

南开大学

计算机学院 预备工作2

定义编译器 & 汇编编程

丁屹、卢麒萱

年级: 2020 级

专业:计算机科学与技术

摘要

关键字: CFG 设计 ARM 汇编

景目

一、总	体设计及分工	1
二 、 C	FG 描述	1
(-)	标识符及数字	1
(二)	变量/常量声明	1
(三)	数组声明	2
(四)	赋值语句	2
(五)	复合语句	2
(六)	if 分支语句	2
(七)	for/while 循环语句	2
$(/ \setminus)$	算术表达式	3
(九)	关系运算表达式	3
(十)	逻辑表达式	3
(+-) 函数	3
三 、A	RM 汇编	4
(-)	SysY 程序设计	4
(二)	卢麒萱的工作	4
(三)	丁屹的工作	8

一、 总体设计及分工

经过讨论,确定我们实现的编译器 SysY 语言特性有:

- 1. 变量/常量声明: 支持 int、float 类型
- 2. 数组声明: 支持 int、float 类型及数组指针
- 3. 赋值语句
- 4. 复合语句
- 5. if 分支语句
- 6. for/while 循环语句
- 7. 算术表达式: 支持加减乘除、按位与或
- 8. 关系运算表达式: 支持不等、等于、大于、小于
- 9. 逻辑表达式: 支持逻辑与或
- 10. 函数

二、 CFG 描述

(一) 标识符及数字

$$id
ightarrow id - nondigit \mid id \ id - nondigit \mid id \ number$$

$$id - nondigit
ightarrow \{_{\bf A} - {\bf Za} - {\bf z}\}$$

$$digit
ightarrow number \ digit$$

$$number
ightarrow \ \{{\bf 0} - {\bf 9}\}$$

id 表示标识符,以下都认为为终结符, num 表示数字字符,digit 表示数字。同名标识符的约定:

- 全局变量和局部变量的作用域可以重叠,重叠部分局部变量优先;同名局部变量的作用域 不能重叠;
- 变量名可以与函数名相同。

(二) 变量/常量声明

$$type \rightarrow \mathbf{int} \mid \mathbf{float}$$
 $idlist \rightarrow idlist, \mathbf{id} \mid \mathbf{id}$
 $decl \rightarrow type \ idlist$

其中 type 代表变量类型, decl 代表声明语句, idlist 代表标识符列表。

二、 CFG 描述 编译系统原理报告

(三) 数组声明

$$\begin{split} arr_decl \to type \ arr_idlist \\ arr_idlist \to arr_idlist, \mathbf{id}[arr_size] \mid \mathbf{id}[arr_size] \mid *\mathbf{id} \\ arr_size \to \mathbf{id} \mid \mathbf{number} \mid \epsilon \end{split}$$

其中 arr_decl 代表数组声明语句, arr_idlist 表示数组标识符列表, arr_size 表示数组定义大小。

(四) 赋值语句

$$unary_expr \rightarrow digit \mid \mathbf{id}$$

$$assign_expr \rightarrow unary_expr = assign_expr \mid logical_expr$$

$$unary_expr \rightarrow prim_expr \mid funcname(paralist) \mid unary_op \ unary_expr$$

$$prim_expr \rightarrow (add_expr) \mid Lval \mid number$$

$$L_val \rightarrow \mathbf{id} \ [exp]$$

$$unary_op \rightarrow \{+,-,!\}$$

logical_expr 为逻辑表达式, unary_expr 为一元表达式, assign_expr 为赋值表达式, 这里最后一个产生式第二个右部为逻辑表达式的原因是因为逻辑表达式的逻辑与和逻辑或优先级高于赋值表达式但却低于关系表达式的比较和算术表达式的各种运算, prim_expr 为基本表达式, L_val 为左值表达式, unary_op 为单目运算符, '!' 仅出现在条件表达式中。

