

# Lab

# Tìm hiểu về Kỹ thuật Dịch ngược (tt)

Reverse Engineering (cont)

Thực hành Lập trình Hệ thống

Lưu hành nội bộ

# A. TỔNG QUAN

#### A.1 Muc tiêu

- Tìm hiểu và làm quen với Kỹ thuật Dịch ngược (tiếp theo).
- Cài đặt và tìm hiểu trình Disassembler IDA.
- Tìm hiểu cách remote debugger trên desktop.

#### A.2 Môi trường

#### A.2.1 Chuẩn bị môi trường

- 01 máy Linux (máy ảo) dùng để làm môi trường chạy file cần phân tích.
- 01 máy Windows chạy chương trình IDA Pro (6.6 hoặc 7.0) cho quá trình phân tích.

## A.2.2 Các tập tin được cung cấp sẵn

- File cần phân tích: **bomb** – là một file thực thi Linux 32-bit.

### A.3 Liên quan

- Sinh viên cần vận dụng kiến thức trong Chương 3 (Lý thuyết).
- Tìm hiểu thêm các kiến thức về remote debugger trên desktop khi sử dụng IDA.

## **B. THỰC HÀNH**

Bài thực hành này yêu cầu sinh viên thực hiện phân tích file thực thi **bomb**. File thực thi Linux 32-bit này được thiết kế như 1 trò chơi phá bomb. Quả bomb này gồm nhiều pha, mỗi pha yêu cầu nhập vào một input riêng để ngắt pha, nếu sai bomb sẽ nổ và trò chơi kết thúc. Sinh viên cần áp dụng kỹ thuật dịch ngược và sử dụng IDA để tìm ra các input để ngắt pha và ngăn quả bomb phát nổ.

# B.1 Yêu cầu 1 - Thiết lập môi trường

## B.1.1 Tuỳ chọn 1 – Thiết lập môi trường phân tích tĩnh file

Sinh viên có thể sử dụng môi trường đã được hướng dẫn thiết lập từ bài thực hành 3 – Tìm hiểu cơ bản về Kỹ thuật dịch ngược để phân tích file **bomb** trong bài thực hành này. Việc phân tích file bomb sẽ thực hiện trên IDA Pro ở máy Windows và thực thi thử nghiệm trên máy Linux. Sinh viên cần đảm bảo:

- Với máy Windows:
  - + Cài đặt được IDA Pro trên máy Windows dùng để phân tích file thực thi.
  - + Trên máy Windows có file **bomb** và mở được với IDA Pro phiên bản phù hợp.
- Với máy Linux:
  - + Thiết lập được môi trường thực thi file 32-bit trên máy ảo Linux.

# B.1.2 Tuỳ chọn 2 - Thiết lập môi trường remote debug

Ở tuỳ chọn này, sinh viên có thể thiết lập môi trường phân tích nâng cao với tính năng debug file của IDA Pro. Tuy nhiên do cần phân tích file thực thi Linux 32-bit (không chạy



được trên Windows), cần thiết lập môi trường *Remote debug* kết nối từ IDA Pro trên máy Windows sang tiến trình thực thi file cần phân tích trên máy Linux.

## Bước 1. Kiểm tra phiên bản môi trường Linux và cài đặt các package nếu cần thiết

Bước kiểm tra này đảm bảo môi trường Linux có thể thực thi file **bomb** (32-bit) và để xác đinh các file IDA Pro cần thiết cho quá trình thiết lập môi trường.

Sau bước này, cần đảm bảo thực thi được file bomb với lênh:

```
$ chmod +x bomb
$ ./bomb
```

#### Bước 2. Kiểm tra kết nối giữa máy ảo Linux và máy Windows

Lưu ý: Máy ảo Linux ví du dưới đây được tao với VMWare Workstation.

