器描述文件${target}.[ch]，第10章介绍从GIMPLE到RTL，第11章介绍RTL处理及其优化，第12章介绍支持新的目标处理器。

本书是作者结合自身科研工作实践和科研兴趣，花费了三年多的时间，通过对GCC4.4.0的源代码进行刻苦研读，是自己在学习、分析编译系统的经验总结，实例丰富，实践性强。

源码分析系列

深入分析GCC

王亚刚◎编著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

深入分析 GCC / 王亚刚编著 . —北京：机械工业出版社，2017.1
（源码分析系列）

ISBN 978-7-111-55632-9

I. 深… II. 王… III. 应用软件 IV. TP317

中国版本图书馆 CIP 数据核字（2016）第 317737 号

深入分析 GCC

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）

责任编辑：张梦玲　　责任校对：董纪丽

印　　刷：北京诚信伟业印刷有限公司　　版　　次：2017 年 2 月第 1 版第 1 次印刷

开　　本：186mm × 240mm　1/16　　印　　张：34.25

书　　号：ISBN 978-7-111-55632-9　　定　　价：99.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：（010）88379426　88361066　　投稿热线：（010）88379604

购书热线：（010）68326294　88379649　68995259　　读者信箱：hzit@hzbook.com

Talk is cheap, show me the code.

——Linus Torralds

前　言

GCC（GNU Compiler Collection，GNU 编译器套件）是一套由 GNU 开发的程序设计语言编译工具，是 GNU 工程中最重要的组成部分。经过近 30 年的发展，GCC 不仅支持众多的前端编程语言，还支持各种主流的处理器平台和操作系统平台，成为公认的跨平台编译器的事实标准，也成为编译器设计的成功典范。

作为一名 GCC 编译器的使用者和源码阅读的爱好者，我一直想写一本关于 GCC 的技术书。

2002 年，我开始在 Linux 系统上进行一些软件开发，GCC 就是我使用的编译工具。我一直对从源代码到目标代码的转换过程充满好奇，想知道在这个过程中 GCC 到底都做了些什么？GCC 是如何设计的，那些成千上万个 GCC 的源代码文件都表示了什么意义？那时我常常幻想，要是能透彻地分析和理解 GCC 源代码，多好！从那时起，在教学科研之余，我偶尔会翻阅一下 GCC 的相关源代码，可是看着繁多的 GCC 源代码，也常常感觉手足无措，真有一种“老虎吃天，无法下爪”的尴尬。于是分析 GCC 源代码的事情被搁置了，然而那种一探究竟的心情总是挥之不去。

2012 年开始，我有了较多的闲暇时间，在经过一段彷徨之后，分析 GCC 源代码的冲动又一次浮现在脑海。我知道，这次是要来真的了，我要做点自己喜欢的事。

Why?

我有空余时间了，我要干些自己感兴趣的事情。在我创建的 GCC 爱好者交流群中经常有朋友问，有没有介绍 GCC 的资料呀？大多人都会说，有——请看官方文档！我也去看了看，没错，GCC 有比较详细的官方文档，包括 gccinternal 及用户手册等。然而，这些文档的内容庞杂，缺乏系统分析 GCC 设计框架和工作流程的内容，并且大多的内容对读者来讲都是零散的，让初学者无所适从。于是我想，为什么不分析一下 GCC 系统，把 GCC 的设计实现用一种更清晰明了、更系统的方法介绍给 GCC 的爱好者呢？

What?

本书将围绕 GCC 编译过程，详细介绍从源代码到 AST、从 AST 到 GIMPLE、从 GIMPLE

到 RTL，以及从 RTL 到最终的目标机器汇编代码的详细过程，涉及各个阶段中间表示的详细分析、生成过程。本书提供了大量的图表和实例，展示了 GCC 编译系统的总体工作流程和工作细节。本书的另外一个特点是结合 GCC 4.4.0 的源代码进行分析，使读者在了解编译原理的基础上进一步掌握其实现的总体流程和细节，让更多读者对编译技术的认识不再只停留在理论层面，而是向其展示一个编译系统实例的实现过程。

How?