(五) 复合语句

$$block \rightarrow \{blockitem\}$$

 $blockitem \rightarrow blockitem; blockitem \mid decl \mid stmt$

其中 block 表示一个复合语句块, blockitem 表示中间内容, 由一或多条表达式或分支循环语句组成。

(六) if 分支语句

$$stmt \rightarrow if(expr) \ stmt \ else \ stmt$$

其中 stmt 为语句, expr 为表达式。

(七) for/while 循环语句

$$stmt \rightarrow \mathbf{while}(expr) \ stmt$$

 $stmt \rightarrow \mathbf{for}(expr; expr; expr) \ stmt$

其中 stmt 为语句, expr 为表达式。

二、 CFG 描述 编译系统原理报告

(八) 算术表达式

$$mul_expr
ightarrow unary_expr \mid mul_expr \ \mathbf{mul_op} \ unary_expr$$

$$mul_op
ightarrow \{*,/\}$$

$$add_expr
ightarrow mul_expr \mid add_expr \ \mathbf{add_op} \ mul_expr$$

$$add_op
ightarrow \{+,-\}$$

其中 mul_expr 为乘除模运算表达式, mul_op 为乘除模符号, add_expr 为加减运算表达式, add op 为加减运算符。

(九) 关系运算表达式

$$\begin{split} rel_expr &\rightarrow add_expr \mid rel_expr \ rel_op \ add_expr \\ &rel_op \rightarrow \{<,>,<=,>=\} \\ equ_expr \rightarrow rel_expr \mid equ_expr \ equ_op \ rel_expr \\ &equ_op \rightarrow \{==,!=\} \end{split}$$

其中 rel_expr 为包含大于、小于、大于等于、小于等于的关系表达式, equ_expr 为包含等于和不等于的关系表达式。

(十) 逻辑表达式

$$logical_expr \rightarrow lor_expr$$

$$lor_expr \rightarrow land_expr \mid lor_expr \; lor_op \; land_expr$$

$$lor_op \rightarrow \{||\}$$

$$land_expr \rightarrow equ_expr \mid land_expr \; land_op \; equ_expr$$

$$land \; op \rightarrow \{\&\&\}$$

(十一) 函数

$$\begin{split} funcdef &\to type\ funcname(paralist)\ stmt \\ paralist &\to paralist, paradef \mid paradef \mid \epsilon \\ paradef &\to type\ \mathbf{id} \\ funcname &\to \mathbf{id} \end{split}$$

其中 paralist 代表参数列表,paradef 代表参数声明,funcname 代表函数名,funcdef 代表函数声明语句。

三、 ARM 汇编

(一) SysY 程序设计

为满足上述选定语言特性,我们二人分别设计了两个 SysY 程序,涵盖了我们定义的所有语言特性。这里使用了标准库的 scanf、printf,后续实现 SysY 时会替换成 syslib 的 I/O 函数。

(二) 卢麒萱的工作

编写的 SysY 示例程序如下:

```
#include <stdio.h>
   int a, b;
   float c[5] = \{1, 2, 3, 4, 5\};
   int f(int a, int b) {
     if(a > b) return 1;
     else return 0;
   }
10
   int main() {
11
     scanf("%d%d", &a, &b);
     const int d = 0;
13
       int d = f(a, b);
15
       if(d == 1) {
         for(int i = 0; i < 5; i ++)
17
            d = d * c[i];
          printf("%d\n", d);
19
       } else d = 1;
20
21
     printf("%d\n", d);
```

