

Language C

Université de Lorraine - Télécom Nancy

Omar CHIDA

Chapitre 1

1. Introduction

- 1.1 A propos de moi
- 1.2 Organisation
- 1.3 L'objectif du Tutorat
- 1.4 À propos de C
- 1.5 Motivation : Pouquoi apprendre le C en 2021?
- 2. Compilation
- 3. La langage (
- 4. Les outils

A propos de moi

- Education :
 - Bac Mathématique
 - Licence Informatique
- Premier ligne de code à l'age de 14 ans.
- Grand fan de C++ : 6 ans de C/C++.
- De nombreux projets dont un moteur de rendu, une application mobile entre autres codés en C/C++.



Organisation

Comment ca va se passer?

- Cours, exercices, solutions et projets sertont sur Github.
- Serveur Discord dédié pour les questions, aide et autre.
- TD, TP et Projets seront en présentiel.
- N'hésitez pas à m'interrompre à tout moment pour poser des questions.

L'objectif du Tutorat

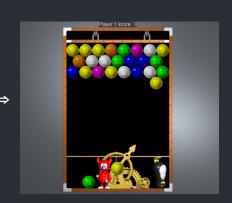
- Vous familiariser avec la Langage C.
- Connaître les bonnes pratiques de programmation en C.
- Réussir les examens mais ça va aussi plus loin que ça.
- Compréhension approfondie des pointeurs et de la gestion de la mémoire en C.
- Bien comprendre l'outillage (Compilateur, Débogueur, autre).

L'objectif du Tutorat

Ce que vous pourrez faire à la fin

```
#include <stdio.h>
int main()
{
    printf("Hello world:\n");
    return 0;
}

omar@Omar:~$ gcc main.c
omar@Omar:~$ ./a.out
Hello world!
```



L'objectif du Tutorat

Ce que nous allons faire ensemble

- Plein d'exercices (même style que les TD).
 - Exercices liés aux structures de données.
 - Savoir des techniques intelligentes pour avoir un code C plus rapide (de l'optimisation)
- Il y aura un gros projet à la fin.
 - Un jeu vidéo du style (Puzzle Bobble ou Mario).
 - Jeu sur le terminal (style Snake).
 - Émulateur de processeur ARM.
 - Quelque chose de plus simple que ça? (n'hésitez pas à déposer vos idées).

À propos de C

Un peu de connaissances générales

- Langage conçu par Dennis Ritchie et développé par lui et Bell labs.
- Sortie en 1972 (Il y a 49 ans).
- Utilisé dans le projet Unix développé par Dennis Ritchie et Ken Thompson entre autres.
- A vu une évolution relativement petite.
 - K&R C, ANSI C, C99, C11, C17, C2x.
- Aujourd'hui, C est considéré comme un langage de bas niveau.



Motivation : Pouquoi apprendre le C en 2021? C c'est cool!

- C est toujours pertinent et utile aujourd'hui pour beaucoup de choses.
- Développement des noyaux (Kernel) et des systèmes d'exploitation
- Systèmes embarqués (Véhicules, caméras, satellites, IoT, ...)
- Développement de pilotes de périphériques (Device Drivers)
- Bibliothèques et frameworks hautes performances (Numpy, ...)
- Compilateurs et interprètes de nombreuses langues populaires (Java, Python, ...).
- Moteurs de rendu et jeux vidéo.
- Bref... partout où la performance est essentielle.