Trên terminal của máy ảo, gõ lệnh sau để hiển thị các IP hiện có của máy ảo:

#### \$ ifconfig

Trên máy Windows, mở terminal và gõ lênh sau để xem IP:

#### \$ ipconfig

Quan sát kết quả hiển thị, tìm 2 IP thuộc cùng 1 mạng của máy thật và máy ảo. Ví dụ ở hình trên là 2 IP là **192.168.229.132** và **192.168.229.1** (Cùng mạng Vmnet8 của VMWare). Thực hiện lệnh **ping** để kiểm tra, nếu ping thành công thì 2 máy đã được kết nối.

```
C:\Windows\system32\cmd.exe

Microsoft Windows [Version 10.0.19042.630]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\Users\Thu-Hien Do>ping 192.168.229.132

Pinging 192.168.229.132 with 32 bytes of data:
Reply from 192.168.229.132: bytes=32 time<1ms TTL=64
```



Khi đó IP bên máy ảo có thể sử dụng ở **Bước 7** khi kết nối IDA Pro với môi trường thực thi trên máy ảo.

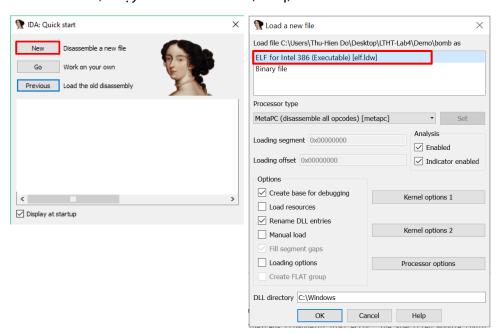
- **<u>Bước 3.</u>** Sao chép file **bomb** vào máy ảo Linux ở một thư mục nhất định.
- <u>Bước 4.</u> Vào thư mục cài đặt trên Windows của IDA, thường là **C:\Program Files\** hay **C:\Program Files (x86)**. Trong thư mục **dbgsrv** sao chép file **linux\_server** vào cùng thư mục trên máy Linux với file **bomb**.



**Bước 5.** Trên máy Linux, với terminal đang ở thư mục chứa file **bomb** và **linux\_server** của IDA, thực hiện các lệnh bên dưới để mở debug server bên máy Linux.



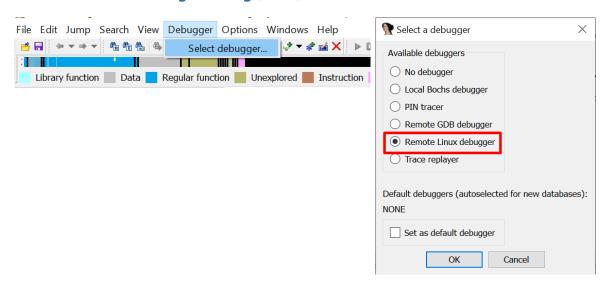
**<u>Bước 6.</u>** Trên Windows, chạy IDA bản 32 bit (idaq) và mở file **bomb**.



Bước 7. Sau khi load file thành công, chọn tab Debugger → Select Debugger và click chọn *Remote Linux Debugger* để thiết lập môi trường debug đến máy đích là máy Linux ở xa.

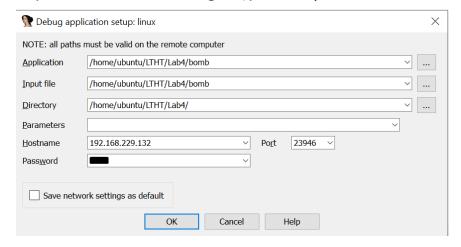
## ட

#### Lab 5: Reverse Engineering (cont)

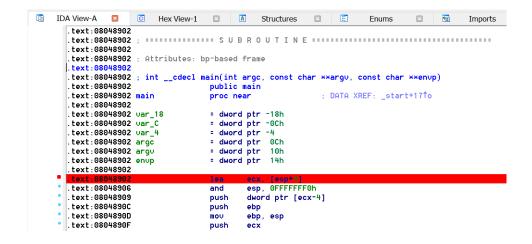


Sau đó, mở cửa sổ cài đặt debugger với **Debugger > Process Options** và điền các thông tin như sau:

- Directory: đường dẫn chứa file **bomb** trên máy Linux
- Application và Input file: đường dẫn đến file **bomb** trên máy Linux
- Hostname: địa chỉ IP của máy ảo Linux đã tìm thấy ở **Bước 2** (port để mặc định)
- Password: password tài khoản đăng nhập của máy ảo Linux.