编写对应 ARM 汇编代码及注释如下,这里的循环恰好可以展开,使用 NEON 指令 SIMD 化。

```
.arch armv7-a

.fpu vfpv3-d16

.text

.global c

.bss

.type c, %object
```

```
.size c, 20
   a:
     .space 4
10
     .global b
11
     .type b, %object
12
     .size b, 4
13
   b:
14
     .space 4
15
     .global c
16
     .data
17
     .type c, %object
     .size c, 20
19
     @ 浮点数 1-5 在内存中存放的地址按照 "IEEE 754 标准"单精度格式编码的浮点数,对应的十进制值
21
     .word 1065353216
     .word 1073741824
23
     .word 1077936128
     .word 1082130432
25
     .word 1084227584
26
     .text
27
     .global f
28
     .type f, %function
29
   @ 函数部分
30
   f:
31
     sub sp, sp, #12
32
     str r0, [sp, #4]
     str r1, [sp]
34
     ldr r2, [sp, #4]
35
     ldr r3, [sp]
36
     cmp r2, r3
37
     ble .L1
     movs r3, #1
39
     b .end
40
   .L1:
41
     movs r3, #0
   .end:
43
     mov r0, r3
44
     adds sp, sp, #12
45
     mov pc, lr
   @ 读入字符串
47
   .str0:
     .ascii "%d%d\000"
49
    .align 2
   .str1:
```

```
.ascii "%d\012\000"
52
     .text
53
     .global main
54
     .syntax unified
55
     .thumb
56
     .type main, %function
57
   @ 主函数
   main:
59
     push {fp, lr}
     sub sp, sp, #16
61
     add fp, sp, #0
62
     ldr r3, ._bridge
63
   .LPICO:
     add r3, pc
65
     mov r2, r3
     ldr r3, ._bridge+4
67
   .LPIC1:
68
     add r3, pc
     mov r1, r3
70
     ldr r3, ._bridge+8
71
    .LPIC2:
72
     add r3, pc
73
     mov r0, r3
74
     bl scanf
     movs r3, #0
76
     str r3, [fp, #4]
     ldr r3, ._bridge+12
78
   .LPIC3:
     add r3, pc
80
     ldr r2, [r3]
81
     ldr r3, ._bridge+16
    .LPIC4:
83
     add r3, pc
     ldr r3, [r3]
85
     mov r1, r3
     mov r0, r2
87
     bl f
     str r0, [fp, #12]
89
     ldr r3, [fp, #12]
     cmp r3, #1
91
     bne .L5
     movs r3, #0
93
     str r3, [fp, #8]
     b .L6
```

```
.L7:
      ldr r3, [fp, #12]
97
      vmov s15, r3
98
      vcvt.f32.s32 s14, s15
99
      ldr r2, ._bridge+20
100
    .LPIC5:
101
      add r2, pc
102
      ldr r3, [fp, #8]
103
      lsls r3, r3, #2
104
      add r3, r3, r2
105
      vldr.32 s15, [r3]
106
      vmul.f32 s15, s14, s15
107
      vcvt.s32.f32 s15, s15
108
      vmov r3, s15
109
      str r3, [fp, #12]
110
      ldr r3, [fp, #8]
111
      adds r3, r3, #1
112
      str r3, [fp, #8]
113
    .L6:
114
      ldr r3, [fp, #8]
115
      cmp r3, #4
116
      ble .L7
117
      ldr r1, [fp, #12]
118
      ldr r3, ._bridge+24
119
    .LPIC6:
120
      add r3, pc
121
      mov r0, r3
122
      bl printf
123
      b .L8
124
    .L5:
125
      movs r3, #1
126
      str r3, [fp, #12]
127
    .L8:
128
      ldr r1, [fp, #4]
129
      ldr r3, ._bridge+28
    .LPIC7:
131
      add r3, pc
132
      mov r0, r3
133
      bl printf
134
      movs r3, #0
135
      mov r0, r3
136
      adds fp, fp, #16
137
      mov sp, fp
138
      pop {fp, pc}
139
```

```
@ 桥接隐性全局变量的地址
140
    ._bridge:
141
      .word b-(.LPICO+4)
142
      .word a-(.LPIC1+4)
143
      .word .str0-(.LPIC2+4)
144
      .word a-(.LPIC3+4)
145
      .word b-(.LPIC4+4)
      .word c-(.LPIC5+4)
147
      .word .str1-(.LPIC6+4)
      .word .str1-(.LPIC7+4)
149
      .size main, .-main
```

测试代码正确性,编译静态库 sylib.a 后,通过 arm-linux-gnueabihf-gcc sample.arm.s sylib.a -o sample.arm 进行链接生成可执行代码,再使用 qemu 运行,测试如图1所示。