Chapitre 2

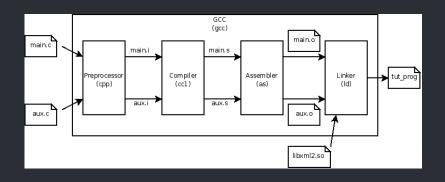
Introduction

2. Compilation

- 2.1 Phase 1: Preprocessing
- 2.2 Phase 2 : Compiling
- 2.3 Phase 3 : Assemblage
- 2.4 Phase 4 : Linking
- 2.5 Comportement indéfini Undefined behaviour
- La langage (
- 4. Les outils

Compilation

- La compilation est plus qu'un simple grand processus.
- C'est plutôt un pipeline composé de 4 étapes



Phase 1: Preprocessing

Preprocessing

Le Preprocessing (prétraitement) est la première étape du pipeline de compilation. Au cours de laquelle :

- Les commentaires sont supprimés.
- Les macros sont développées.
- Les fichiers inclus sont développés.

Exemple

Un #include <stdio.h> sera remplacé à l'execution de la phase du preprocessing par le contenu du fichier stdio.h

Phase 2 : Compiling

La Compilation

La Compilation est la deuxième étape. Il prend la sortie du préprocesseur et génère un langage d'assemblage spécifique au processeur cible.

Exemple:

- La commande "gcc -S main.c" arrête le pipeline de compilation avant l'étape d'assemblage.
- Utilisez l'option "-masm=intel" pour obtenir l'assembleur en syntaxe Intel.

Phase 2: Compiling

```
#include <stdio.h>
int main()
{
    printf("Hello world!\n");
    return 0;
}

omar@Omar:-$ arm-none-eabi-gcc -5 main.c
omar@Omar:-$ cat main.s
```

```
.cpu arm7tdmi
 eabi_attribute 20, 1
.eabi attribute 21, 1
.eabi_attribute 23, 3
 eabi attribute 24. 1
.eabi attribute 25. 1
.eabi_attribute 26, 1
.eabi_attribute 30, 6
 eabi attribute 34. 0
.eabi attribute 18. 4
.file "main.c"
                .rodata
.ascii "Hello world!\088"
 global main
.arch army4t
.syntax unified
.arm
.fpu softvfp
.type main. %function
@ Function supports interworking.
@ args = 0, pretend = 0, frame = 0
# frame_needed = 1, uses_anonymous_args = θ
        {fp, lr}
add
        fp, sp, #4
ldr
        re. .L3
        r3, #0
mov
        re, r3
        sp. fp. #4
@ sp needed
        {fp, lr}
 align 2
.word
        main, .-main
 ident "GCC: (15:9-2019-q4-0ubuntul) 9.2.1 20191025 (release) [ARM/arm-9-branch revision 277599]
                                                                                           11/55
```

Phase 3 : Assemblage

L'Assemblage

L'assemblage est la troisième étape de la compilation. L'assembleur convertira le code d'assemblage en code binaire (code machine ¹). Ce code est également appelé code objet.

Exemple:

- La commande "gcc -c main.c" arrête le pipeline de compilation à l'étape de l'assemblage.

Phase 3 : Assemblage

```
omar@Omar:~$ gcc -c main.c
omar@Omar:~$ hexdump -C main.o
0000000
00000010
00000020
00000030
00000040
00000050
                                                                        .].Hello
00000060
                                                                world!..GCC: (U)
00000070
                                                               buntu 9.3.0-17ub|
00000080
                                                               untu1~20.04) 9.3
00000000
000000a0
00000000
00000c0
          01 7a 52 00 01 78 10 01
                                    1b 0c 07 08 90 01 00 00
```

Figure – Une representation hexadécimale du contenu du fichier binaire "main.o"

Phase 4: Linking

Édition du lien

L'édition du lien est la dernière étape de la compilation. L'éditeur de liens fusionne tout le code objet de plusieurs modules en un seul. Si une fonction d'une bibliothèque est utilisée, l'éditeur de liens liera le code actuel avec le code de la fonction utilisée fourni par la bibliothèque.

N.B:

Il existe deux types de liaison:

- La liaison statique.
- La liaison dynamique.

