Béatrice Creusillet

bcreusillet@quarkslab.com> Philippe Virouleau <pvirouleau@quarkslab.com>

Compilation pour le reverseur

Table of Contents

Avant de rentrer dans le vif du sujet

Introduction générale

Ma première passe

Pré-requis

- Linux avec gcc, clang, ida (version gratuite)
- ou Machine virtuelle:
 - Vagrant
 - VirtualBox
 - Fichier Vagrantfile envoyé précédemment
 - vagrant up
 - quitter la VM
 - vagrant up
 - ► Si ida n'est pas chargé: ./idafree70_linux.run --mode unattended
 - ► En cas de pb de clavier, dans un terminal de la VM sudo dpkg-reconfigure keyboard-configuration
- ► TDs: git clone git@github.com:bac-qb/compil.git
- vagrant upload compil/codes

Table of Contents

Avant de rentrer dans le vif du sujet

Introduction générale

Compilation pour le reverseur, vraiment ?

Compilation vs. interprétation

Compilation statique : du code source au binaire

Options de compilation

Différents compilateurs, différents binaires

Un peu de mécanique : les internes d'un compilateur

Ma première passe

Table of Contents

Avant de rentrer dans le vif du sujet

Introduction générale

Compilation pour le reverseur, vraiment ?

Compilation vs. interprétation

Compilation statique : du code source au binaire

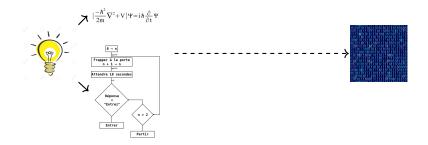
Options de compilation

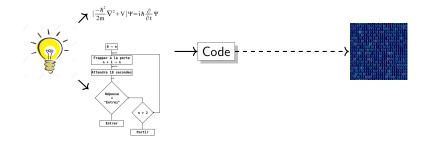
Différents compilateurs, différents binaires

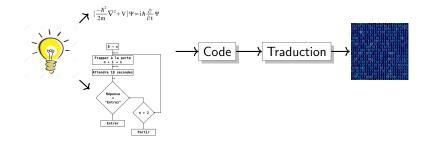
Un peu de mécanique : les internes d'un compilateur

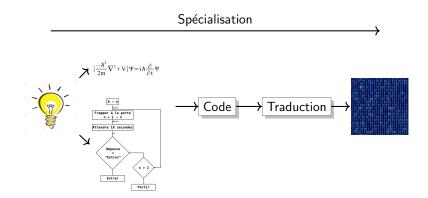
Ma première passe











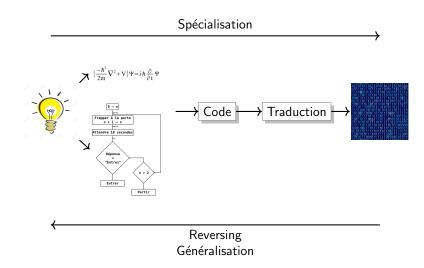


Table of Contents

Avant de rentrer dans le vif du sujet

Introduction générale

Compilation pour le reverseur, vraiment ?

Compilation vs. interprétation

Compilation statique : du code source au binaire

Options de compilation

Différents compilateurs, différents binaires

Un peu de mécanique : les internes d'un compilateur

Ma première passe

- ► Compilation : traduction statique en binaire
 - réutilisable
 - pour des inputs variés
 - adapté à un environnement particulier (windows, linux, osx,...)
 - ▶ ex: C, C++, Fortran

- ► Compilation : traduction statique en binaire
 - réutilisable
 - pour des inputs variés
 - ▶ adapté à un environnement particulier (windows, linux, osx,...)
 - ▶ ex: C, C++, Fortran

- ▶ Interprétation : traduction dynamique = à la volée
 - non-réutilisable
 - optimisée pour des inputs particuliers
 - ex: Basic, R, bash

- ► Compilation : traduction statique en binaire
 - réutilisable
 - pour des inputs variés
 - adapté à un environnement particulier (windows, linux, osx,...)
 - ▶ ex: C, C++, Fortran
- Pré-compilation : traduction statique vers un 'byte-code'
 - byte-code générique
 - adaptation à l'environnement au run-time
 - ex: Java

- ▶ Interprétation : traduction dynamique = à la volée
 - non-réutilisable
 - optimisée pour des inputs particuliers
 - ex: Basic, R, bash

