

编译器认知实验报告

实验目的

本实验的目的是了解工业界常用的编译器GCC和LLVM，熟悉编译器的安装和使用过程，观察编译器工作过程中生成的中间文件的格式和内容，了解编译器的优化效果，为编译器的学习和构造奠定基础。

实验内容

在Linux平台上安装和运行工业界常用的编译器GCC和LLVM，如果系统中没有安装，则需要首先安装编译器，安装完成后编写简单的测试程序，使用编译器编译，并观察中间输出结果。

实现的具体过程和步骤

- 对于GCC 编译器，完成编译器安装和测试程序编写后，按如下步骤完成：
 - 查看编译器的版本
 - 使用编译器编译单个文件
 - 使用编译器编译链接多个文件
 - 查看预处理结果： `gcc -E main.c -o main.i`
 - 查看语法分析树： `gcc -fdump-tree-all main.c`
 - 查看中间代码生成结果： `gcc -fdump-rtl-all main.c`
 - 查看生成的目标代码（汇编代码）： `gcc -S main.c -o main.s`
- 对于LLVM 编译器，完成编译器安装和测试程序编写后，按如下步骤完成：
 - 查看编译器的版本
 - 使用编译器编译单个文件
 - 使用编译器编译链接多个文件
 - 查看编译流程和阶段： `clang -ccc-print-phases main.c -c`
 - 查看词法分析结果： `clang main.c -Xclang -dump-tokens`
 - 查看词法分析结果2： `clang main.c -Xclang -dump-raw-tokens`
 - 查看语法分析结果： `clang main.c -Xclang -ast-dump`
 - 查看语法分析结果2： `clang main.c -Xclang -ast-view`
 - 查看编译优化的结果： `clang main.c -S -mllvm -print-after-all`
 - 查看生成的目标代码结果： `clang main.c -S`