GCC 源代码涉及的内容非常庞杂，很难在一本书中全面描述，因此本书以 GCC 中间表示为主线，详细分析 GCC 从源代码开始，直到生成目标机器汇编代码的整个过程中所使用的三种中间表示（AST、GIMPLE 及 RTL），并对这三种中间表示的基本概念、生成过程进行详细的描述，对基于 GIMPLE 和 RTL 的优化处理进行介绍，从而描述一条从源代码到目标机器汇编代码的清晰路线图。

Who?

本书以热爱编译系统理论及其实现的在校大学生、研究生为主要读者对象，也可以作为企业中研发编译系统以及进行编译系统移植的研发工程师的有益参考。

在编写这本书的时候，有一种精神支撑着我，我相信“兴趣”加上“坚持”就是胜利！分析 GCC 不是一年半载的事情，需要 3 年、5 年，甚至更长时间，不过我可以坚持，我要用我的坚持换来对 GCC 的深入分析，让更多的 GCC 爱好者熟悉它、接触它、了解它，更多地参与 GCC 的开发与维护。

感谢我的爱人和孩子，给了我家的温暖和亲情。感谢病榻上的父亲，虽然他不能和我说话，但他那一双大手，依然经常抚摸在我的头上。感谢年老体弱的母亲，感谢她一直照顾我的父亲，让我知道什么是坚持，什么是不离不弃！感谢西安邮电大学 GPU 项目组的各位同事在本书的写作中提出的宝贵建议。

本书的写作得到国家自然科学基金重点项目（项目编号：61136002）以及陕西省教育厅科研计划项目（项目编号：14JK1674）资助。

鉴于作者水平有限，在分析和写作本书的过程中也引入了一些个人观点，因此难免有一些理解的偏差和错误，敬请读者批评指正并不吝赐教，如有意见和建议，请联系作者 lazy_linux@126.com，在此一并感谢！

王亚刚

2016 年 10 月于西安邮电大学

目　录

前言

第 1 章　GCC 概述 …… 1

1.1　GCC 的产生与发展 …… 1

1.2　GCC 的特点 …… 2

1.3　GCC 代码分析 …… 3

第 2 章　GCC 源代码分析工具 …… 4

2.1　vim+ctags 代码阅读工具 …… 4

2.2　GNU gdb 调试工具 …… 6

2.3　GNU binutils 工具 …… 8

2.4　shell 工具及 graphviz 绘图工具 …… 11

2.5　GCC 调试选项 …… 13

第 3 章　GCC 总体结构 …… 16

3.1　GCC 的目录结构 …… 16

3.2　GCC 的逻辑结构 …… 18

3.3　GCC 源代码编译 …… 20

3.3.1　配置 …… 21

3.3.2　编译 …… 23

3.3.3　安装 …… 25

第 4 章　从源代码到 AST/GENERIC …… 26

4.1　抽象语法树 …… 26

4.2　树节点的声明 …… 28

4.3　树节点结构 …… 33

4.3.1　struct tree_base …… 35

4.3.2　struct tree_common …… 36

4.3.3　常量节点 …… 38

4.3.4　标识符节点 …… 42

4.3.5　声明节点 …… 44

4.3.6　struct tree_decl_minimal …… 46

4.3.7　struct tree_decl_common …… 46

4.3.8　struct tree_field_decl …… 49

4.3.9　struct tree_decl_with_rtl …… 55

4.3.10　struct tree_label_decl …… 55

4.3.11　struct tree_result_decl …… 56

4.3.12　struct tree_const_decl …… 57

4.3.13　struct tree_parm_decl …… 57

4.3.14　struct tree_decl_with_vis …… 59

4.3.15　struct tree_var_decl …… 59

4.3.16　struct tree_decl_non_common …… 62

4.3.17　struct tree_function_decl …… 62

4.3.18　struct tree_type_decl …… 64

4.3.19　类型节点 …… 67

4.3.20　tree_list 节点 …… 68

4.3.21　表达式节点 …… 71

4.3.22　语句节点 …… 73

4.3.23 其他树节点 …… 75
4.4 AST 输出及图示 …… 76
4.5 AST 的生成 …… 83
4.5.1 词法分析 …… 84
4.5.2 词法分析过程 …… 90
4.5.3 语法分析 …… 98
4.5.4 语法分析过程 …… 99
4.5.5 c_parse_file …… 103
4.5.6 c_parser_translation_unit …… 105
4.5.7 c_parser_external_declaration …… 105
4.5.8 c_parser_declaration_or_fndef …… 107
4.5.9 c_parser_declspecs …… 112
4.6 小结 …… 114

