

Introduction à l'exploitation binaire (pwn)

Ce cours vous donne les bases de **l'exploitation binaire** (basée sur la **pile**, basée sur le **tas**, et sur les **formats de chaîne**), ainsi que l'étude de certaines fonctions vulnérables du langage C.

- [Le format ELF\(Executable and Linking Format\)](#)
 - [Format ELF](#)
 - [Les Registres 86_64](#)
 - [Analyse d'un Binaire](#)
 - [Fonctions vulnérables de la libc](#)
 - [Exploiter un Buffer Overflow](#)

Le format ELF(Executable and Linking Format)

le format ELF(base)

Le format ELF(Executable and Linking Format)

Format ELF

ELF(Executable and Linking Format) est un format de fichier standard dans les os de type unix-like(linux en général) pour les fichiers exécutables, les bibliothèques partagées et les fichiers Objets.

Structure d'un Fichier ELF (à connaitre pour ce cours)

un fichier **ELF** contient plusieurs sections et segments qui servent à gérer les différentes parties d'un programme en mémoire

1. En-tête ELF(ELF header) :

situé au début du fichier, il contient des informations sur le type du fichier, l'architecture, les adresses d'entrée, et les offsets pour les autres sections du fichier.

Pour afficher l'en-tête d'un ELF on peut utiliser la commande `< readelf -h nom_du_binaire>`

exemples:

ici nous avons les l'en-tête de la **libc** dont le type est DYN pour dynamique est un **fichier objet partagé**

```
[h_cams@cams12 ~]$ readelf -h /usr/lib/libc.so.6
En-tête ELF:
  Magique:      7f 45 4c 46 02 01 01 03 00 00 00 00 00 00 00 00
  Classe:      ELF64
  Données:      complément à 2, système à octets de poids faible d'abord (little endian)
  Version:      1 (actuelle)
  OS/ABI:      UNIX - GNU
  Version ABI:  0
  Type:        DYN (fichier objet partagé)
  Machine:     Advanced Micro Devices X86-64
  Version:     0x1
  Adresse du point d'entrée: 0x25fe0
  Début des en-têtes de programme : 64 (octets dans le fichier)
  Début des en-têtes de section : 2010488 (octets dans le fichier)
  Fanions:     0x0
  Taille de cet en-tête: 64 (octets)
  Taille de l'en-tête du programme: 56 (octets)
  Nombre d'en-tête du programme: 14
  Taille des en-têtes de section: 64 (octets)
  Nombre d'en-têtes de section: 63
  Table d'index des chaînes d'en-tête de section: 62
```

et notre deuxième exécutable est un programme c qui est de type **fichier exécutable**

```
[h_cams@cams12 c]$ readelf -h buffer
En-tête ELF:
  Magique:   7f 45 4c 46 02 01 01 00 00 00 00 00 00 00 00
  Classe:                                ELF64
  Données:                                complément à 2, système à octets de poids faible d'abord (little endian)
  Version:                                1 (actuelle)
  OS/ABI:                                UNIX - System V
  Version ABI:                            0
  Type:                                    EXEC (fichier exécutable)
  Machine:                                Advanced Micro Devices X86-64
  Version:                                0x1
  Adresse du point d'entrée:              0x401060
  Début des en-têtes de programme :      64 (octets dans le fichier)
  Début des en-têtes de section :        13640 (octets dans le fichier)
  Fanions:                                0x0
  Taille de cet en-tête:                  64 (octets)
  Taille de l'en-tête du programme:      56 (octets)
  Nombre d'en-tête du programme:         13
  Taille des en-têtes de section:         64 (octets)
  Nombre d'en-têtes de section:          30
  Table d'index des chaînes d'en-tête de section: 29
[h_cams@cams12 c]$
```

Sur ce screenshot, nous pouvons voir différentes informations sur ce binaire, comme son type, le nombre d'en-têtes, le point d'entrée, etc.

Program Header Table : Il contient les informations nécessaires au chargeur du système pour mapper le fichier ELF en mémoire, en indiquant quelles parties du fichier correspondent aux segments de mémoire (stack, heap, section, etc.).

1. **Sections** : ce sont des parties d'un ELF qui ont un rôle bien défini :

.text : Contient le code exécutable (instructions du programme).

.data : Contient les variables (locales et globales) initialisées du programme .

.bss : Contient les variables (locales et globales) non initialisées.

.rodata : Contient les données en lecture seule, comme les chaînes de caractères constantes. ex: les chaînes que vous écrivez par exemple dans vos fonctions d'affichage de messages sur la sortie standard

.symtab et **.dynsym** : Tables de symboles utilisées pour identifier les fonctions et les variables dans le programme.