GCC运行结果分析

- 查看编译器版本

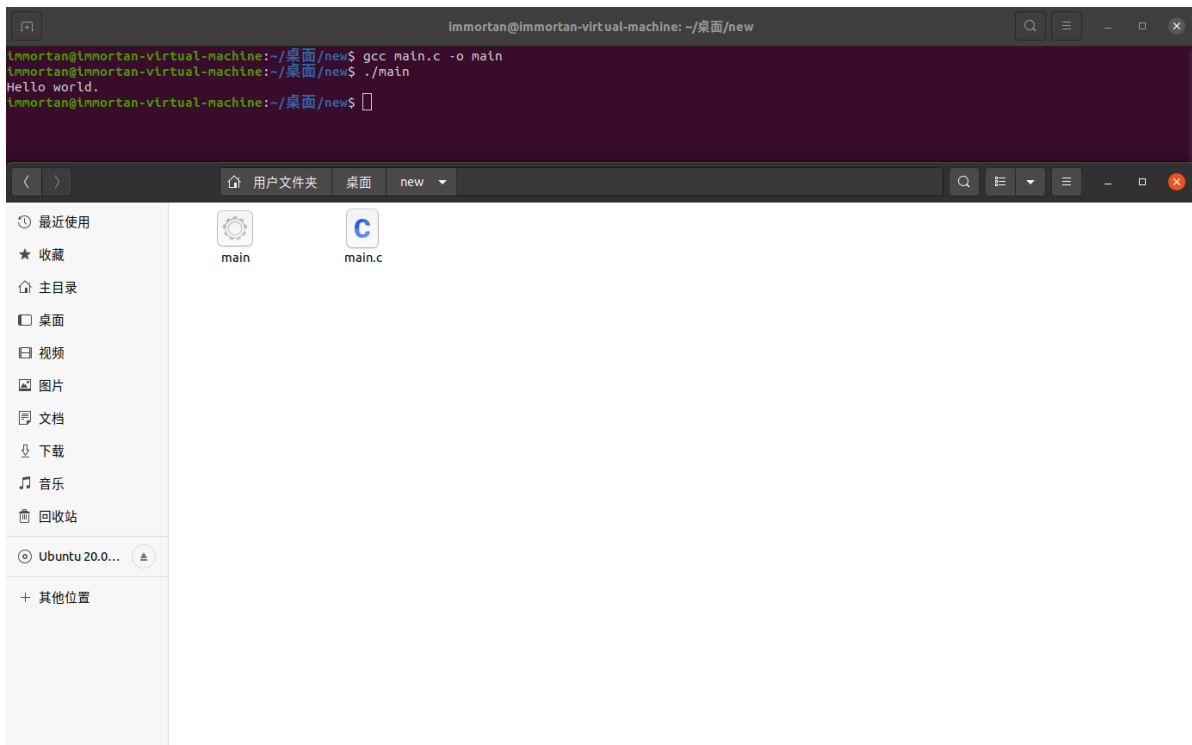
```
Immortan@Immortan-virtual-machine: ~/桌面
Immortan@Immortan-virtual-machine:~/桌面$ gcc -v
Using built-in specs.
COLLECT_GCC=gcc
COLLECT_LTO_WRAPPER=/usr/lib/gcc/x86_64-linux-gnu/9/lto-wrapper
OFFLOAD_TARGET_NAMES=nvptx-none:hsa
OFFLOAD_TARGET_DEFAULT=1
Target: x86_64-linux-gnu
Configured with: ../src/configure -v --with-pkgversion='Ubuntu 9.4.0-1ubuntu1-20.04' --with-bugurl=file:///usr/share/doc/gcc-9/README.Bugs --enable-lang
uages=c,ada,c++,go,brig,d,fortran,objc,obj-c++,gm2 --prefix=/usr --with-gcc-major-version-only --program-suffix=-9 --program-prefix=x86_64-linux-gnu- --
enable-shared --enable-linker-build-id --libexecdir=/usr/lib --without-included-gettext --enable-threads=posix --libdir=/usr/lib --enable-nls --enable-c
locale=gnu --enable-libstdcxx-debug --enable-libstdcxx-time=yes --with-default-libstdcxx-abi=new --enable-gnu-unique-object --disable-vtable-verify --en
able-plugin --enable-default-pie --with-system-zlib --with-target-system-zlib=auto --enable-objc-gc=auto --enable-multiarch --disable-werror --with-arch
-x32=686 --with-abi=m64 --with-multilib-list=m32,m64,mx32 --enable-multilib --with-tune=generic --enable-offload-targets=nvptx-none=/build/gcc-9-yTrUTS/
gcc-9-9.4.0/debian/tmp-nvptx/usr,hsa --without-cuda-driver --enable-checking=release --build=x86_64-linux-gnu --host=x86_64-linux-gnu --target=x86_64-li
nux-gnu
Thread model: posix
gcc version 9.4.0 (Ubuntu 9.4.0-1ubuntu1-20.04)
Immortan@Immortan-virtual-machine:~/桌面$ gcc --version
gcc (Ubuntu 9.4.0-1ubuntu1-20.04) 9.4.0
Copyright (C) 2019 Free Software Foundation, Inc.
This is free software; see the source for copying conditions. There is NO
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

Immortan@Immortan-virtual-machine:~/桌面$
```

- 编译单个文件

```
#include<stdio.h>

int main() {
    printf("Hello world.\n");
    return 0;
}
```