第 5 章 从 AST/GENERIC 到 GIMPLE …… 115

5.1 GIMPLE …… 115
5.2 GIMPLE 语句 …… 119
5.3 GIMPLE 的表示与存储 …… 122
5.4 GIMPLE 语句的操作数 …… 128
5.5 GIMPLE 语句序列的基本操作 …… 132
5.6 GIMPLE 的生成 …… 135
5.6.1 gimplify_function_tree …… 136
5.6.2 gimplify_body …… 138
5.6.3 gimlify_parameters …… 139
5.6.4 gimplify_stmt …… 144
5.6.5 gimplify_expr …… 144
5.7 GIMPLE 转换实例 …… 157
5.7.1 BIND_EXPR 节点的 GIMPLE 生成 …… 158
5.7.2 STATEMENT_LIST_EXPR 节点的 GIMPLE 生成 …… 159
5.7.3 MODIFY_EXPR 节点的 GIMPLE 生成 …… 160
5.7.4 POSTINCREMENT_EXPR 节点的 GIMPLE 生成 …… 162
5.8 实例分析 …… 172
5.9 小结 …… 176

第 6 章 GIMPLE 处理及其优化 … 177

6.1 GCC Pass …… 177
6.1.1 核心数据结构 …… 177
6.1.2 Pass 的类型 …… 179
6.1.3 Pass 链的初始化 …… 182
6.1.4 Pass 的执行 …… 184
6.2 Pass 列表 …… 187
6.3 GIMPLE Pass 实例 …… 193
6.3.1 pass_remove_useless_stmts … 193
6.3.2 pass_lower_cf …… 195
6.3.3 pass_build_cfg …… 197
6.3.4 pass_build_cgraph_edges …… 203
6.3.5 pass_build_ssa …… 205
6.3.6 pass_all_optimizations …… 206
6.3.7 pass_expand …… 207
6.4 小结 …… 207

第 7 章 RTL …… 208

7.1 RTL 中的对象类型 …… 209
7.2 RTX_CODE …… 210
7.3 RTX 类型 …… 210
7.4 RTX 输出格式 …… 212
7.5 RTX 操作数 …… 213
7.6 RTX 的机器模式 …… 216

7.7 RTX 的存储 …… 219
7.8 RTX 表达式 …… 222
7.8.1 常量 …… 225
7.8.2 寄存器和内存 …… 227
7.8.3 算术运算 …… 228
7.8.4 比较运算 …… 230
7.8.5 副作用 …… 230
7.9 IR-RTL …… 232
7.9.1 INSN …… 233
7.9.2 JUMP_INSN …… 234
7.9.3 CALL_INSN …… 235
7.9.4 BARRIER …… 235
7.9.5 CODE_LABEL …… 236
7.9.6 NOTE …… 237
7.10 小结 …… 238

第 8 章 机器描述文件 ${target}.md …… 239

8.1 机器描述文件 …… 240
8.2 指令模板 …… 241
8.2.1 模板名称 …… 242
8.2.2 RTL 模板 …… 246
8.2.3 条件 …… 256
8.2.4 输出模板 …… 256
8.2.5 属性 …… 256
8.3 定义 RTL 序列 …… 257
8.4 指令拆分 …… 263
8.5 枚举器 …… 266
8.5.1 mode 枚举器 …… 266
8.5.2 code 枚举器 …… 268
8.6 窥孔优化 …… 269
8.6.1 define_peephole …… 269
8.6.2 define_peephole2 …… 270
8.7 小结 …… 271