- ► Compilation : traduction statique en binaire
 - réutilisable
 - pour des inputs variés
 - ▶ adapté à un environnement particulier (windows, linux, osx,...)
 - ▶ ex: C, C++, Fortran
- Pré-compilation : traduction statique vers un 'byte-code'
 - byte-code générique
 - adaptation à l'environnement au run-time
 - ex: Java
- ▶ JIT : Just-in-Time compilation
 - traduction dynamique partiellement 'réutilisable'
 - parfois limitée aux hot spots
 - ex: Matlab, Julia, Java
- ▶ Interprétation : traduction dynamique = à la volée
 - non-réutilisable
 - optimisée pour des inputs particuliers
 - ex: Basic, R, bash

Compilateurs C/C++

Compilateurs commerciaux

- ► Intel C/C++ compiler (icc) (Windows, linux, MacOs, Intel)
- ► Visual Studio C++ (Windows, Microsoft)
- ► IAR C/C++ compiler (Windows, Linux, others, IAR Systems)
- Edison Design Group (Windows, Linux, others, EDG)
- VisualAge C++, XL C/C++ (Windows, LInux, Aix, OS/2, OS/400,... IBM)

OpenSource ou Freeware

- ► GNU C/C++ (gcc, MinGW) (Windows, linux, MacOS, GNU Project)
- Clang/LLVM (clang) (Windows, linux, MacOS, LLVM Project)
- **...**

Table of Contents

Avant de rentrer dans le vif du sujet

Introduction générale

Compilation pour le reverseur, vraiment ? Compilation vs. interprétation

Compilation statique : du code source au binaire

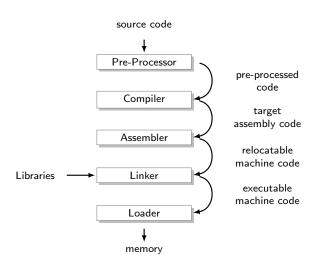
Options de compilation

Différents compilateurs, différents binaires

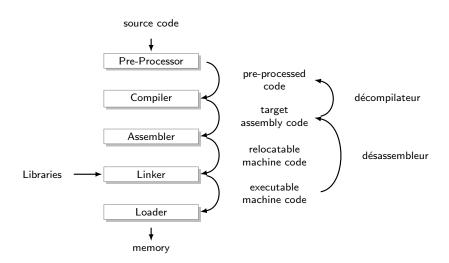
Un peu de mécanique : les internes d'un compilateur

Ma première passe

Une traduction en plusieurs étapes



Une traduction en plusieurs étapes



TD 1 : les étapes de la génération de binaire

welcome.c:

```
#include <stdio.h>
void welcome(char const* nom) {
printf("welcome %s!\n", nom);
}
```

main_welcome.c :

```
#ifindef DEFAULT
#define DEFAULT "everybody"
#endif

extern void welcome(char const*);

int main(int argc, char const* argv[]) {
  if (argc == 1) { welcome(DEFAULT); }
  else if (argc == 2) { welcome(argv[1]); }
  else { return 1; }
  return 0;
}
```

Version par défaut :

```
clang welcome.c main_welcome.c
// ./a.out
// ./a.out alice
```

Version par défaut :

```
clang welcome.c main_welcome.c
   ./a.out
   ./a.out alice
```

```
welcome everybody!
welcome alice!
```

Version par défaut :

```
clang welcome.c main_welcome.c
   ./a.out
   ./a.out alice
```

```
welcome everybody!
welcome alice!
```

Version avec définition de variable de pré-processing :

```
clang welcome.c main_welcome.c -DDEFAULT='"bob"'
./a.out
./a.out alice
```

Version par défaut :

```
clang welcome.c main_welcome.c

./a.out

./a.out alice
```

```
welcome everybody!
welcome alice!
```

Version avec définition de variable de pré-processing :

```
clang welcome.c main_welcome.c -DDEFAULT='"bob"'
./a.out
./a.out alice
```

```
welcome bob!
welcome alice!
```

clang -E main_welcome.c

```
1 clang -E main_welcome.c
```

```
int main(int argc, char const* argv[]) {
  if (argc == 1) { welcome("everybody"); }
  else if (argc == 2) { welcome(argv[1]); }
  else { return 1; }
  return 0;
}
```

```
1 clang -E main_welcome.c
```

```
int main(int argc, char const* argv[]) {
  if (argc == 1) { welcome("everybody"); }
  else if (argc == 2) { welcome(argv[1]); }
  else { return 1; }
  return 0;
}
```

```
clang -E main_welcome.c -DDEFAULT='"bob"'
```

```
1 clang -E main_welcome.c
```

```
int main(int argc, char const* argv[]) {
   if (argc == 1) { welcome("everybody"); }
   else if (argc == 2) { welcome(argv[1]); }
   else { return 1; }
   return 0;
}
```

```
clang -E main_welcome.c -DDEFAULT='"bob"'
```

```
int main(int argc, char const* argv[]) {
   if (argc == 1) { welcome("bob"); }
   else if (argc == 2) { welcome(argv[1]); }
   else { return 1; }
   return 0;
}
```