- 编译链接多个文件
 - main.c

```
#include<stdio.h>

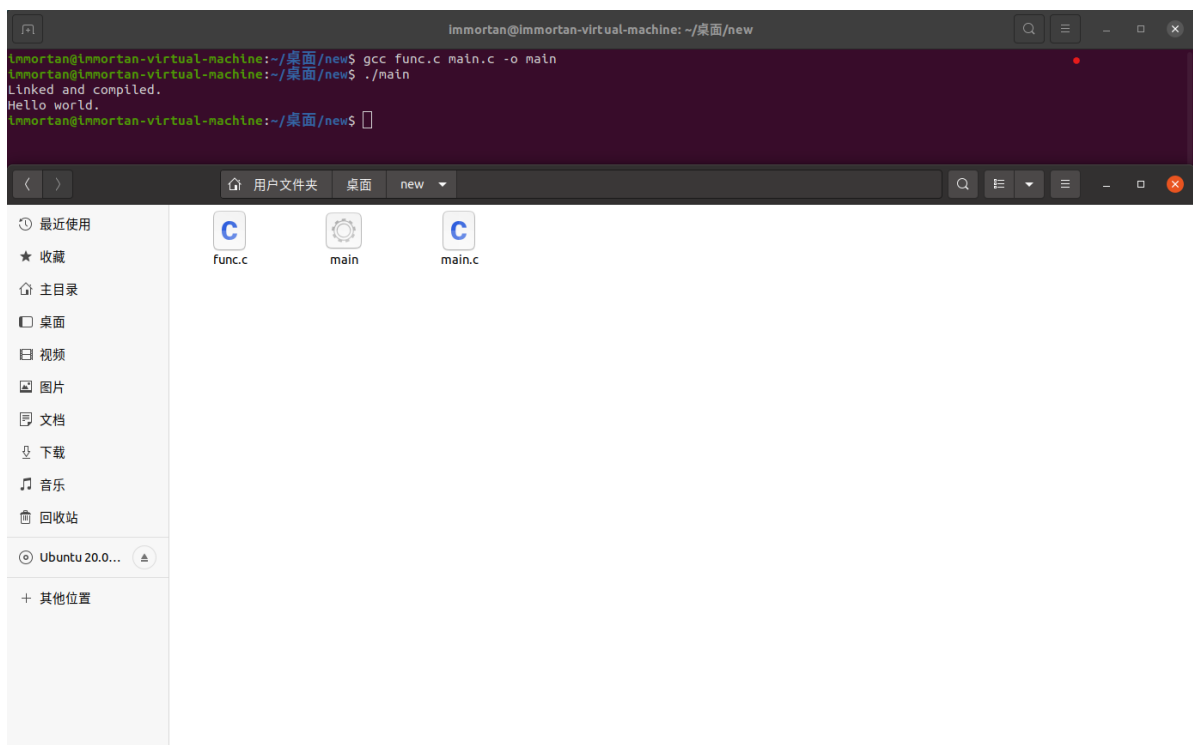
extern void func();

int main() {
    func();
    printf("Hello world.\n");
    return 0;
}
```