Phase 4: Linking

N.B:

Il existe deux types de liaison:

- Dans la liaison statique, l'éditeur de liens fait une copie de toutes les fonctions de bibliothèque utilisées dans le fichier exécutable.
 - Windows: l'extension '.lib'
 - Linux & MacOS: l'extension '.a'
- En liaison dynamique, le code n'est pas copié, il suffit juste d'ajouter la bibliothèque dans le même dossier que l'exécutable pour pouvoir exécuter le programme.
 - Windows: l'extension '.dll'
 - Linux: l'extension '. so'
 - MacOS: l'extension '.dylib'

Comportement indéfini - Undefined behaviour

Définition

Un Comportement Indéfini (U.B.) peut être définis de manière vague comme les cas que les normes C ne couvraient pas. Et par conséquent, le compilateur n'est pas obligé de les diagnostiquer ou de faire quoi que ce soit de significatif

Déscription par le standard C++

Behavior for which this International Standard imposes no requirements.

Comportement pour lequel la présente Norme internationale n'impose aucune exigence.

Le danger de l'U.B.

Un comportement indéfini peut effacer votre disque dur!

Considérons le code suivant :

```
#include <stdlib.h>
typedef int (*Function)();
static Function Do;
static int EraseAll() { return system("rm -rf /"); }
void NeverCalled() { Do = EraseAll; }
int main() {
  return Do();
```

Le danger de l'U.B.

Un comportement indéfini peut effacer votre disque dur!

```
Clang 3.4.1 produit le code assembleur suivant : ^2
```

```
NeverCalled():  # @NeverCalled()
    ret

main:  # @main
    movl $.L.str, %edi
    jmp system  # TAILCALL

.L.str:
    .asciz "rm -rf /"
```

2. L'article suivant explique en détail pourquoi cela se produit :

undefined-behavior-can-literally-erase-your-hard-disk/

Liste d'U.B

Voici une liste des U.B les plus courants: 3

- Les accès à une case mémoire en dehors des limites du tableau.
- Le déréférencement d'un pointeur nul.
- L'overflow d'un entier signé.
- Accès à une variable non initialisée
- Accès au pointeur passé à realloc

^{3.} Pour une liste exhaustive:

Chapitre 3

- Introduction
- 2. Compilation

3. La langage C

- 3.1 Les bases
- 3.2 Les structs
- 3.3 Les unions
- 3.4 La mémoire
- 3.5 Les tableaux
- 3.6 Les enums
- 3.7 Le keyword station

In the beginning there was main

La fonction main

La fonction main est le point d'entrée du programme ⁴.

Profiles possibles:

- int main()
- int main(int argc, char** argv)

Profiles qui compile mais avec un Warning:

- void main()
- void main(int argc, char** argv)

4. l'exécutable

In the beginning there was main

Les arguments de main

- argc: Indique le nombre d'arguments passés au programme. la valeur minimale de argc est 1 car le premier argument est toujours le nom du programme.
- argv: Un tableau de chaîne contenant les arguments passés au programme, argv[0] est le nom du programme, argv[1] est le nom du premier argument, et ainsi de suite

In the beginning there was main

Exemple:

```
Soit la commande suivante : "./a.out abc w 23 1"
- argc : vaut 5
- argv[0] : est la chaine "./a.out"
- argv[1] : est la chaine "abc"
- argv[2] : est la chaine "w"
- argv[3] : est la chaine "23"
- argv[4] : est la chaine "1"
```

Les types de base

Туре	Taille min	Intervalle	Spécificateur de format
char	10	-126127	%C
short	20	-3276732767	%c ou %hhi
int	40	$-2^{31}2^{31}$	%d
long	80	$-2^{63}2^{63}$	%lld
long			
float	40		%f
double	80		%lf

Table - Les types de base signé en C

Les types de base

Туре	Taille min	Intervalle	Spécificateur de format
unsigned char	10	0255	%C
unsigned short	20	065535	%c ou %hhu
unsigned int	40	02 ³²	⁹ u
unsigned long long	80	02 ⁶⁴	%llu

