



Assembleur 101

x86/x64

EPITA · APPING2 S7

Auteur(s)	Hugo Sibony
------------------	-------------

Date	23/12/2025
-------------	------------

Table des matières

1	C'est quoi l'assembleur ?	3
1.1	Exemple simple	3
1.2	Structure d'un fichier assembleur	3
2	Les registres : ta mémoire rapide	4
2.1	Registres 64-bit (amd64)	4
2.2	Registres 32-bit (i386)	4
2.3	Sous-parties des registres	5
3	Les deux syntaxes : AT&T vs Intel	6
3.1	Tableau comparatif	6
3.2	Exemples côte à côte	6
3.3	Suffixes de taille (AT&T)	7
3.4	Template Intel avec GAS	7
4	La pile : comment ça marche	8
4.1	Opérations de base	8
4.2	Visualisation de la pile après un CALL	8
5	Boucles et conditions	9
5.1	Les flags du CPU	9
5.2	Mettre un registre à zéro	9
5.3	Comparer deux valeurs : <code>cmp</code>	9
5.4	Tester si un registre est zéro : <code>test</code>	9
5.5	Sauts conditionnels	10
5.6	Exemple : condition simple	10
5.7	Exemple : boucle while	11
5.8	Exemple : parcourir une chaîne	11
5.9	Exemple : boucle for avec tableau	12
5.10	Astuce : <code>lea</code> pour l'arithmétique	12
6	L'ABI : le contrat entre fonctions	13
6.1	ABI i386 (32-bit)	13
6.2	ABI amd64 (64-bit)	13
6.3	Sauvegarder les registres : quand et comment ?	13
7	Appeler des fonctions C depuis l'assembleur	14
7.1	Recette	14
7.2	Alignement de la pile	14
7.3	@PLT et @GOTPCREL	14
7.4	<code>lea</code> vs <code>mov</code>	15
7.5	Exemple : appeler <code>puts</code>	15
7.6	<code>printf</code> avec plusieurs arguments	15
7.7	Fonctions fichiers de la <code>libc</code>	15
7.8	Exemple : afficher un fichier	16
8	Les syscalls : parler au noyau Linux	17
8.1	Différence avec les fonctions C	17
8.2	Structure d'un syscall	17
8.3	Numéros et arguments	17

8.4	File descriptors standards	18
8.5	Flags pour open()	18
8.6	Flags pour mmap()	18
8.7	mmap() pour allouer de la mémoire	19
8.8	Gestion des erreurs	19
9	La division	20
9.1	Division 64-bit	20
9.2	Division signée	20
10	Debugger avec GDB	21
10.1	Lancer GDB	21
10.2	Exemple de session	21
11	Aide-mémoire	22
11.1	Conventions d'appel	22
11.2	Callee-saved (à sauvegarder)	22
11.3	Syscalls	22
11.4	Templates	22

1 C'est quoi l'assembleur ?

L'assembleur est le langage le plus proche du processeur. Chaque instruction correspond directement à une opération du CPU.

1.1 Exemple simple

En C :

```
int x = 5;
x = x + 3;
```

En assembleur, on manipule directement les registres du CPU :

```
movl $5, %eax    # Met 5 dans le registre EAX
addl $3, %eax    # Ajoute 3 à EAX → EAX contient maintenant 8
```

1.2 Structure d'un fichier assembleur

```
.section .data          # Section des données (variables globales)
message: .asciz "Hello" # Une chaîne de caractères

.section .text          # Section du code
.globl ma_fonction      # Rend le symbole visible pour le linker

ma_fonction:            # Étiquette = début de la fonction
    # ton code ici
    ret                 # Retour à l'appelant
```

★ Directives importantes

Directive	Rôle
.section .text	Code exécutable
.section .data	Données initialisées (modifiables)
.section .rodata	Données en lecture seule (constantes, chaînes)
.globl nom	Exporte le symbole (obligatoire pour que C trouve ta fonction !)
.ascii "..."	Chaîne sans <code>\0</code> final
.asciz "..."	Chaîne avec <code>\0</code> final
.long valeur	Valeur 32-bit
.quad valeur	Valeur 64-bit

2 Les registres : ta mémoire rapide

Les registres sont des « variables » directement dans le CPU. Ils sont ultra-rapides.