第 9 章 机器描述文件 ${target}.[ch] …… 272

9.1 targetm …… 272
9.1.1 struct gcc_target 的定义 …… 273
9.1.2 targetm 的初始化 …… 277
9.2 编译驱动及选项 …… 279
9.2.1 编译选项 …… 280
9.2.2 SPEC 语言及 SPEC 文件 …… 281
9.2.3 机器相关的编译选项 …… 285
9.3 存储布局 …… 286
9.3.1 位顺序和字节顺序 …… 286
9.3.2 类型宽度 …… 287
9.3.3 机器模式提升 …… 287
9.3.4 存储对齐 …… 288
9.3.5 编程语言中数据类型的存储布局 …… 289
9.4 寄存器使用 …… 290
9.4.1 寄存器的基本描述 …… 290
9.4.2 寄存器分配顺序 …… 297
9.4.3 机器模式 …… 298
9.4.4 寄存器类型 …… 300
9.5 堆栈及函数调用规范描述 …… 307
9.5.1 堆栈的基本特性 …… 309
9.5.2 寄存器消除 …… 313
9.5.3 函数栈帧的管理 …… 315
9.5.4 参数传递 …… 316
9.5.5 函数返回值 …… 318
9.5.6 i386 机器栈帧 …… 318
9.6 寻址方式 …… 325
9.7 汇编代码分区 …… 326
9.8 定义输出的汇编语言 …… 333

9.8.1 汇编代码文件的框架 …… 333
9.8.2 数据输出 …… 336
9.8.3 未初始化数据输出 …… 336
9.8.4 标签输出 …… 338
9.8.5 指令输出 …… 342
9.9 机器描述信息的提取 …… 343
9.9.1 gencode.c …… 347
9.9.2 genattr.c …… 348
9.9.3 genattrtab.c …… 348
9.9.4 genrecog.c …… 349
9.9.5 genflag.c …… 352
9.9.6 genemit.c …… 353
9.9.7 genextract.c …… 354
9.9.8 genopinit.c …… 356
9.9.9 genoutput.c …… 360
9.9.10 genpreds.c …… 362
9.9.11 其他 …… 363
9.10 小结 …… 364

第 10 章 从 GIMPLE 到 RTL … 365

10.1 GIMPLE 序列 …… 365
10.2 典型数据结构 …… 366
10.3 RTL 生成的基本过程 …… 367
10.3.1 变量展开 …… 370
10.3.2 参数及返回值处理 …… 380
10.3.3 初始块的处理 …… 395
10.3.4 基本块的 RTL 生成 …… 398
10.3.5 退出块的处理 …… 410
10.3.6 其他处理 …… 411
10.4 GIMPLE 语句转换成 RTL …… 411
10.4.1 GIMPLE 语句转换的一般过程 …… 412
10.4.2 GIMPLE_GOTO 语句的 RTL 生成 …… 415
10.4.3 GIMPLE_ASSIGN 语句的 RTL 生成 …… 417
10.5 小结 …… 432

第 11 章 RTL 处理及优化 …… 433

11.1 RTL 处理过程 …… 433
11.2 特殊虚拟寄存器的实例化 …… 435
11.3 指令调度 …… 437
11.3.1 指令调度算法 …… 439
11.3.2 GCC 指令调度的实现 …… 440
11.3.3 指令调度实例 1 …… 442
11.3.4 指令调度实例 2 …… 459
11.4 统一寄存器分配 …… 460
11.4.1 基本术语 …… 461
11.4.2 寄存器分配的主要流程 … 463
11.4.3 代码分析 …… 466
11.4.4 寄存器分配实例 1 …… 468
11.4.5 寄存器分配实例 2 …… 483
11.5 汇编代码生成 …… 494
11.5.1 汇编代码文件的结构 …… 495
11.5.2 从 RTL 到汇编代码 …… 499
11.6 小结 …… 502

第 12 章 支持新的目标处理器 … 503

12.1 GCC 移植 …… 503
12.2 PAAG 处理器 …… 504
12.2.1 PAAG 处理器指令集 …… 505
12.2.2 应用二进制接口 …… 505
12.3 GCC 移植的基本步骤 …… 506
12.4 PAAG 机器描述文件 (paag.md) …… 507

12.5 paag.[ch] 文件 …… 512
12.5.1 存储布局 …… 512
12.5.2 寄存器使用规范 …… 513
12.5.3 堆栈布局及堆栈指针 …… 514
12.5.4 函数调用规范 …… 515
12.5.5 寻址方式 …… 519
12.5.6 汇编代码输出 …… 521
12.5.7 杂项 …… 523
12.6 PAAG 后端注册 …… 523
12.7 GCC 移植测试 …… 524
12.8 小结 …… 526
参考文献 …… 527
索引 …… 529