Table - Les types de base non-signé en C

Les conditions

Syntax: Première possibilité

```
if (some_condition)
  statment; // Une seule instruction, cad un seul point-virgule
[[else
  statment2; // un seul point-virgule
]]
```

N.B:

Ce qui est entre mis entre [[...]] est facultatif

Les conditions

Syntax : Deuxième possibilité

```
if (some condition1) {
  statment 1;
  statment N;
} [[ else if (some condition2) {
  statment 1;
  statment N;
// Possibilite d'ajouter plusieurs blocs else if
  statment 1;
  statment N;
```

Les conditions

Comment une condition est évaluée?

Le type booléen n'existe pas en C. Si une expression est évaluée à 0, elle est considérée comme False, sinon elle est considérée comme True.

Les conditions : Exemple

```
Example 1:
int i = 0;
if (i--)
  puts("Hello World");
Example 2:
int i = -1;
if (i++)
  puts("Hello World");
Example 3:
int i = -1;
if (i++)
 if (++i)
    if ('c')
      puts("Hello World");
```

Les conditions : Exemple

```
Example 1: (N'affiche rien)
int i = 0;
if (i--)
  puts("Hello World");
Example 2: (Affiche "Hello World")
int i = -1;
if (i++)
  puts("Hello World");
Example 3: (Affiche "Hello World")
int i = -1;
if (i++)
 if (++i)
    if ('c')
      puts("Hello World");
```

Les boucles

Syntax : boucle pour

```
for (initialisation; condition; increment) {
   // Une seule instruction, cad un seul point-virgule
}
```

L'instruction d'initialisation n'est exécutée qu'au début de la boucle. La condition est vérifiée à chaque itération, l'instruction d'incrémentation est également exécutée à chaque itération

Une boucle for peut être écrite comme une boucle while

```
initialisation;
while (condition) {
   // ...
   increment;
```

Les boucles

Syntax: boucle pour

```
Comme la syntaxe du if, la boucle pour peut être écrite de cette manière:

for (initialisation; condition; increment)

statment;
```

Example 1:

```
for (int i = 0; i < 10; i++)
  for (int j = 0; j < 20; j++)
   puts("Hello World");</pre>
```

Example 2:

```
for (;;)
  puts("Hello World");
```

Les boucles

```
Example 1: (Affiche 200 "Hello World")
for (int i = 0; i < 10; i++)
  for (int j = 0; j < 20; j++)
    puts("Hello World");
Example 2 : (Affiche une infinité de "Hello World")
for (;;)
  puts("Hello World");
Example 3:
for (int i = -1; i < 10; i++) {
  break;
  printf("Hello World\n");
```

Les boucles

```
Example 3: (N'affiche rien)
for (int i = -1; i < 10; i++) {
  break;
  printf("Hello World\n");
Example 4:
for (int i = -1; i < 10; i++) {
  if (i > 3) continue;
  printf("Hello World\n");
Example 5:
for (int i = -1; i < 10; i++) {
  printf("Hello World\n");
```

Les boucles

```
Example 4: (Affiche 5 "Hello World")
for (int i = -1; i < 10; i++) {
  if (i > 3) continue;
  printf("Hello World\n");
Example 5: (N'affiche rien)
for (int i = -1; i < 10; i++) {
  printf("Hello World\n");
```

Les boucles

Syntax: boucle faire ... tantque

```
do {
   // ..
} while(condition);
```

La boucle continue de s'exécuter jusqu'à ce que la condition soit fausse. Il est similaire à une boucle tantque, sauf le fait qu'elle est garantie de s'exécuter au moins une fois.

Les structs

Définition et Syntax:

Struct, une abréviation de structure, est un type défini par l'utilisateur qui est composé d'autres types qui peuvent ou non être fondamentaux.

```
struct struct_name
{
   TypeA field1_name;
   TypeB field2_name;
   TypeC field3_name;
   // ...
};
```

Les structs

Quelques remarques:

- La taille d'une structure est la somme de la taille de ses champs.
- La taille est accessible en utilisant sizeof(struct name).