2.1 Registres 64-bit (amd64)

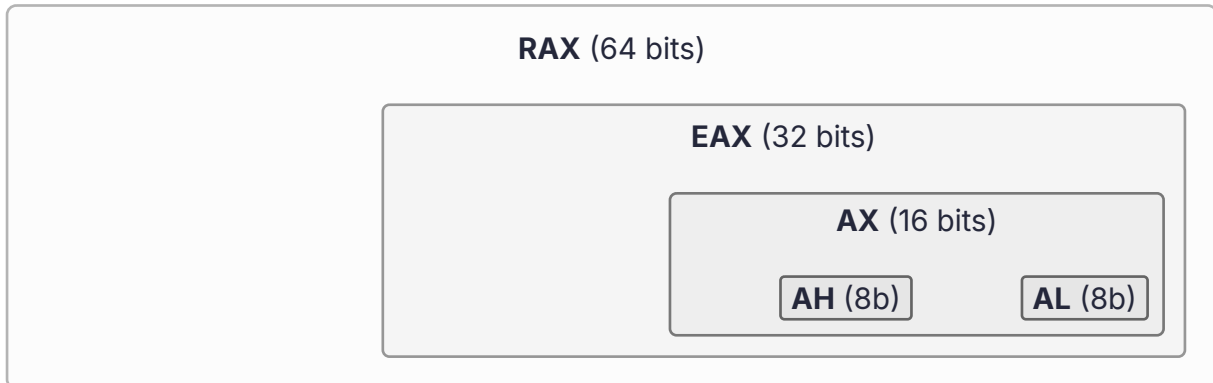
Registre	Rôle
RAX	Valeur de retour des fonctions
RBX	Usage général (À SAUVEGARDER !)
RCX	4ème argument / compteur
RDX	3ème argument / utilisé par les divisions
RSI	2ème argument
RDI	1er argument
RSP	Pointeur de pile (NE JAMAIS PERDRE !)
RBP	Pointeur de base (À SAUVEGARDER !)
R8	5ème argument
R9	6ème argument
R10-R11	Temporaires
R12-R15	Usage général (À SAUVEGARDER !)

2.2 Registres 32-bit (i386)

Registre	Rôle
EAX	Valeur de retour
EBX	Usage général (À SAUVEGARDER !)
ECX	Compteur
EDX	Données / utilisé par les divisions
ESI	Source (À SAUVEGARDER !)
EDI	Destination (À SAUVEGARDER !)
ESP	Pointeur de pile
EBP	Pointeur de base (À SAUVEGARDER !)

2.3 Sous-parties des registres

Un registre 64-bit peut être accédé par morceaux :



! Important

ATTENTION : Écrire dans EAX met les 32 bits hauts de RAX à zéro automatiquement !

```
movq $0xFFFFFFFFFFFFFFFF, %rax # RAX = 0xFFFFFFFFFFFFFFFF
movl $1, %eax # RAX = 0x0000000000000001 (pas 0xFFFFFFFF00000001!)
```

3 Les deux syntaxes : AT&T vs Intel

Tu dois savoir écrire dans les deux syntaxes. Voici les différences :

3.1 Tableau comparatif

Aspect	AT&T	Intel
Ordre	source, destination	destination, source
Registres	<code>%rax</code>	<code>rax</code>
Immédiats	<code>\$42</code>	<code>42</code>
Mémoire	<code>8(%rsp)</code>	<code>[rsp + 8]</code>
Taille	Suffixe : <code>movl</code> , <code>movq</code>	Mot-clé : <code>DWORD</code> , <code>QWORD</code>