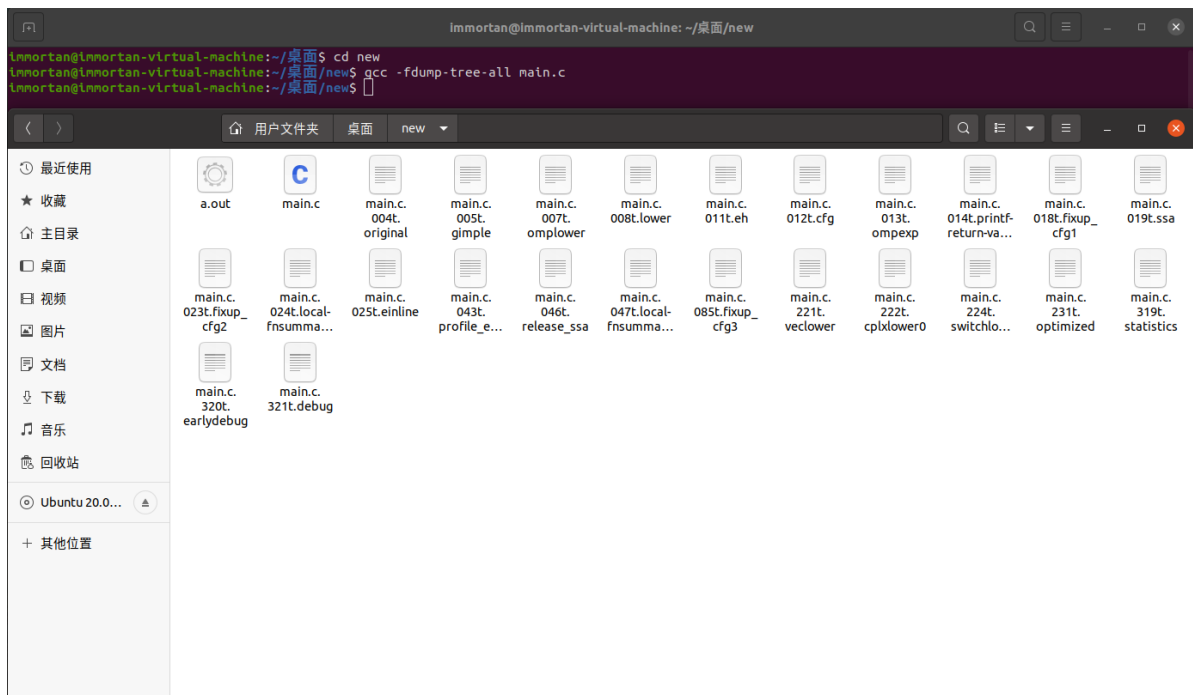
o func.c

```
#include<stdio.h>

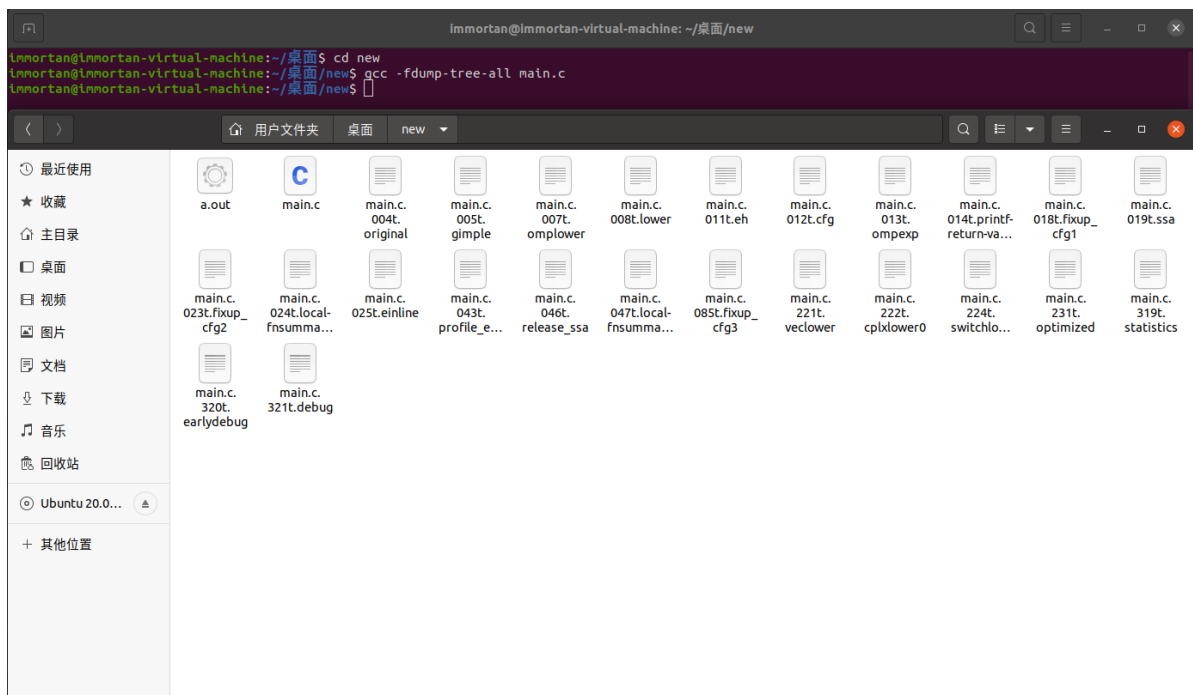
void func() {
    printf("Linked and compiled.\n");
    return;
}
```