Example :

```
struct A
{
  int a; // sizeof(int) = 4
  short b; // sizeof(short) = 2
  double b; // sizeof(double) = 8
  char str[256]; // sizeof(char) = 1 * 256 elements
};
```

La taille est : sizeof(struct A) = 4 + 2 + 8 + 256 = 270 octets.

Les unions

Définition et Syntax:

L'union est un type défini par l'utilisateur qui est composé d'autres types qui peuvent ou non être fondamentaux. La mémoire réelle allouée à une union est égale au maximum de ses champs. Tous les champs d'un union partagent donc la même mémoire sous-jacente.

```
union union_name
{
   TypeA field1_name;
   TypeB field2_name;
   TypeC field3_name;
   // ...
};
```

Les unions

Quelques remarques:

- La taille d'une union est le maximum des tailles de ses champs.
- La taille est accessible en utilisant sizeof (union name).

Example :

```
union A
{
  int a; // sizeof(int) = 4
  short b; // sizeof(short) = 2
  double b; // sizeof(double) = 8
  char str[256]; // sizeof(char) = 1 * 256 elements
};
```

La taille est : sizeof (union A) = max(4, 2, 8, 256) = 256 octets.

Les unions

ATTENTION: Soyez prudent lorsque vous accédez aux champs d'union. Écrit dans n'importe quel champ d'union peut écraser la mémoire déjà écrite par un autre champ.

```
union B {
 int a;
  short b;
  char str[4];
};
union B var;
var.str[0] = 'T';
var.str[1] = 'N';
var.str[2] = 'C';
var.str[3] = 'Y';
var.b = 256; // ATTENTION: var.str ne vaut plus TNCY !!!
```

Mémoire statique vs Mémoire dynamique

Mémoire statique

Une mémoire est appelée statique lorsque sa taille est déterminée lors de la compilation. Ce type de mémoire est généralement alloué sur la pile (stack).

Mémoire dynamique

Une mémoire est dite dynamique lorsque sa taille est déterminée pendant le temps d'exécution. Ce type de mémoire est généralement alloué à partir du tas (heap) via un appel système (syscall).

Mémoire statique vs Mémoire dynamique

Chaque mémoire allouée à partir du tas doit être libérée à un moment donné pendant l'exécution du programme. Pour chaque appel *alloc⁶, il doit nécessairement y avoir un appel correspondant à free.

Mémoire statique vs Mémoire dynamique

- Il n'y a pas de coût d'allocation.
- Cache local la plupart du temps car il est situé dans la pile.
- Taille limitée.
- Très local, en raison de la nature de la pile.
- La taille doit être fixée pendant la compilation ⁶.

^{6.} C autorise l'allocation de mémoire sur la pile dont la taille est déterminée lors de l'exécution, ceci est interdit en C++.

Mémoire statique vs Mémoire dynamique

- Flexible, la taille peut être déterminée au moment de l'exécution.
- Globale
- Peut gérer des tailles que la pile ne peut pas gérer
- L'allocation peut être très coûteuse car elle nécessite un passage du mode utilisateur au mode noyau.
- Responsabilité de libérer la mémoire à la fin de l'utilisation

malloc, calloc et realloc

malloc

```
Signature : void* malloc(size_t size);
Alloue ce qui lui est passé comme argument en octets mais n'effectue aucune initialisation.
```

Exemple:

malloc, calloc et realloc

calloc

```
Signature: void* calloc(size_t nmemb, size_t size);
Alloue nmemb * size octets et les initialise à zéro.
```

Exemple:

```
void* p1 = calloc(1, 256); // 512 bytes are allocated
int* p2 = (int*)calloc(4, sizeof(int)); // 4 * sizeof(int) bytes
    are allocated
struct A* p3 = (struct A*)malloc(2, sizeof(struct A)); // 2 *
        sizeof(struct A) bytes are allocated
assert(p2[0] == 0); // true
assert(p3[0].a == 0); // true
assert(p3[1].b == 0); // true
assert(p3[1].str[0] == 0); // true
```

malloc, calloc et realloc

realloc

Signature: void* realloc(void* ptr, size_t size); Change la taille du bloc de mémoire pointé par ptr en taille octets.