图 1: 汇编代码 1 测试

根据程序设定, a <= b 和 a > b 得到的结果均正确。

(三) 丁屹的工作

编写的 SysY 示例程序如下,使用了计算斐波那契数列的代码加以扩展。fib 函数计算 n 步以后的斐波那契数, a, b 指针指向的变量保存最后两个斐波那契数,中间计算步骤保存在全局静态 trace 数组中。程序会向标准输入读取初始两个斐波那契数,计算步长固定为 10。最后会向标准输出打印最后两个斐波那契数的和差积商以及按位与或的结果和计算步骤。

```
#include <stdio.h>
   int trace[18];
   int* ptrace
                         = trace;
   float (*nonsense)[2] = trace;
   float (*ns)[2][3]
                         = trace;
   float nnss[5][4][3][2];
   void fib(int n, int* a, int* b) {
     for (int i = 0; i < n; ++i) {
       int t = *b:
11
       *b += *a;
12
       trace[i] = *b;
13
```

```
*a
                   = t;
14
      }
15
    }
16
17
    int main() {
18
      int
                   a = 1, b = 1;
19
      const int n
                              = 10;
20
                   s_{fmt}[19] = "%d%d";
21
      const char p_{fmt}[19] = "%d %d %d %d %d %d %d %n";
22
      scanf(s_fmt, &a, &b);
23
      if (a > 0 \&\& 0 < b
24
           | \ | \ (a >= 0 \ | \ | \ 0 <= b) \&\& \ a > -1 \&\& \ -1 < b \&\& \ (a \ != 0 \ | \ 1 \ == b)) \ \{
25
        // a > 0, b > 0 or a == 0, b == 1
        fib(n, &a, &b);
27
        printf(p_fmt, a + b, a - b, a * b, a / b, a & b, a | b);
        int i = 0;
29
        while (1) {
30
           printf("%d ", trace[i]);
31
          ++i;
32
          if (i == 10) {
33
             printf("\n");
34
             break;
35
           } else continue;
36
           printf("unreachable!\n");
        }
38
      } else {
        char err[] = "invalid input\n";
40
        printf(err);
41
        return 1;
42
      }
43
    }
```