- gcc -E main.c -o main.i
生成预处理文件，其中包含头文件的展开。



- `gcc -fdump-rtl-all main.c`
生成中间代码。在gcc中，中间代码用指令序列表示。



- `gcc -S main.c -o main.s`
生成汇编代码

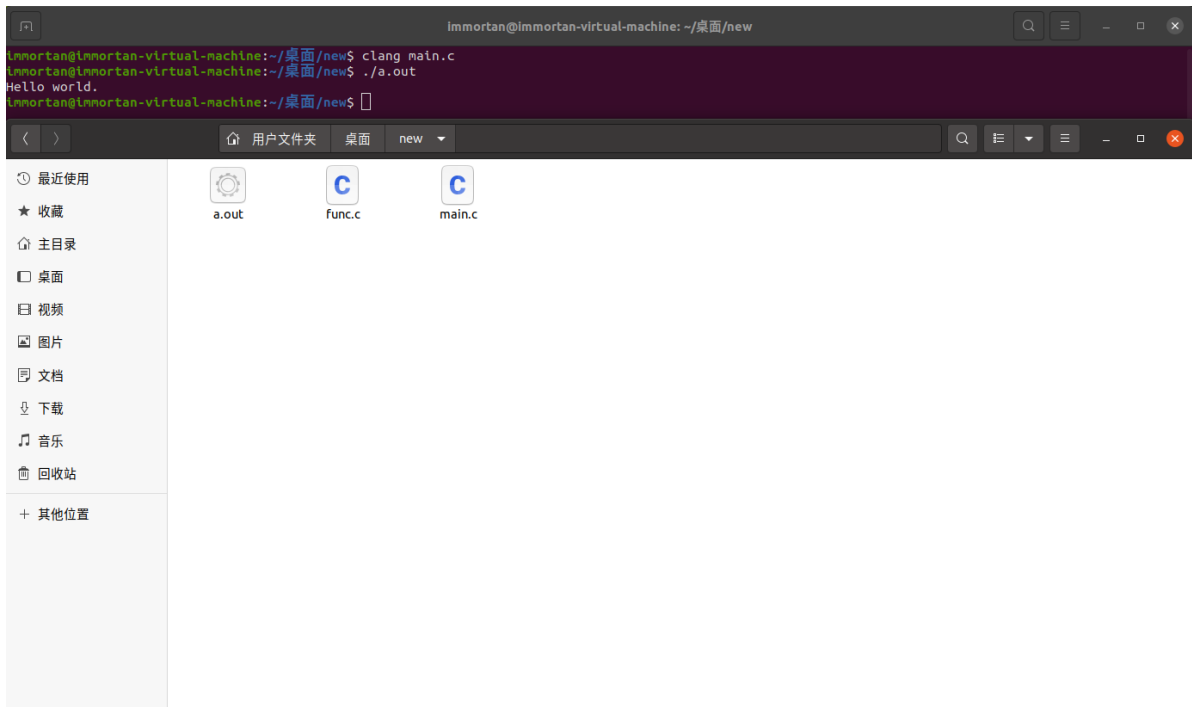
```
1 .file "main.c"
2 .text
3 .section .rodata
4 .LC0:
5 .string "Hello world."
6 .text
7 .globl main
8 .type main, @function
9 main:
10 .LFB0:
11 .cfi_startproc
12 endbr4
13 pushq %rbp
14 .cfi_def_cfa_offset 16
15 .cfi_offset 0, -16
16 movq %rsp, %rbp
17 .cfi_def_cfa_register 6
18 leaq .LC0(%rip), %rdi
19 call printf@plt
20 movl $0, %eax
21 popq %rbp
22 .cfi_def_cfa 7, 8
23 ret
24 .cfi_endproc
25 .LFE0:
26 .size main, .-main
27 .ident "GCC: (Ubuntu 9.4.0-1ubuntu1-20.04) 9.4.0"
28 .section .note.GNU-stack,"",@progbits
29 .section .note.gnu.property,"",@progbits
30 .align 8
31 .long 1f - 0f
32 .long 4f - 3f
33 .long 5
34 0:
35 .string "GNU"
36 1:
37 .align 8
38 .long 0xc0000002
39 .long 3f - 2f
40 2:
41 .long 0x3
42 3:
43 .align 8
44 4:
```