Un symbole est une chaîne de caractères dans un fichier binaire, associée à une adresse mémoire, une variable, ou une fonction.

.rel et .rela : Sections utilisées pour le relogement (relocation) des adresses lors du chargement du programme ou de l'utilisation de bibliothèques partagées.

Vous pouvez le voir dans cet exemple, certaines sections du binaire C, avec la sortie de GDB ci-dessous :

```
gdb> info files
Symbols from "/home/h_cams/programme/c/buffer".
Native process:
  Using the running image of child process 16197.
  While running this, GDB does not access memory from...
Local exec file:
  `/home/h_cams/programme/c/buffer', file type elf64-x86-64.
Entry point: 0x401060
0x0000000000400318 - 0x0000000000400334 is .interp
0x0000000000400338 - 0x0000000000400378 is .note.gnu.property
0x0000000000400378 - 0x000000000040039c is .note.gnu.build-id
0x000000000040039c - 0x00000000004003bc is .note.ABI-tag
0x00000000004003c0 - 0x00000000004003dc is .gnu.hash
0x00000000004003e0 - 0x00000000004004a0 is .dynsym
0x00000000004004a0 - 0x000000000040053e is .dynstr
0x000000000040053e - 0x000000000040054e is .gnu.version
0x0000000000400550 - 0x0000000000400590 is .gnu.version_r
0x0000000000400590 - 0x00000000004005f0 is .rela.dyn
0x00000000004005f0 - 0x0000000000400638 is .rela.plt
0x0000000000401000 - 0x000000000040101b is .init
0x0000000000401020 - 0x0000000000401060 is .plt
0x0000000000401060 - 0x0000000000401192 is .text
0x0000000000401194 - 0x00000000004011a1 is .fini
0x0000000000402000 - 0x000000000040200e is .rodata
0x0000000000402010 - 0x0000000000402044 is .eh_frame_hdr
0x0000000000402048 - 0x00000000004020f4 is .eh_frame
0x0000000000403de8 - 0x0000000000403df0 is .init_array
0x0000000000403df0 - 0x0000000000403df8 is .fini_array
0x0000000000403df8 - 0x0000000000403fc8 is .dynamic
0x0000000000403fc8 - 0x0000000000403fe8 is .got
0x0000000000403fe8 - 0x0000000000404018 is .got.plt
0x0000000000404018 - 0x0000000000404028 is .data
0x0000000000404028 - 0x0000000000404030 is .bss
```

Les types de fichiers ELF (liste non exhaustive)

- **Executable** : Un fichier ELF pouvant être directement exécuté par le système d'exploitation.
- **Shared Object** : Une bibliothèque partagée, utilisée à la fois au moment de la liaison dynamique et de l'exécution pour accéder à certaines fonctions qui ne se trouvent pas

dans notre programme, comme *"printf"* en C par exemple. Ce sont généralement des fichiers ayant une extension *.so*, comme la *libc*

Chargement et exécution d'un ELF

Lorsqu'un fichier ELF est exécuté :

1. Le **chargeur** du système lit le **program header** pour charger les segments appropriés en mémoire.
2. Il alloue de la mémoire pour la pile (stack) et le tas (heap).
3. Il place les segments de code (*.text*) et de données (*.data*, *.bss*) dans les sections de la mémoire correspondantes.
4. Enfin, il commence l'exécution en sautant à l'adresse spécifiée dans le **point d'entrée** du fichier ELF (généralement défini dans le **ELF Header**).

Mappage d'un **ELF** en mémoire :

Représentation de la mémoire pour un programme en cours d'exécution :

Espace du noyau (Kernel)	

Stack	
(variables locales, adresses	
de retour, pile des appels)	

Heap	
(mémoire dynamique, malloc, etc)	

Uninitialized Data (.bss)	
(variables non initialisées)	

Initialized Data (.data)	
(variables initialisées)	

Code (.text)	
(instructions exécutables)	

Espace réservé au programme	

Espace kernel : gère les ressources, assure la gestion des erreurs, exécute les appels système, etc.

Stack : est une structure de données qui contient les variables locales et les adresses de retour des fonctions. Elle croît vers le bas.

Heap: est une Zone de mémoire utilisée pour les allocations dynamiques (malloc, calloc). Elle croît vers le haut, en partant du bas

ce n'est pas vraiment la structure d'un elf vu qu'il contient d'autres choses donc juste le minimum !