编写对应 ARM 汇编代码及注释如下:

```
.arch armv7-a
.fpu vfpv3-d16

.text
.global trace
.bss
.type trace, %object
.size trace, 72

trace:
.space 72
```

```
11
    .global ptrace
12
    .section .data.rel.local,"aw"
13
    .type ptrace, %object
14
    .size ptrace, 4
   ptrace:
16
    .word trace
18
    .global nonsense
    .type nonsense, %object
20
    .size nonsense, 4
21
   nonsense:
22
    .word trace
24
    .global ns
25
    .type ns, %object
26
    .size ns, 4
27
   ns:
28
    .word trace
29
30
    .global nnss
31
    .bss
32
    .type nnss, %object
33
    .size nnss, 480
   nnss:
35
    .space 480
37
38
    .text
    .global fib
39
    .type fib, %function
40
41
      str fp, [sp, #-4]!
42
      add fp, sp, #0
43
      sub sp, sp, #28
44
      str r0, [fp, #-16]
45
      @ int n
46
      str r1, [fp, #-20]
      0 int* a
48
      str r2, [fp, #-24]
      @ int* b
50
      mov r3, #0
      str r3, [fp, #-8]
52
      0 int i = 0
      b .L0
54
```

```
@ for (int i = 0; i < n; ++i)
55
    .L1:
56
      ldr r3, [fp, #-24]
57
      @ b
58
      ldr r3, [r3]
59
      @ *b
60
      str r3, [fp, #-12]
61
      0 int t = *b
62
      ldr r2, [fp, #-20]
      @ a
64
      ldr r2, [r2]
65
      0 *a
66
      add r3, r2, r3
      ldr r2, [fp, #-24]
68
      str r3, [r2]
      0 *b += *a
70
      ldr r1, [r2]
71
      @ *b
72
      ldr r2, [fp, #-12]
73
      0 *a = t
      ldr r3, .LTR
75
    .LPICO:
76
      add r3, pc, r3
77
      ldr r2, [fp, #-8]
      @ i
79
      str r1, [r3, r2, lsl #2]
      @ trace[i] = *b
81
      ldr r3, [fp, #-20]
      @ a
      ldr r2, [fp, #-12]
      0 t
85
      str r2, [r3]
86
      0 *a = t
      ldr r3, [fp, #-8]
88
      add r3, r3, #1
      str r3, [fp, #-8]
90
      @ ++i
91
    .LO:
92
      ldr r2, [fp, #-8]
      @ i
94
      ldr r3, [fp, #-16]
95
      @ n
96
      cmp r2, r3
      blt .L1
```

```
0 \text{ for } (\ldots; i < n; \ldots)
99
       add sp, fp, #0
100
       ldr fp, [sp], #4
101
       bx lr
102
       @ return
103
104
     .LTR:
105
     .word trace-(.LPICO+8)
106
     .size fib, .-fib
107
     .align 2
108
109
     .global __aeabi_idiv
110
     .section .rodata
111
     .align 2
112
     .LUNMSTRO:
114
     .ascii "%d \000"
115
     .align 2
116
117
     .LSFMT:
118
     .ascii "%d%d\000"
119
     .space 14
120
     .align 2
121
122
     .LPFMT:
123
     .ascii "%d %d %d %d %d %d\012\000"
     .align 2
125
126
     .LINVSTR:
127
     .ascii "invalid input\012\000"
128
     .text
129
     .align 2
130
131
     .global main
132
     .type main, %function
    main:
134
       push {r4, r5, r6, fp, lr}
135
       add fp, sp, #16
136
       sub sp, sp, #92
       mov r3, #1
138
       str r3, [fp, #-32]
139
       0 = 1
140
       str r3, [fp, #-36]
141
       0 b = 1
142
```

```
mov r3, #10
143
       str r3, [fp, #-28]
144
       0 n = 10
145
       ldr r2, .LPICT
146
     .LPIC1:
147
       add r2, pc, r2
148
       0 \text{ char s_fmt[19]} = "%d%d"
149
       sub r3, fp, #56
150
       ldm r2, {r0, r1}
151
       str r0, [r3]
152
       add r3, r3, #4
       strb r1, [r3]
154
       sub r3, fp, #51
155
       mov r2, #0
156
       str r2, [r3]
       str r2, [r3, #4]
158
       str r2, [r3, #8]
159
       strh r2, [r3, #12]
160
       ldr r3, .LPICT+4
161
       @ const char p_fmt[19] = "%d %d %d %d %d %d %d\n"
162
     .LPIC2:
163
       add r3, pc, r3
164
       sub ip, fp, #76
165
       mov lr, r3
       ldmia lr!, {r0, r1, r2, r3}
167
       stmia ip!, {r0, r1, r2, r3}
       ldr r3, [lr]
169
       strh r3, [ip]
170
       add ip, ip, #2
171
       lsr r3, r3, #16