3.2 Exemples côte à côte

Mettre une valeur dans un registre :

AT&T

```
movl $42, %eax
movq $100, %rax
```

Intel

```
mov eax, 42
mov rax, 100
```

Copier un registre dans un autre :

AT&T

```
movq %rdi, %rax
```

Intel

```
mov rax, rdi
```

Lire depuis la mémoire (pile) :

AT&T

```
movl 4(%esp), %eax
```

Intel

```
mov eax, [esp + 4]
```

Écrire en mémoire :

AT&T

```
movq %rax, (%rdi)
movq %rax, 8(%rdi)
```

Intel

```
mov [rdi], rax
mov [rdi + 8], rax
```

3.3 Suffixes de taille (AT&T)

Suffixe	Taille	Registre exemple
b	1 octet (byte)	AL , BL , CL , R8B
w	2 octets (word)	AX , BX , CX , R8W
l	4 octets (long/dword)	EAX , EBX , ECX , R8D
q	8 octets (quad)	RAX , RBX , RCX , R8

3.4 Template Intel avec GAS

Pour utiliser la syntaxe Intel avec l'assembleur GNU (GAS) :

```
.intel_syntax noprefix    # Active Intel, désactive les préfixes % et $

.text
.globl ma_fonction
ma_fonction:
    # ton code en syntaxe Intel ici
    ret
```


4 La pile : comment ça marche

La pile est une zone mémoire qui **grandit vers le bas** (adresses décroissantes).

4.1 Opérations de base

```
# PUSH: Empile une valeur
pushq %rax
# Équivalent à:
# 1. RSP = RSP - 8      (la pile grandit vers le bas)
# 2. Écrit RAX à l'adresse RSP

# POP: Dépile une valeur
popq %rax
# Équivalent à:
# 1. Lit la valeur à l'adresse RSP dans RAX
# 2. RSP = RSP + 8      (la pile rétrécit)
```

4.2 Visualisation de la pile après un CALL

Quand le C appelle ta fonction `add(3, 4)` en 32-bit :

```
Adresses hautes
|
v
+-----+
|  4  | ← ESP + 8  (2ème argument)
+-----+
|  3  | ← ESP + 4  (1er argument)
+-----+
| RetAddr | ← ESP      (adresse de retour, poussée par CALL)
+-----+
^
|
Adresses basses
```

! Important

C'est pour ça que le 1er argument est à ESP+4 et pas ESP+0 ! L'instruction CALL pousse automatiquement l'adresse de retour.

5 Boucles et conditions

5.1 Les flags du CPU

Le CPU possède des « flags » (drapeaux) mis à jour par certaines instructions :

- **ZF** (Zero Flag) : mis à 1 si le résultat est zéro
- **SF** (Sign Flag) : mis à 1 si le résultat est négatif

Les instructions de saut (`jz` , `jnz` , etc.) lisent ces flags pour décider si elles sautent.

5.2 Mettre un registre à zéro

```
xorq %rax, %rax      # RAX = RAX XOR RAX = 0
```

★ Astuce

Pourquoi `xor` et pas `movq $0, %rax` ? Plus court et plus rapide.

5.3 Comparer deux valeurs : `cmp`

```
cmpq %rsi, %rdi      # Calcule RDI - RSI, met les flags, jette le résultat
```

Après `cmpq %rsi, %rdi` :

- Si `RDI == RSI` → `ZF = 1`
- Si `RDI < RSI` → `SF = 1` (pour nombres signés)

5.4 Tester si un registre est zéro : `test`

```
testq %rdi, %rdi     # Calcule RDI AND RDI, met les flags, jette le résultat
```

`RDI AND RDI` = `RDI`. Donc :

- Si `RDI == 0` → résultat = 0 → **ZF = 1**
- Si `RDI != 0` → résultat != 0 → **ZF = 0**

C'est la façon standard de tester si un registre vaut zéro.