Le pré-processeur : inclusion de fichiers header

```
1 clang -E welcome.c | wc -l
```

```
1 797
```

Le pré-processeur : inclusion de fichiers header

```
1 clang -E welcome.c | wc -l
```

```
1 797
```

```
1 clang -E welcome.c
```

Le pré-processeur : inclusion de fichiers header

```
1 clang -E welcome.c | wc -l
```

```
1 797
```

```
1 clang -E welcome.c
```

- les fichiers sont intégralement inclus, récursivement
- attention aux include redondants

```
#ifndef MY_HEADER_H_
#define MY_HEADER_H_
... // header content
#endif // MY_HEADER_H_
```

attention à l'expansion de macros !

Le compilateur : génération de code assembleur

- 1 clang -S main_welcome.c
- more main_welcome.s

L'assemblage: génération de code objet

```
1 as main_welcome.s -o main_welcome.o
2 file main_welcome.o
3 nm main_welcome.o
```

L'assemblage: génération de code objet

```
1 as main_welcome.s -o main_welcome.o
2 file main_welcome.o
3 nm main_welcome.o
```

L'assemblage: génération de code objet

```
1 as main_welcome.s -o main_welcome.o
2 file main_welcome.o
3 nm main_welcome.o
```

```
1 clang -c welcome.c -o welcome.o
2 file welcome.o
3 nm welcome.o
```

L'assemblage: génération de code objet

```
1 as main_welcome.s -o main_welcome.o
2 file main_welcome.o
3 nm main_welcome.o
```

```
main_welcome.o: ELF 64-bit LSB relocatable, x86-64,
    version 1 (SYSV), not stripped

U _GLOBAL_OFFSET_TABLE_

O000000000000000 T main
U welcome
```

```
1 clang -c welcome.c -o welcome.o
2 file welcome.o
3 nm welcome.o
```

```
velcome.o: ELF 64-bit LSB relocatable, x86-64, version
1 (SYSV), not stripped

U _GLOBAL_OFFSET_TABLE_
U printf
0000000000000000 T welcome
```

ld welcome.o main_welcome.o

```
1 ld welcome.o main_welcome.o
```

ld welcome.o main_welcome.o

```
1 ld welcome.o main_welcome.o -lc
2 file a.out
```

ld welcome.o main_welcome.o

```
1 ld welcome.o main_welcome.o -lc
2 file a.out
```

clang -v welcome.o main_welcome.o 2>&1 | grep ld

```
clang -v welcome.o main_welcome.o 2>&1 | grep ld
```

```
"/usr/bin/ld" -z relro --hash-style=gnu --eh-frame-hdr
    -m elf_x86_64 -dynamic-linker /lib64/ld-linux-x86
   -64.so.2 -o a.out /usr/bin/../lib/gcc/x86_64-linux
   -gnu/7.3.0/../../x86_64-linux-gnu/crt1.o /usr/
   bin/../lib/gcc/x86_64-linux-gnu/7.3.0/../../
   x86_64-linux-gnu/crti.o /usr/bin/../lib/gcc/x86_64
   -linux-gnu/7.3.0/crtbegin.o -L/usr/bin/../lib/gcc/
   x86_64-linux-gnu/7.3.0 -L/usr/bin/../lib/gcc/
   x86_64-linux-gnu/7.3.0/../../x86_64-linux-gnu -
   L/lib/x86_64-linux-gnu -L/lib/../lib64 -L/usr/lib/
   x86_64-linux-gnu -L/usr/bin/../lib/gcc/x86_64-
   linux-gnu/7.3.0/../.. -L/usr/lib/llvm-6.0/bin
   /../lib -L/lib -L/usr/lib welcome.o main_welcome.o
    -lgcc --as-needed -lgcc_s --no-as-needed -lc -
   lgcc --as-needed -lgcc_s --no-as-needed /usr/bin
   /../lib/gcc/x86_64-linux-gnu/7.3.0/crtend.o /usr/
   bin/../lib/gcc/x86_64-linux-gnu/7.3.0/../../
   x86_64-linux-gnu/crtn.o
```

Table of Contents

Avant de rentrer dans le vif du sujet

Introduction générale

Compilation pour le reverseur, vraiment ?