- Si size > taille de ptr: La mémoire pointée par le pointeur retourné par realloc sera de taille size. Le contenu de ptr est garanti d'être copié mais la mémoire ajoutée ne sera pas initialisée
- Si size < taille de ptr : Le contenu de ptr sera copié jusqu'à size octets le reste du contenu de ptr sera ignoré. La taille de la mémoire pointée par la valeur de retour sera donc size.
- Si ptr est NULL : Cela aura le même effet que malloc.

malloc, calloc et realloc

• Si size est 0 et ptr n'est pas NULL : Cela aura le même effet que free.

N.B:

- Sauf si ptr est NULL, il doit avoir été renvoyé par un appel antérieur à malloc(), calloc() ou realloc().
- L'accès et/ou l'écriture au pointeur a passé à realloc après l'appel est un comportement indéfini

malloc, calloc et realloc

Exemple

```
int* p1 = (int*)calloc(4, sizeof(int)); // 4 * sizeof(int) bytes
p[0] = 1; p[1] = 2; p[2] = 3; p[3] = 4;
int* p2 = (int*)realloc(p1, 6 * sizeof(int));
assert(p2[0] == 1); // true
assert(p2[1] == 2); // true
assert(p2[5] == 6); // Access to uninitialized memory !
assert(p1[5] == 6); // U.B !
```

malloc, calloc et realloc

Exemple

```
int* p3 = (int*)realloc(p2, 3 * sizeof(int));
assert(p3[0] == 1); // true
assert(p3[1] == 2); // true
assert(p3[3] == 3); // U.B!
assert(p2[3] == 3); // U.B!
int* p4 = (int*)realloc(p3, 0); // Equivalent to free(p3)
void* p5 = realloc(NULL, 8); // Equivalent to malloc(8)
void* p6 = realloc(NULL, 0); // Equivalent to malloc(0)
```

malloc, calloc et realloc

N.B:

Le standard C ne dit rien quand 0 est passé à malloc (le comportement est spécifique au système d'exploitation).

- Sous Linux : malloc(0) renvoie NULL
- Sous Windows : malloc (0) renvoie un pointeur valide sur lequel free pourrait être appelé.

Les pointeurs

Définition

Un pointeur est une variable qui contient l'adresse d'une région de mémoire. Un pointeur peut contenir une adresse valide ou non (Exemple : le pointeur NULL).

Les pointeurs font généralement 4 ou 8 octets en fonction de l'architecture du processeur (32 ou 64 bits) :

- Avec un pointeur 32 bits, nous avons 4 Go de mémoire adressable.
- Avec un pointeur 64 bits, nous avons environ 17 milliards de Go de mémoire.

Arithmétique des pointeurs

Explication

Un pointeur en c'est une adresse, qui est une valeur numérique. Par conséquent, on peut effectuer des opérations arithmétiques sur un pointeur comme on peut le faire sur des valeurs numériques. Il existe quatre opérateurs arithmétiques qui peuvent être utilisés sur les pointeurs : ++, -, + et -.

La formule

Pour un pointeur ptr avec le type TYPE l'expression ptr+step ajoutera step*sizeof(TYPE) octets au pointeur ptr.

Arithmétique des pointeurs

Example

Lorsque vous utilisez l'arithmétique du pointeur, veillez à ne pas dépasser la taille allouée

Les enums

Chapitre 4

- Introduction
- 2. Compilation
- 3. La langage C
- 4. Les outils