LLVM运行结果分析

- clang -v
获取clang版本信息和配置信息。

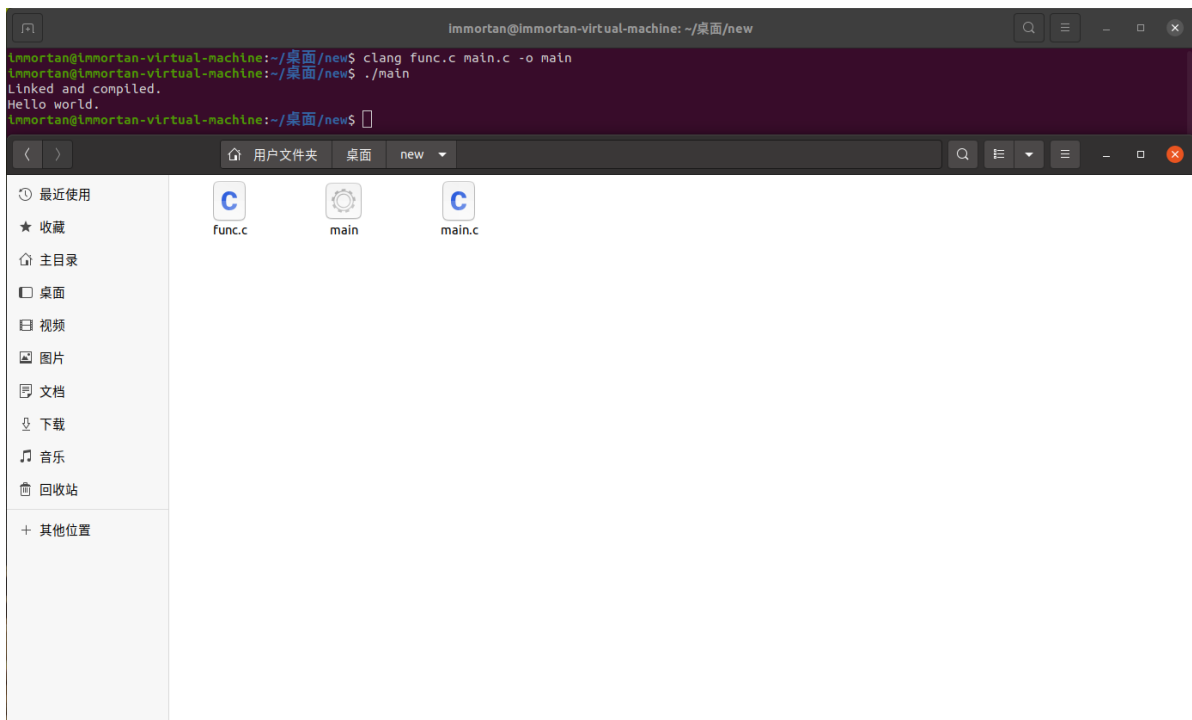
```
immortan@immortan-virtual-machine: ~/桌面/new
immortan@immortan-virtual-machine:~/桌面/new$ clang -v
clang version 10.0.0-4ubuntu1
Target: x86_64-pc-linux-gnu
Thread model: posix
InstalledDir: /usr/bin
Found candidate GCC installation: /usr/bin/../lib/gcc/x86_64-linux-gnu/9
Found candidate GCC installation: /usr/lib/gcc/x86_64-linux-gnu/9
Selected GCC installation: /usr/bin/../lib/gcc/x86_64-linux-gnu/9
Candidate multilib: .;@m64
Selected multilib: .;@m64
immortan@immortan-virtual-machine:~/桌面/new$
```