Compilation vs. interprétation

Compilation statique : du code source au binaire

Options de compilation

Différents compilateurs, différents binaires

Un peu de mécanique : les internes d'un compilateur

Ma première passe

Piloter son compilateur : les options

Contrôle des inputs/outputs

- fichiers sources : paramètres positionnels
- bibliothèques : -1<libname>
- chemins d'accès
 - fichiers headers : -I<path>
 - bibliothèques : -L<path>
- standard du language : -std=c++11 par ex.
- warnings : -W<warning> (ex: -Wall, -Werror)
- ▶ infos: --help, --version, -print-search-dirs, ...

Passer des options aux différentes étapes

- pré-processeur: -Wp,<option>
- assembleur: -Wa,<option>
- ▶ linker: -W1,<option>

Piloter l'optimiseur

Liste des optimisations :

▶ gcc/g++:

```
g++ --help=optimizers
```

clang/llvm:

```
clang -help
opt -help
```

Optimisations globales : -0<level>

- -00 : pas d'optimisations, pratique pour le debug
- ► -03 : plus fort niveau d'optimisations
- ► -0s : -02, mais minimise la taille du code
- -Ofast: -03 + -ffastmath

TD 2 : Piloter l'optimiseur

```
#ifdef BIG N
  #define n 100
3 #else
  #define n 5
  #endif
  int main() {
7
  int a[n];
   for (int i = 0; i < n; ++i) { a[i] = i+1; }
10
  int s = 0:
  for (int i = 0; i < n; ++i) { s +=a [i]: }
12
13
  if (s==(n*(n+1))/2) { return s; }
15 return -1;
16 }
```

- avec #define n 5
 Comparer les assembleurs générés avec -00 et -02
- avec #define n 100
 Comparer les binaires générés avec -00 et -03

TD 3 : les options pour le debug (1)

```
#include <assert.h>
#define MAX_SIZE 512

int count_char(const char* str) {
   int c = 0;
   while ( (c < MAX_SIZE) && (str[c] != '\0')) { ++c; }
   return c;
}

int main(int argc, char** argv) {
   assert(argc == 2 && "compliant number of arguments");
   return count_char(argv[1]);
}</pre>
```

```
clang debug.c -o debug
//debug toto
cho $?
//debug
```

TD 3 : les options pour le debug (2)

L'impact de -DNDEBUG

```
clang debug.c -o debug-n -DNDEBUG
./debug-n toto
cecho $?
./debug-n
```

Comparer les binaires générés avec et sans -DNDEBUG

L'impact de -g

```
1 clang debug.c -o debug-ng -DNDEBUG -g
```

Comparer les binaires générés avec et sans -g

Table of Contents

Avant de rentrer dans le vif du sujet

Introduction générale

Compilation pour le reverseur, vraiment ?

Compilation vs. interprétation

Compilation statique : du code source au binaire

Options de compilation

Différents compilateurs, différents binaires

Un peu de mécanique : les internes d'un compilateur

Ma première passe

TD 4: clang vs. gcc

Reprendre le programme array.c du TD 2

Comparer les assembleurs générés par gcc et clang pour n = 100 et différents niveaux d'option -0<1evel> :

td4/gcc_vs_clang.sh :

```
LEVEL=$1

# assembleur

clang array.c -DBIG_N -O$LEVEL -S -o
array_clang_O$LEVEL.s

ccarray.c -DBIG_N -O$LEVEL -S -o array_gcc_O$LEVEL.s

binaire

clang array.c -DBIG_N -O$LEVEL -o array_clang_O$LEVEL

ccarray.c -DBIG_N -O$LEVEL -o array_clang_O$LEVEL

ccarray.c -DBIG_N -O$LEVEL -o array_gcc_O$LEVEL
```

```
bash gcc_vs_clang.sh 1
bash gcc_vs_clang.sh 2
bash gcc_vs_clang.sh 3
```

Comparer aussi les versions générées par clang avec -02 et gcc avec -03.

Jouer avec les compilateurs

```
https://godbolt.org/
Outil permettant de visionner les assembleurs générés par différents compilateurs pour différentes platformes (x86_64, armxx,...) (développé par Matt Godbolt)
https://cppinsights.io/
Outil basé sur Clang, qui permet de voir comment le compilateur transforme le code (développé par Andreas Fertig)
```

Table of Contents

Avant de rentrer dans le vif du sujet

Introduction générale

Compilation pour le reverseur, vraiment ?