172
       strb r3, [ip]
173
       sub r2, fp, #36
174
       sub r1, fp, #32
       sub r3, fp, #56
176
       mov r0, r3
       bl scanf(PLT)
178
       @ scanf(s_fmt, &a, &b)
179
       ldr r3, [fp, #-32]
180
       cmp r3, #0
181
       ble .L2
182
       @ a > 0
183
       ldr r3, [fp, #-36]
184
       cmp r3, #0
185
       bgt .L3
186
```

```
@ 0 < b
187
     .L2:
188
       ldr r3, [fp, #-32]
189
       cmp r3, #0
190
       bge .L4
191
       0 a >= 0
192
       ldr r3, [fp, #-36]
193
       cmp r3, #0
194
       blt .LINV
195
       @ 0 <= b
196
     .L4:
197
       ldr r3, [fp, #-32]
198
       cmp r3, #-1
199
       ble .LINV
200
       0 \ a > -1
201
       ldr r3, [fp, #-36]
202
       cmp r3, #-1
203
       ble .LINV
204
       @ -1 < b
205
       ldr r3, [fp, #-32]
206
       cmp r3, #0
207
       bne .L3
208
       @ a != 0
209
       ldr r3, [fp, #-36]
       cmp r3, #1
211
       bne .LINV
       0 1 == b
213
     .L3:
214
       @ exec fib
215
       sub r2, fp, #36
216
       @ &b
217
       sub r3, fp, #32
218
       mov r1, r3
219
       0 &a
220
       ldr r0, [fp, #-28]
221
       @ n
222
       bl fib(PLT)
223
       @ fib
224
       ldr r2, [fp, #-32]
       @ a
226
       ldr r3, [fp, #-36]
       @ b
228
       add r4, r2, r3
229
       @ a + b
230
```



```
sub r5, r2, r3
231
       @ a - b
232
       mul r6, r2, r3
233
       @ a * b
234
       mov r1, r3
235
       @ r1 = b
236
       mov r0, r2
237
       0 \text{ r0} = a
238
       bl __aeabi_idiv(PLT)
239
       @ r0 /= r1 -> a / b
240
       mov ip, r0
241
       ldr r1, [fp, #-32]
242
       @ a
       ldr r2, [fp, #-36]
244
       @ b
       and r3, r1, r2
246
       @ a & b
247
       orr r2, r1, r2
248
       @ a | b
249
       sub r0, fp, #76
250
       @ push p_fmt
251
       str r2, [sp, #8]
252
       @ push a | b
253
       str r3, [sp, #4]
254
       @ push a & b
255
       str ip, [sp]
       @ push a / b
257
       mov r3, r6
258
       @ push a * b
259
       mov r2, r5
260
       @ push a - b
261
       mov r1, r4
262
       @ push a + b
263
       bl printf(PLT)
264
       mov r3, #0
       str r3, [fp, #-24]
266
     .L5:
267
       ldr r3, .LPICT+8
268
       @ "%d "
     .LPIC3:
270
       add r3, pc, r3
       ldr r2, [fp, #-24]
272
       ldr r3, [r3, r2, lsl #2]
273
       mov r1, r3
^{274}
```



```
@ push "%d "
275
      ldr r0, .LPICT+12
     .LPIC4:
277
      add r0, pc, r0
278
      @ push trace[i]
279
      bl printf(PLT)
280
      ldr r3, [fp, #-24]
      add r3, r3, #1
282
      str r3, [fp, #-24]
283
      @ ++i
284
      cmp r3, #10
      bne .L5
286
      @ i != 10
      mov r0, #10
288
      bl putchar(PLT)
      @ printf("\n") -> putchar(10)
290
      mov r0, #0
291
      b .L6
292
      @ printf("unreachable!\n") can be deleted
293
    .LINV:
294
      ldr r3, .LPICT+16
295
      @ "invalid input\n"
296
     .LPIC5:
297
      add r3, pc, r3
      sub ip, fp, #92
299
      ldm r3, {r0, r1, r2, r3}
      stmia ip!, {r0, r1, r2}
301
      strh r3, [ip]
302
      add ip, ip, #2
303
      lsr r3, r3, #16
304
      strb r3, [ip]
305
      sub r3, fp, #92
306
      mov r0, r3
307
      bl printf(PLT)
308
      mov r0, #1
     .L6:
310
      sub sp, fp, #16
311
      pop {r4, r5, r6, fp, pc}
312
     .LPICT:
       .word .LSFMT-(.LPIC1+8)
314
       .word .LPFMT-(.LPIC2+8)
       .word trace-(.LPIC3+8)
316
       .word .LUNMSTRO-(.LPIC4+8)
317
       .word .LINVSTR-(.LPIC5+8)
```

319 .size main, .-main

测试代码正确性,使用 Makefile 的目标 arm-build-run 自动构建,使用 qemu-arm 环境模拟运行,测试如图2所示

图 2: 汇编代码 2 测试