- 编译单个文件
源码同gcc



- 编译链接多个文件

源码同gcc



- clang -ccc-print-phases main.c -c

```
immortan@immortan-virtual-machine: ~/桌面/new
immortan@immortan-virtual-machine:~/桌面/new$ clang -ccc-print-phases main.c -c
+ 0: input, "main.c", c
+ 1: preprocessor, {0}, cpp-output
+ 2: compiler, {1}, ir
+ 3: backend, {2}, assembler
4: assembler, {3}, object
immortan@immortan-virtual-machine:~/桌面/new$
```

- clang main.c -Xclang -dump-tokens

```
immortan@immortan-virtual-machine: ~/桌面/new
+ 0: input, "main.c", c
+ 1: preprocessor, {0}, cpp-output
+ 2: compiler, {1}, ir
+ 3: backend, {2}, assembler
4: assembler, {3}, object
immortan@immortan-virtual-machine:~/桌面/new$ clang main.c -Xclang -dump-tokens
+ 0: input, "main.c", c
+ 1: preprocessor, {0}, cpp-output
+ 2: compiler, {1}, ir
+ 3: backend, {2}, assembler
4: assembler, {3}, object
immortan@immortan-virtual-machine:~/桌面/new$
```

- clang main.c -Xclang -dump-raw-tokens


```
renamable $rdi = MOV64r1 @.str
sal = MOV8r1 0
CALL64pcr13 @printf, <regmask $bh $bl $bp $bh $bp $bh $ebp $ebx $shp $rbp $rbx $r12 $r13 $r14 $r15 $r12b $r13b $r14b $r15b $r12bh $r13bh $r14bh $r15bh $r12d $r13d $r14d $r15d $r12w $r13w $r14w $r15w $r12hw and 3 more....>, Implicit $rsp, Implicit $ssp, Implicit killed sal, Implicit killed $rdi, Implicit-def $eax
renamable $ecx = XOR32rr under $ecx(tied-def 0), under $ecx, Implicit-def $eflags
MOV32w $rbp, 1, $noreg, -4, $noreg, killed $eax :: (store 4 into %r1)
$eax = MOV32rr killed $ecx
$rsp = frame-destroy ADD64r18 $rsp(tied-def 0), 16, Implicit-def dead $eflags
$rbp = frame-destroy POP64r Implicit-def $rsp, Implicit $rsp
CFI_INSTRUCTION def_cfa $rsp, 8
RETIQ Implicit killed $eax

# End machine code for function main.

# *** IR Dump After Live DEBUG_VALUE analysis ***:
# Machine code for function main: NoPMIs, TracksLiveness, NoVRegs
Frame Objects:
fl#-1: size=8, align=16, fixed, at location [SP-8]
fl#0: size=4, align=4, at location [SP-12]
fl#1: size=4, align=4, at location [SP-16]

bb.0 (Nir-Block.0):
frame-setup PUSH64r Killed $rbp, Implicit-def $rsp, Implicit $rsp
CFI_INSTRUCTION def_cfa_offset 16
CFI_INSTRUCTION offset $rbp, 16
$rbp = frame-setup MOV64rr $rsp
CFI_INSTRUCTION def_cfa_register $rbp
$rsp = frame-setup SUB64r18 $rsp(tied-def 0), 16, Implicit-def dead $eflags
MOV32w $rbp, 1, $noreg, -4, $noreg, 0 :: (store 4 into %r1)
renamable $rdi = MOV64r1 @.str
sal = MOV8r1 0
CALL64pcr13 @printf, <regmask $bh $bl $bp $bh $bp $bh $ebp $ebx $shp $rbp $rbx $r12 $r13 $r14 $r15 $r12b $r13b $r14b $r15b $r12bh $r13bh $r14bh $r15bh $r12d $r13d $r14d $r15d $r12w $r13w $r14w $r15w $r12hw and 3 more....>, Implicit $rsp, Implicit $ssp, Implicit killed sal, Implicit killed $rdi, Implicit-def $eax
renamable $ecx = XOR32rr under $ecx(tied-def 0), under $ecx, Implicit-def $eflags
MOV32w $rbp, 1, $noreg, -8, $noreg, killed $eax :: (store 4 into %stack.1)
$eax = MOV32rr killed $ecx
$rsp = frame-destroy ADD64r18 $rsp(tied-def 0), 16, Implicit-def dead $eflags
$rbp = frame-destroy POP64r Implicit-def $rsp, Implicit $rsp
CFI_INSTRUCTION def_cfa $rsp, 8
RETIQ Implicit killed $eax

# End machine code for function main.

# *** IR Dump After Check CFA Info and Insert CFI Instructions If needed ***:
# Machine code for function main: NoPMIs, TracksLiveness, NoVRegs
Frame Objects:
fl#-1: size=8, align=16, fixed, at location [SP-8]
fl#0: size=4, align=4, at location [SP-12]
fl#1: size=4, align=4, at location [SP-16]

bb.0 (Nir-Block.0):
frame-setup PUSH64r Killed $rbp, Implicit-def $rsp, Implicit $rsp
CFI_INSTRUCTION def_cfa_offset 16
CFI_INSTRUCTION offset $rbp, 16
$rbp = frame-setup MOV64rr $rsp
CFI_INSTRUCTION def_cfa_register $rbp
$rsp = frame-setup SUB64r18 $rsp(tied-def 0), 16, Implicit-def dead $eflags
MOV32w $rbp, 1, $noreg, -4, $noreg, 0 :: (store 4 into %r1)
renamable $rdi = MOV64r1 @.str
sal = MOV8r1 0
CALL64pcr13 @printf, <regmask $bh $bl $bp $bh $bp $bh $ebp $ebx $shp $rbp $rbx $r12 $r13 $r14 $r15 $r12b $r13b $r14b $r15b $r12bh $r13bh $r14bh $r15bh $r12d $r13d $r14d $r15d $r12w $r13w $r14w $r15w $r12hw and 3 more....>, Implicit $rsp, Implicit $ssp, Implicit killed sal, Implicit killed $rdi, Implicit-def $eax
renamable $ecx = XOR32rr under $ecx(tied-def 0), under $ecx, Implicit-def $eflags
MOV32w $rbp, 1, $noreg, -4, $noreg, killed $eax :: (store 4 into %stack.1)
$eax = MOV32rr killed $ecx
$rsp = frame-destroy ADD64r18 $rsp(tied-def 0), 16, Implicit-def dead $eflags
$rbp = frame-destroy POP64r Implicit-def $rsp, Implicit $rsp
CFI_INSTRUCTION def_cfa $rsp, 8
RETIQ Implicit killed $eax

# End machine code for function main.
```