Compilation vs. interprétation

Compilation statique : du code source au binaire

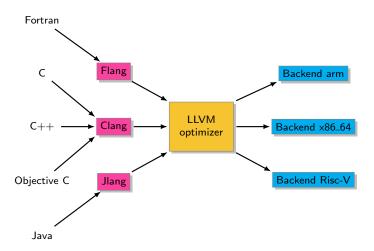
Options de compilation

Différents compilateurs, différents binaires

Un peu de mécanique : les internes d'un compilateur

Ma première passe

Clang/LLVM



L'IR LLVM

- Pseudo assembleur
- ► Nombre illimité de registres virtuels = *Values*
- Assignation unique des registres virtuels

```
#ifindef DEFAULT
2 #define DEFAULT "everybody"
3 #endif
4
5 extern void welcome(char const*);
6
7 int main(int argc, char const* argv[]) {
8    if         (argc == 1) { welcome(DEFAULT); }
9    else if (argc == 2) { welcome(argv[1]); }
10    else { return 1; }
11    return 0;
12 }
```

```
1 clang main_welcome.c -S -emit-llvm -o -
```

```
define i32 @main(i32. i8**) #0 {
     %3 = alloca i32, align 4
     %4 = alloca i32, align 4
     %5 = alloca i8**, align 8
      store i32 0, i32* %3, align 4
      store i32 %0, i32* %4, align 4
     store i8** %1. i8*** %5. align 8
     \%6 = 10ad i32, i32* \%4, align 4
8
     %7 = icmp eq i32 %6, 1
9
     br i1 %7, label %8, label %9
10
11
12
   : <label>:8:
                                                  : preds = %2
      call void @welcome(i8* getelementptr inbounds ([10 x i8], [10
13
           x i8]* @.str, i32 0, i32 0))
     br label %18
14
15
16
    . . .
17
18
   : <label>:18:
                                                  ; preds = %17, %8
19
    store i32 0, i32* %3, align 4
     br label %19
20
   : <label>:19:
                                                  ; preds = %18, %16
22
     %20 = load i32, i32 * %3, align 4
23
     ret i32 %20
24
25
```

```
#ifdef BIG_N
2 #define n 100
3 #else
4 #define n 5
  #endif
  int main() {
7
  int a[n];
   for (int i = 0; i < n; ++i) { a[i] = i+1; }
10
  int s = 0;
11
  for (int i = 0; i < n; ++i) { s +=a [i]; }
12
13
  if (s==(n*(n+1))/2) { return s; }
14
15 return -1:
16 }
```

```
define i32 @main() local unnamed addr #0 {
2
      . . .
      br label %3
3
5
    : <label>:3:
                                                             ; preds = %3,
          %0
      %4 = phi i64 [ 0, %0 ], [ %5, %3 ]
7
      \%5 = add nuw nsw i64 \%4, 1
      \%6 = getelementptr inbounds [5 x i32], [5 x i32]* \%1, i64 0,
8
           i64 %4
9
      . . .
      br i1 %8, label %9, label %3
10
11
                                                             ; preds = %3
12
    ; <label >: 9:
      br label %13
13
14
                                                             ; preds = %13
    : <label >: 10:
15
16
      . . .
      ret i32 %12
17
18
    : <label>:13:
                                                             ; preds = %9,
19
          %13
      %14 = phi i64 [ %19, %13 ], [ 0, %9 ]
20
      %15 = phi i32 [ %18, %13 ], [ 0, %9 ]
21
      %16 = \text{getelementptr inbounds } [5 \times i32], [5 \times i32] * %1, i64 0,
22
            i64 %14
      %17 = load i32, i32* %16, align 4, !tbaa !2
23
24
      . . .
      br i1 %20. label %10. label %13
25
26
```





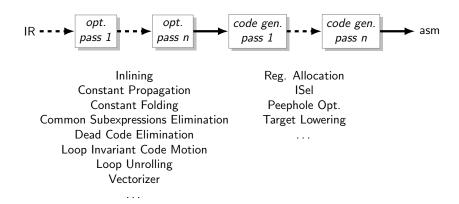
Inlining
Constant Propagation
Constant Folding
Common Subexpressions Elimination
Dead Code Elimination
Loop Invariant Code Motion
Loop Unrolling
Vectorizer

- - -



Inlining
Constant Propagation
Constant Folding
Common Subexpressions Elimination
Dead Code Elimination
Loop Invariant Code Motion
Loop Unrolling
Vectorizer

Reg. Allocation ISel Peephole Opt. Target Lowering



https://llvm.org/docs/Passes.html https://github.com/llvm-mirror/llvm/blob/master/lib/Transforms/IPO/ PassManagerBuilder.cpp

Table of Contents

Avant de rentrer dans le vif du sujet

Introduction générale

Ma première passe