5.5 Sauts conditionnels

Instruction	Condition (après <code>cmp b, a</code> ou <code>test</code>)
<code>jz</code> / <code>jz</code>	$a == b$ (ou résultat $== 0$)
<code>jne</code> / <code>jnz</code>	$a != b$ (ou résultat $!= 0$)
<code>jl</code> / <code>jg</code>	$a < b$ / $a > b$ (signé)
<code>jle</code> / <code>jge</code>	$a \leq b$ / $a \geq b$ (signé)
<code>jb</code> / <code>ja</code>	$a < b$ / $a > b$ (non-signé)
<code>jbe</code> / <code>jae</code>	$a \leq b$ / $a \geq b$ (non-signé)
<code>js</code>	résultat négatif
<code>jmp</code>	toujours

★ Astuce

Signé vs non-signé : `jl` / `jg` pour int, `jb` / `ja` pour unsigned

5.6 Exemple : condition simple

En C :

```
if (x == 0) return -1;
return x;
```

En assembleur :

```
testq %rdi, %rdi      # RDI AND RDI → ZF=1 si RDI==0
jnz not_zero          # saute si ZF=0 (donc si RDI ≠ 0)
movq $-1, %rax
ret
not_zero:
movq %rdi, %rax
ret
```

5.7 Exemple : boucle while

En C :

```
int sum = 0;
while (n > 0) { sum += n; n--; }
return sum;
```

En assembleur :

```
xorq %rax, %rax      # sum = 0 (xor avec soi-même = 0)
loop:
testq %rdi, %rdi     # n == 0 ?
jz done              # si ZF=1 (n==0), termine
addq %rdi, %rax       # sum += n
decq %rdi             # n--
jmp loop
done:
ret
```

5.8 Exemple : parcourir une chaîne

En C :

```
int len = 0;
while (*s != '\0') { len++; s++; }
return len;
```

En assembleur :

```
xorq %rax, %rax      # len = 0
loop:
movzbl (%rdi), %ecx   # charge 1 octet à l'adresse RDI dans ECX
testb %cl, %cl        # octet == 0 ?
jz done
incq %rax             # len++
incq %rdi             # s++ (passe au caractère suivant)
jmp loop
done:
ret
```

i Information

`movzbl (%rdi), %ecx` = « Move Zero-extend Byte to Long »

- Lit 1 octet à l'adresse RDI
- L'étend à 32 bits en ajoutant des zéros (zero-extend)
- Le met dans ECX

Équivalent Intel : `movzx ecx, BYTE PTR [rdi]`

5.9 Exemple : boucle for avec tableau

En C :

```
int sum = 0;
for (int i = 0; i < len; i++) { sum += arr[i]; }
return sum;
```

En assembleur :

```
xorl %eax, %eax      # sum = 0
xorl %ecx, %ecx      # i = 0
loop:
  cmpl %esi, %ecx     # compare i avec len
  jge done            # si i ≥ len, termine
  movl (%rdi,%rcx,4), %edx # edx = arr[i]
  addl %edx, %eax      # sum += arr[i]
  incl %ecx            # i++
  jmp loop
done:
  ret
```

★ Adressage indexé

$(\%rdi, \%rcx, 4) = \text{adresse } RDI + RCX * 4$

- RDI = adresse de base du tableau
- RCX = index
- 4 = taille d'un élément (4 octets pour `int`)

5.10 Astuce : lea pour l'arithmétique

`lea` calcule une adresse sans accéder à la mémoire :

AT&T

```
leaq (%rdi,%rsi), %rax # addition
leaq (%rdi,%rdi,4), %rax # * 5
leaq 1(%rdi), %rax     # + 1
```

Intel

```
lea rax, [rdi + rsi]
lea rax, [rdi + rdi*4]
lea rax, [rdi + 1]
```

★ Multiplication par 10 (utile pour atoi)

`result = result * 10 + digit`

```
leaq (%rax,%rax,4), %rax # rax *= 5
leaq (%rdx,%rax,2), %rax # rax = digit + rax*2 (donc digit + ancien*10)
```

6 L'ABI : le contrat entre fonctions

L'ABI (Application Binary Interface) définit les règles pour que les fonctions puissent s'appeler correctement.