**<u>Bước 8.</u>** Thử đặt breakpoint tại lệnh assembly đầu tiên của hàm **main** để debug.



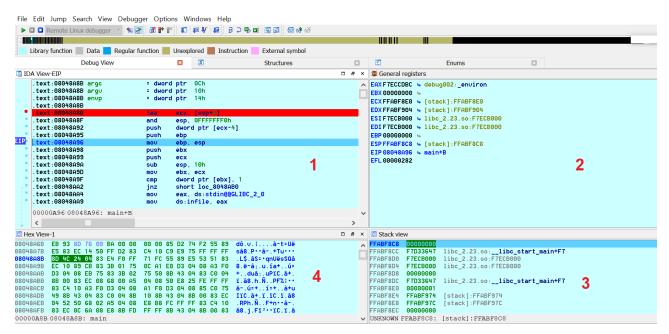


**<u>Bước 9.</u>** Lưu ý: cần thực thi file **linux\_server**trên máy ảo Linux như ở **Bước 7** trước khi thực hiện bước này.

Chạy thử chương trình trên IDA Pro với dấu dể kiểm tra remote debug. Nếu quan sát thấy các kết quả dưới đây thì quá trình thiết lập môi trường remote debug đã hoàn tất.

- Trên máy ảo Linux: có thông báo nhận được kết nối từ máy thật

- Giao diện IDA Pro: chuyển sang màn hình nền xanh dương nhạt



Quan sát giao diện trên IDA Pro khi chuyển sang màu xanh nhạt của chế độ debug sẽ hiển thị nhiều cửa sổ mô tả trạng thái của hệ thống khi đang chạy chương trình **bomb**:

- Cửa sổ (1) **Debug view** hiển thị mã assembly, cùng vị trí đặt breakpoint (dòng tô màu đỏ) và vị trí đang thực hiện debug (dòng tô màu xanh dương).
- Cửa sổ (2) General registers hiển thị giá trị hiện tại của các thanh ghi.
- Cửa sổ (3) Stack view hiển thị trạng thái của stack khi thực thi chương trình.
- Cửa sổ (4) Hex view hiển thị dạng mã hex nội dung file thực thi

Trên IDA, dùng icon trên thanh công cụ để xem hoạt động qua từng dòng code của các hàm. *Lưu ý*: với các hàm phổ biến ta đã biết chức năng của chúng, ở lệnh call tương ứng với thể dùng ký hiệu dể thực thi mà không cần vào đoạn code cụ thể của từng hàm.



#### **B.2** Yêu cầu 2 – Thực hành giải các pha (phá bomb)

#### B.2.1 Tổng quan về các pha

File thực thi **bomb** là một chương trình dạng command line, đã được lập trình sẵn và có hoạt động như một quả bomb. Quả bomb này gồm 6 pha, mỗi pha yêu cầu một input riêng để ngắt pha. Nếu input sai, bomb sẽ nổ và trò chơi kết thúc.

```
■● ubuntu@ubuntu:~/LTHT/Lab4
ubuntu@ubuntu:~/LTHT/Lab4$ ./bomb
Welcome to my fiendish little bomb. You have 6 phases with
which to blow yourself up. Have a nice day!
Hello world

BOOM!!!
The bomb has blown up.
ubuntu@ubuntu:~/LTHT/Lab4$
```

Trong trường hợp nhập đúng input, chương trình sẽ in ra thông báo chúc mừng tương ứng và chờ nhập input cho phase kế tiếp.

```
■●  ubuntu@ubuntu:~/LTHT/Lab4

ubuntu@ubuntu:~/LTHT/Lab4$ ./bomb

Welcome to my fiendish little bomb. You have 6 phases with
which to blow yourself up. Have a nice day!

Phase 1 defused. How about the next one?
```

Các pha có thể được cấu trúc lại thành các nhóm dựa trên độ khó như sau:

- *Bô pha cơ bản*: Pha 1, 2, 3
- <u>Bô pha trung bình khó</u>: Pha 4 và 5
- <u>Bộ pha rất khó</u>: Pha 6
- Môt pha bí mât 😂

Đô khó của mỗi pha được đánh giá bằng dấu (\*) như sau:

- Pha 1:\*
- Pha 2: \*\*
- Pha 3: \*\*\*
- Pha 4: \*\*\*\*\*
- Pha 5: \*\*\*\*
- Pha 6: \*\*\*\*\*

Trong phần này, sinh viên được yêu cầu giải mỗi pha của Bomb khi thực thi file "bomb" và nộp kèm báo cáo quá trình phân tích, thực hiện.