Les Registres 86_64

Les registres sont des espaces mémoire situés dans le processeur, dont le rôle est de permettre un accès rapide aux données. Dans un ordinateur, ils constituent la mémoire la plus petite, mais aussi la plus rapide.

Lorsque notre ordinateur effectue des calculs, les données sont stockées dans les registres du processeur. Par exemple, lorsqu'on effectue une addition, le processeur utilise des registres pour stocker temporairement les valeurs des nombres à additionner.

```
ma_fonction addition(a:entier , b:entier){  
    retourner a+b  
}
```

dans l'exemple de la fonction ci-dessus, le processeur stocke la valeur de a dans un registre supposons Z et pour faire l'addition il ajoute b à Z qui contient déjà la valeur de a donc $Z = Z + b$

Types de registres :

Ils existe plusieurs types de registres en x64. Nous avons :

Les registres généraux : globalement se sont des registres qui sont utiliser pour contenir des données (variables, adresses mémoire, ...) :

.RAX (accumulateur) : Principalement utilisé pour effectuer des calculs arithmétiques et logiques. Il est également utilisé pour stocker la valeur de retour d'une fonction.

.RBX (base register) : N'a pas de rôle spécifique, mais peut être utilisé pour stocker des valeurs temporaires ou manipuler des indices de tableaux.

.RCX (compteur) : Généralement utilisé comme compteur de boucle ou pour certaines instructions répétitives (rep, loop).

.RDX (data register) : Utilisé pour stocker la partie haute du résultat lors d'opérations nécessitant plus de 64 bits (par exemple, une multiplication). Il est souvent combiné avec RAX dans ce cas.

.R8 à R15 : Registres généralistes comme RBX, utilisés pour stocker des valeurs temporaires et pour passer des arguments aux fonctions en x86_64 (R8 et R9 étant les 5^e et 6^e arguments).

.RSI (Source Index) & RDI (Destination Index) : Utilisés pour le passage des arguments aux fonctions (RSI pour le deuxième argument, RDI pour le premier argument). Peuvent être utilisés aussi comme RBX

.RFLAGS : est un registre qui contient des valeurs indiquant l'état du processeur lors de l'exécution d'un programme.

Pour le Pwn voici les registres qui sont importants :

Les registres de Pile:

RBP (base pointer)

Le registre RBP est utilisé pour sauvegarder l'adresse du cadre de pile (stack frame) dans une fonction. Le cadre de pile représente l'état de la pile au moment de l'appel d'une fonction.

Grâce à RBP, il est possible de localiser les variables locales et d'y accéder facilement, car il définit la base de la pile pour cette fonction particulière. Il se situe juste avant les variables locales d'une fonction et après le RIP (instruction pointer)

RSP (stack pointer)

Le registre RSP pointe vers le sommet de la pile, c'est-à-dire l'endroit où la prochaine valeur sera empilée ou dépilée. Lors de l'exécution d'une fonction, il change dynamiquement, car chaque nouvelle valeur ajoutée à la pile fait que RSP diminue (puisque la pile croît vers les adresses basses).

“ Le registre RIP contient l'adresse de la prochaine instruction à exécuter. Il pointe vers l'emplacement mémoire où le processeur doit se rendre pour récupérer la prochaine instruction à exécuter dans le programme.

C'est un registre clé dans le contrôle du flux d'exécution, car il permet au processeur de savoir où il en est dans l'exécution du programme et quelle sera l'instruction suivante à exécuter.

RIP et attaques :

Dans le contexte des attaques comme le buffer overflow, le RIP joue un rôle crucial. Lorsqu'un payload malveillant est écrit dans la pile, il peut modifier le RIP pour rediriger l'exécution du programme vers un code malveillant. Cela se fait souvent en écrivant l'adresse qui contient le début de notre payload dans le RIP, amenant le processeur à exécuter ce code malveillant.

Le but de ces attaques est souvent de faire en sorte que le RIP pointe vers une adresse sur laquelle

on a le contrôle, afin d'exécuter une série d'instructions bien précises (selon notre objectif et les morceaux de code disponibles dans le binaire) par le processeur.

Le format ELF(Executable and Linking Format)

Analyse d'un Binaire

Outils pour analyse du binaire

Le format ELF(Executable and Linking Format)

Fonctions vulnérables de la libc

Ici nous verrons quelques fonctions vulnérables de la libc aux attaques du type buffer overflow.

Le format ELF(Executable and Linking Format)

Exploiter un Buffer Overflow

Exploitation buffer overflow