6.1 ABI i386 (32-bit)

Catégorie	Valeur
Arguments	Sur la pile : [ESP+4] , [ESP+8] , [ESP+12] ...
Retour	EAX
Libres	EAX, ECX, EDX
À sauvegarder	EBX, ESI, EDI, EBP

6.2 ABI amd64 (64-bit)

Catégorie	Valeur
Arguments	RDI, RSI, RDX, RCX, R8, R9 (dans cet ordre)
Retour	RAX
Libres	RAX, RCX, RDX, RSI, RDI, R8-R11
À sauvegarder	RBX, RBP, R12-R15
Pile	RSP multiple de 16 avant <code>call</code>

6.3 Sauvegarder les registres : quand et comment ?

Si ta fonction utilise un registre « callee-saved » (comme RBX), tu DOIS :

1. Le sauvegarder au début (avec `push`)
2. Le restaurer à la fin (avec `pop`)

```
ma_fonction:
    pushq %rbx           # Sauvegarde RBX

    # ... tu peux utiliser RBX ici ...

    popq %rbx           # Restaure RBX
    ret
```

Si tu utilises uniquement des registres « caller-saved » (RAX, RDI, RSI...), pas besoin de les sauvegarder.

7 Appeler des fonctions C depuis l'assembleur

7.1 Recette

1. Mettre les arguments dans RDI, RSI, RDX, RCX, R8, R9
2. Aligner la pile (RSP doit être multiple de 16 avant `call`)
3. `call fonction@PLT` → résultat dans RAX

7.2 Alignement de la pile

À l'entrée de ta fonction, $RSP = 8 \bmod 16$ (l'adresse de retour a été pushée).

Solution simple : `pushq %rbx` avant le `call` (ajoute 8 octets → RSP multiple de 16).

Solution avec frame pointer (recommandée pour fonctions complexes) :

AT&T :

```
ma_fonction:
    pushq %rbp
    movq %rsp, %rbp
    subq $16, %rsp

    # ... ton code ...
    # Variables locales: -8(%rbp), -16(%rbp)

    leave
    ret
```

Intel :

```
ma_fonction:
    push rbp
    mov rbp, rsp
    sub rsp, 16

    ; ... ton code ...
    ; Variables locales: [rbp-8], [rbp-16]

    leave
    ret
```

★ Astuce

`leave` = restaure RSP depuis RBP puis dépile RBP. Pratique pour nettoyer le stack frame.

7.3 @PLT et @GOTPCREL

Quand tu appelles des fonctions de la libc ou accèdes à des variables globales externes :

```
# Appeler une fonction externe (printf, puts, getline, etc.)
call printf@PLT          # @PLT = Procedure Linkage Table

# Accéder à une variable globale externe (stdin, stdout, stderr)
movq stdin@GOTPCREL(%rip), %rdx # Charge l'ADRESSE de stdin depuis la GOT
movq (%rdx), %rdx             # Charge la VALEUR de stdin (le FILE*)
```

i Information

Pourquoi @PLT ? Le linker dynamique résout l'adresse de la fonction au runtime.

Pourquoi @GOTPCREL ? Les variables globales externes sont dans la Global Offset Table.