#### B.2.2 Hướng dẫn cách giải các pha

*Lưu ý:* Trong bài thực hành sẽ có *nhiều phiên bản* file **bomb** khác nhau, bài hướng dẫn sử dụng 1 phiên bản bất kỳ để phân tích.

• Bước 1: Thực thi thử file bomb trên Linux với lệnh:

#### ./bomb

Sau một số thông báo chào mừng, có thể thấy file tạm dừng để yêu cầu cần nhập một input. Vì ban đầu chưa biết input cần những yêu cầu gì nên có thể nhập một đoạn input bất kì. Lúc này chương trình thông báo sai, quả bomb nổ và thoát.

#### Bước 2: Phân tích file bomb

Có 2 cách để thực hiện bước phân tích file bomb với 2 tuỳ chọn môi trường ở phần **B.1** 

## Tuỳ chọn 1: Phân tích tĩnh file bomb với IDA Pro

Sinh viên đọc code assembly của file bomb như ở Bài thực hành 3. Các phân tích ở bước này với tuỳ chọn 2 vẫn có thể áp dụng với tuỳ chọn 1.

## Tuỳ chọn 2: Debug file bomb với IDA Pro

Sau khi thiết lập môi trường debug như phần B.1.2, với việc thực thi file **linux\_server** trên máy Linux, đặt breakpoint trong mã assembly ở vị trí hàm main trong IDA Pro, và nhấp chọn biểu tượng màu xanh trên thanh công cụ để bắt đầu debug, ta được kết quả sau:

Dùng icon dể chạy từng dòng code của hàm **main** để xem hoạt động của chương trình. Ta có thể thấy đoạn mã in ra các dòng chữ trong lần chạy thử file **bomb** với 2 hàm **puts** như hình dưới.

```
text:08048B2A
                                          sub
                                                     esp, OCh
offset s
text:08048B2D
                                          push
call
                                                                            ; "Welcome to my fiendish little bomb. You"...
text:08048B32
text:08048B37
text:08048B3A
                                                     esp, 0Ch
offset aWhichToBlowYou ; "which to blow yourself up. Have a nice ".
                                          sub
text:08048B3D
text:08048B42
                                          push
call
                                                      puts
text:08048B47
text:08048B4A
text:08048B4F
                                          add
call
mov
                                                     esp, 10h
read_line
[ebp+input], eax
text:08048B52
                                          sub
                                                      esp, OCh
text - 02042855
                                                      [ebp+input]
                                                     esp, 10h
phase_defused
text:08048B5D
```



Hàm **main** ở đây gọi một hàm **read\_line**, có tác dụng chính là để đọc 1 chuỗi input từ người dùng. Hàm **read\_line** luôn đọc vào 1 chuỗi, đồng thời **eax** thường chứa giá trị trả về của 1 hàm, **eax** ngay sau khi thực thi xong **read\_line** sẽ là chuỗi đã nhập. Sau đó giá trị trong **eax** được đưa vào [**ebp + input**], và [**ebp + input**] sẽ được push vào stack để làm tham số trước khi gọi **phase\_1** (pha 1). Như vậy hàm **main** đã dùng **read\_line** để đọc input từ người dùng và truyền input này như tham số của hàm **phase\_1**.

Có thể thấy, sau khi gọi hàm **phase\_1**, một hàm **phase\_defused** sẽ được gọi. Theo tên gọi có vẻ như nó dùng để vô hiệu hóa bomb, là hàm được gọi khi 1 pha được giải thành công. Như vậy, ở hàm **phase\_1** có thể sẽ kiểm tra input người dùng, chỉ cần trả về bình thường là vô hiệu hoá được 1 pha, ngược lại sẽ nổ bom trong chính hàm **phase\_1**. Trình tự tương tự cũng áp dụng cho các hàm **phase\_2**, **phase\_3**,... ứng với các pha còn lại.