- clang main.c -S

```
1 .text
2 .file "main.c"
3 @.labl main # -- Begin Function main
4 .p2align 4, 0x00
5 .type main,@function # @main
6 main:
7 .cfi_startproc
8 # bb.0:
9 pushq %rbp
10 .cfi_def_cfa_offset 16
11 .cfi_offset $rbp, 16
12 movq %rbp, %rsp
13 .cfi_def_cfa_register %rbp
14 subq $16, %rsp
15 movl $4, 4(%rsp)
16 movabsq $1, %rax
17 movl %rax, %eax
18 callq printf
19 xorl %ecx, %ecx
20 movl %ecx, %eax # 4-byte Spill
21 movl %ecx, %eax
22 addq $16, %rsp
23 popq %rbp
24 .cfi_def_cfa $rsp, 8
25 retq
26 .Lfunc_end0:
27 .size main, .Lfunc_end0-main
28 .cfi_endproc
29 .type .Lstr,@object # 8.str
30 .section .rodata.str.1,"aM",@progbits,1
31 .L.str:
32 .asciz "hello world!\n"
33 .size .Lstr, 14
34 .ident "clang version 18.0.8-dubnui"
35 .section ".note.GNU-stack","",@progbits
36 .addrsig
37 .addrsig sym printf
38
```

实验心得体会

初步了解了GCC和LLVM的编译过程，对词法分析、语法分析、中间代码和汇编有了具体的认识。