7.4 lea vs mov

```
lea msg(%rip), %rdi    # RDI = ADRESSE de msg (pour passer une chaîne)
movq var(%rip), %rdi   # RDI = CONTENU de var (pour lire une variable)
```

7.5 Exemple : appeler puts

```
.section .rodata
msg: .asciz "Hello"

.section .text
.globl exemple
exemple:
    pushq %rbx           # Aligne la pile
    leaq msg(%rip), %rdi  # arg1 = adresse de msg
    call puts@PLT
    popq %rbx
    ret
```

7.6 printf avec plusieurs arguments

```
leaq format(%rip), %rdi    # arg1: format string
movq %rcx, %rsi            # arg2: valeur à afficher
xorl %eax, %eax            # RAX = 0 (pas d'arguments flottants)
call printf@PLT
```

! Important

`xorl %eax, %eax` est obligatoire avant printf/scanf (fonctions variadiques).

7.7 Fonctions fichiers de la libc

Fonctions pour manipuler des fichiers :

Fonction	Prototype	Description
fopen	FILE *fopen(char *path, char *mode)	Ouvre un fichier, retourne NULL si erreur
fclose	int fclose(FILE *stream)	Ferme un fichier
fgetc	int fgetc(FILE *stream)	Lit un caractère, retourne -1 (EOF) si fin
fputc	int fputc(int c, FILE *stream)	Écrit un caractère
putchar	int putchar(int c)	Écrit un caractère sur stdout

★ Astuce

Modes fopen : "r" = lecture, "w" = écriture (crée/écrase)

7.8 Exemple : afficher un fichier

```
.section .rodata
mode_r: .asciz "r"

.section .text
.globl cat
cat:
    pushq %rbx                # RDI = filename
    leaq mode_r(%rip), %rsi    # Aligne pile + sauvegarde
    call fopen@PLT
    testq %rax, %rax
    jz error
    movq %rax, %rbx            # FILE* dans RBX
loop:
    movq %rbx, %rdi
    call fgetc@PLT
    cmpl $-1, %eax
    je done
    movl %eax, %edi
    call putchar@PLT
    jmp loop
done:
    movq %rbx, %rdi
    call fclose@PLT
    movl $1, %eax
    popq %rbx
    ret
error:
    xorl %eax, %eax
    popq %rbx
    ret
```

8 Les syscalls : parler au noyau Linux

8.1 Différence avec les fonctions C

	Fonctions C	Syscalls
4ème arg	RCX	R10
Appel	<code>call func</code>	<code>syscall</code>
Erreur	Variable	RAX négatif

8.2 Structure d'un syscall

```

movq $NUMERO, %rax    # Numéro du syscall
movq arg1, %rdi       # 1er argument
movq arg2, %rsi       # 2ème argument
movq arg3, %rdx       # 3ème argument
movq arg4, %r10       # 4ème argument (ATTENTION: R10, pas RCX!)
movq arg5, %r8        # 5ème argument
movq arg6, %r9        # 6ème argument
syscall              # Exécute le syscall
# Résultat dans RAX (négatif = erreur)

```

! Important

ATTENTION : Pour les syscalls, le 4ème argument est dans **R10**, pas RCX !
(L'instruction `syscall` utilise RCX en interne)

8.3 Numéros et arguments

#	Nom	Arguments
0	read	fd, buffer, count → nb octets lus
1	write	fd, buffer, count → nb octets écrits
2	open	path, flags, mode → fd
3	close	fd
9	mmap	addr, len, prot, flags, fd, offset → ptr

8.4 File descriptors standards

Nom	Valeur	Description
stdin	0	entrée standard
stdout	1	sortie standard
stderr	2	erreur standard

8.5 Flags pour open()

Constante	Valeur	Description
O_RDONLY	0	Lecture seule
O_WRONLY	1	Écriture seule
O_RDWR	2	Lecture et écriture
O_CREAT	0x40	Créer si n'existe pas
O_TRUNC	0x200	Tronquer si existe