Nhiệm vụ tiếp theo là đọc code cụ thể của những hàm **phase\_1**, **phase\_2**,... này để hiểu hoạt động của từng pha.

## • Bước 3: Phân tích các hàm để tìm input cho mỗi pha

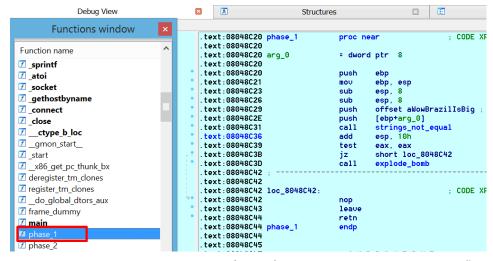
Nhìn chung, ở mỗi pha của bài lab, để phá bom thành công, sinh viên cần xác định được các đặc điểm của input như sau:

- Định dạng của input (kí tự, số hay kết hợp giữa kí tự và số?).
- Số lượng phần tử trong input (bao nhiêu số, ký tự hay chuỗi?)
- Nếu có nhiều phần tử, phân tích mã assembly để tìm quan hệ ràng buộc giữa chúng.

## B.3 Hướng dẫn giải chi tiết pha 1

## • Bước 1: Tìm kiếm định dạng input

Ở trên ta đã tìm thấy function có tên là **phase\_1**, đây là phần mã nguồn của pha 1. Nhấp đúp chuột vào tên hàm để xuất hiện mã assembly của pha.



Phân tích từ hàm **main** trước đó, tham số truyền cho **phase\_1** là một chuỗi. Debug theo từng dòng assembly, ta thấy ở vị trí **[ebp + arg\_0]** (ebp + 8) chính là vị trí tham số này, hay chính là địa chỉ của chuỗi ta đã nhập. Chuỗi này được sử dụng để so sánh luôn nên ở pha 1 yêu cầu nhập vào 1 chuỗi.



## • Bước 2: Phân tích mã nguồn để tìm điều kiện thỏa mãn của input

Từ mã assembly, ta có những hoạt động sau:

- Các địa chỉ của chuỗi input ở [ebp + arg\_0] và một biến aWowBrazilIsBig được push vào stack trước khi gọi strings\_not\_equal. Các giá trị này là tham số của hàm.
- Có thể coi sơ hoạt động hàm **strings\_not\_equal** bằng mã giả bằng cách chọn hàm và nhấn **F5**: hàm này thực hiện so sánh từng phần tử của 2 mảng. Hàm trả về 1 nếu 2 mảng không bằng, 0 nếu ngược lại.

```
Pseudocode
   1<mark>signed</mark> int __cdecl strings_not_equal(int a1, int a2)
      int v2; // ebx@1
     signed int result; // eax@2
int v4; // [sp+8h] [bp-Ch]@3
int v5; // [sp+Ch] [bp-8h]@3
      u2 = string_length(a1);
      if ( v2 == string_length(a2) )
 10
      {
         υ4 = a1;
 12
 13
         while ( *(\_BYTE *) \cup 4 )
           if ( *(_BYTE *)∪4 != *(_BYTE *)∪5 )
 15
 16
           return 1;
++∪4;
 17
           ++05;
 19
        result = 0;
 20
 22
      else
 23
         result = 1:
 25
      return result;
```

- Dòng lệnh **test eax, eax** kiểm tra giá trị trong thanh ghi eax ngay sau khi gọi **string\_not\_equal**, tức là kiểm tra kết quả trả về của hàm này. Nếu bằng 0 sẽ nhảy đến **loc\_8048C42** để kết thúc hàm, còn khác 0 sẽ gọi **explode\_bomb** (nổ bom!). Như vậy, kết quả so sánh 2 mảng input nhập vào và biến **aWowBrazilIsBig** nếu khác nhau sẽ khiến bom nổ, còn lại thì trả về bình thường.
- ➡ Từ đó có thể thấy chuỗi ký tự được nhập cần giống với giá trị biến aWowBrazilIsBig, nhấp đúp vào biến dễ thấy nó có giá trị "Wow! Brazil is big."

Để dễ hơn, có thể nhấn **F5** để xem mã giả của **phase\_1**:

Vậy input cần nhập của pha 1 là: Wow! Brazil is big.



#### • Bước 3: Kiểm tra kết quả

Như vậy, với tham số là chuỗi input người dùng nhập vào, trong các hàm xử lý từng pha tương tự như **phase\_1**, nếu input không thỏa điều kiện kiểm tra tương ứng, chương trình lập tức chạy hàm **explode\_bomb** để nổ bom và thoát chương trình. Ngược lại, hàm trả về bình thường và chương trình gọi tiếp **phase\_defused()**. Vậy yêu cầu cần thực hiện là làm sao để các pha được gọi với input được nhập trả về bình thường.

# C. MỘT SỐ GỢI Ý CÁC PHA CÒN LẠI

## C.1 Định dạng input của mỗi pha

Tất cả các pha của chương trình đều nhận input là chuỗi của người dùng từ hàm **read\_line()**, do dữ liệu nhập từ bàn phím luôn là chuỗi. Tuy vậy, trong mỗi pha sẽ có những đoạn gọi hàm **scanf()** khác nhau để từ một chuỗi input có thể lấy ra được những giá trị cần thiết, như một số, một ký tự, hoặc nhóm các số và ký tự.

Ví dụ như bên dưới, hàm **scanf()** trong pha 2 thực hiện đọc chuỗi input lấy từ **read\_line()** với định dạng **"%d %d %d %d %d %d"**, tức là lấy 6 số nguyên.

## C.2 Nhắc lại một số kiến thức cơ bản về procedure (hàm)

Một số lưu ý về procedure (hàm) trong mã assembly như sau:

- Môt hàm được gọi với lệnh call <tên hàm>
- Một hàm có thể có các tham số truyền vào, các tham số này được truyền bằng những câu lệnh **push** giá trị trước lệnh **call** theo thứ tự ngược với khai báo trong C.

Ví dụ một hàm C có tên **add(int a, int b)** thì trong assembly sẽ tìm thấy những đoạn mã như sau để gọi hàm này:

```
push b
push a
call add
```

- Giá trị trả về của một hàm thường được lưu trong thanh ghi eax. Những xử lý trên giá trị thanh ghi eax ngay sau khi gọi 1 hàm với lệnh call có thể hiểu là xử lý giá trị được trả về từ hàm đó.



## D. YÊU CẦU & ĐÁNH GIÁ

Bài thực hành sẽ có **nhiều phiên bản khác nhau** của file **bomb** với các yêu cầu input ở các pha khác nhau, mỗi file gồm 6 pha. Mỗi nhóm sinh viên sẽ được GVTH giao nhiệm vụ phân tích 1 trong các phiên bản.

Sinh viên thực hành và nộp bài **theo nhóm tối đa 2 sinh viên**. Ở mỗi pha, sinh viên viết báo cáo ngắn gọn về cách thức tìm thấy input cần tìm với các nội dung tối thiểu:

- (1) Định dạng của input (số/ký tự/tổ hợp số ký tự) và số lượng (nếu có).
- (2) Điều kiện ràng buộc của input.
- (3) Kết luận về input của pha.
- (4) Hình ảnh chup kết quả thực thi input đã tìm thấy với file bomb

Mỗi phần cần có hình ảnh (đoạn mã assembly/mã giả/ảnh chụp màn hình) minh chứng. Tên file ghi rõ Lab5-MSSV1-MSSV2-<số pha giải được | Full>.pdf

# Đặt tên thư mục nộp bài theo quy tắc sau: LabX-MSSV1-MSSV2-HoTênSV1-HoTênSV2

Ví du: Lab2-16520901-17520078-NguyenVanA-TranVanC

## E. THAM KHẢO

[1] Randal E. Bryant, David R. O'Hallaron (2011). Computer System: A Programmer's Perspective (CSAPP)
[2] Hướng dẫn sử dụng công cụ dịch ngược IDA Debugger – phần 1 [Online]

https://securitydaily.net/huong-dan-su-dung-cong-cu-dich-nguoc-ma-may-ida-debugger-phan-1/
[3] IDA công cụ hoàn hảo cho các chuyên gia Reverse Engineering [Online]

https://securitydaily.net/ida-cong-cu-hoan-hao-cho-cac-chuyen-gia-reverse-engineering/

HẾT