★ Astuce

Pour ouvrir en écriture avec création : `O_WRONLY | O_CREAT | O_TRUNC = 0x241`

8.6 Flags pour mmap()

Constante	Valeur
PROT_READ	1
PROT_WRITE	2
MAP_PRIVATE	0x02
MAP_ANONYMOUS	0x20

★ Astuce

Pour allouer de la mémoire : `prot = 3 , flags = 0x22 , fd = -1 , offset = 0`

8.7 mmap() pour allouer de la mémoire

```
movq $9, %rax          # syscall mmap
xorq %rdi, %rdi        # addr = NULL
movq $SIZE, %rsi       # len = taille voulue
movq $3, %rdx          # prot = PROT_READ | PROT_WRITE
movq $0x22, %r10       # flags = MAP_PRIVATE | MAP_ANONYMOUS
movq $-1, %r8          # fd = -1
xorq %r9, %r9          # offset = 0
syscall
# RAX = pointeur vers la mémoire allouée (ou négatif si erreur)
```

8.8 Gestion des erreurs

Les syscalls retournent une valeur négative en cas d'erreur :

```
syscall
testq %rax, %rax       # SF=1 si RAX négatif
js erreur              # Jump if Sign (si négatif)
```

9 La division

La division en assembleur est un peu particulière car elle utilise une paire de registres.

9.1 Division 64-bit

```
xorq %rdx, %rdx      # OBLIGATOIRE! Met RDX à 0
divq %rbx            # Divise RDX:RAX par RBX
                    # RAX = quotient
                    # RDX = reste
```

! Important

`xorq %rdx, %rdx` est **obligatoire** avant `divq`. L'instruction divise le nombre 128-bit formé par RDX:RAX. Si RDX contient des valeurs aléatoires, le résultat sera faux.

9.2 Division signée

Pour les nombres signés, utilise `idivq` et `cqto` (sign-extend RAX into RDX:RAX) :

```
cqto                # Étend le signe de RAX dans RDX
idivq %rbx          # Division signée
```

10 Debugger avec GDB

10.1 Lancer GDB

```
gdb ./mon_programme
```

Commande	Description
<code>break add</code>	Breakpoint sur la fonction add
<code>run</code>	Lancer
<code>stepi</code>	1 instruction (entre dans les calls)
<code>nexti</code>	1 instruction (saute les calls)
<code>continue</code>	Continuer
<code>info registers</code>	Voir les registres
<code>p/x \$rax</code>	Afficher RAX en hexa
<code>x/s \$rdi</code>	Afficher la chaîne pointée par RDI
<code>x/10x \$rsp</code>	10 valeurs hexa depuis RSP

10.2 Exemple de session

```
$ gdb ./mon_programme
(gdb) break add          ← point d'arrêt sur ta fonction
(gdb) run                ← lance le programme
Breakpoint 1, add ()
(gdb) info registers rdi rsi ← regarde les arguments
rdi                0x3
rsi                0x4
(gdb) stepi             ← exécute 1 instruction
(gdb) info registers rax ← vérifie le résultat
(gdb) continue          ← continue l'exécution
```

11 Aide-mémoire

11.1 Conventions d'appel

Type	Arguments	Retour
Fonctions C (amd64)	RDI, RSI, RDX, RCX, R8, R9	RAX
Syscalls (amd64)	RDI, RSI, RDX, R10 , R8, R9 (numéro dans RAX)	RAX
i386	[ESP+4], [ESP+8], ...	EAX

11.2 Callee-saved (à sauvegarder)

Arch	Registres
amd64	RBX, RBP, R12-R15
i386	EBX, ESI, EDI, EBP

11.3 Syscalls

#	Nom	Args
0	read	fd, buf, count
1	write	fd, buf, count
2	open	path, flags, mode
3	close	fd
9	mmap	addr, len, prot, flags, fd, off

11.4 Templates

amd64 AT&T

```
.text
.globl func
func:
    ret
```

amd64 Intel

```
.intel_syntax noprefix
.text
.globl func
func:
    ret
```

i386

```
.text
.globl func
func:
    ret
```