

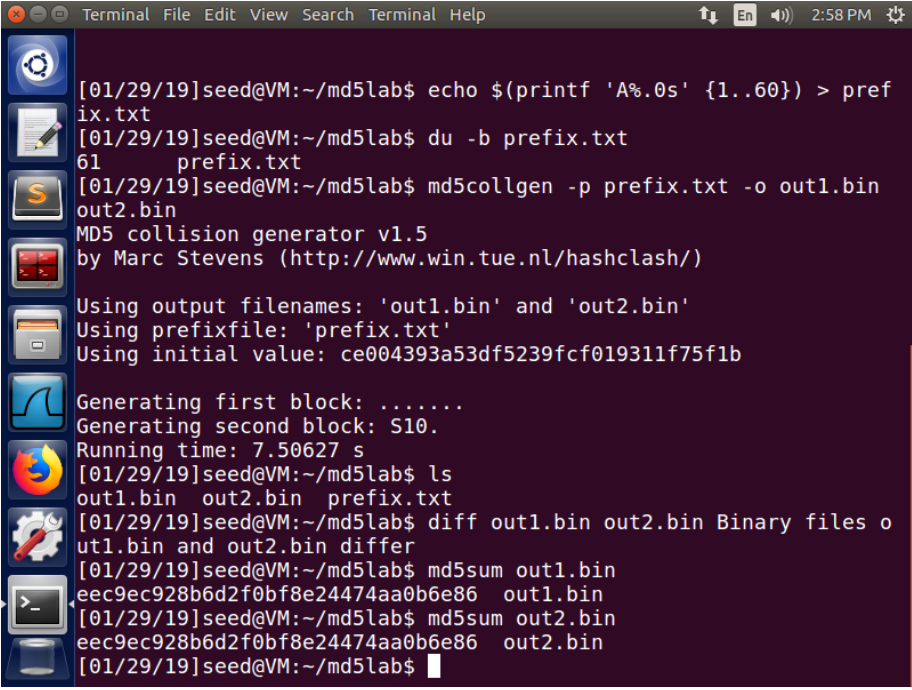
MD5 Collision Attack Lab

Task 1: Generating Two Different Files with the Same MD5 Hash

For this task, we use the following command to write arbitrary A's into the prefix.txt file.

```
$ echo $(printf 'A%.0s' {1..x}) > prefix.txt
```

Question 1. If the length of your prefix file is not multiple of 64, what is going to happen?



```
Terminal File Edit View Search Terminal Help
[01/29/19]seed@VM:~/md5lab$ echo $(printf 'A%.0s' {1..60}) > prefix.txt
[01/29/19]seed@VM:~/md5lab$ du -b prefix.txt
61      prefix.txt
[01/29/19]seed@VM:~/md5lab$ md5collgen -p prefix.txt -o out1.bin out2.bin
MD5 collision generator v1.5
by Marc Stevens (http://www.win.tue.nl/hashclash/)

Using output filenames: 'out1.bin' and 'out2.bin'
Using prefixfile: 'prefix.txt'
Using initial value: ce004393a53df5239fcf019311f75f1b

Generating first block: .....
Generating second block: S10.
Running time: 7.50627 s
[01/29/19]seed@VM:~/md5lab$ ls
out1.bin  out2.bin  prefix.txt
[01/29/19]seed@VM:~/md5lab$ diff out1.bin out2.bin Binary files out1.bin and out2.bin differ
[01/29/19]seed@VM:~/md5lab$ md5sum out1.bin
eec9ec928b6d2f0bf8e24474aa0b6e86  out1.bin
[01/29/19]seed@VM:~/md5lab$ md5sum out2.bin
eec9ec928b6d2f0bf8e24474aa0b6e86  out2.bin
[01/29/19]seed@VM:~/md5lab$
```

As shown in the screenshot, the prefix file has 60 A's. Although out1.bin are different from out2.bin, they have the same md5 hash string.

Question 2. Create a prefix file with exactly 64 bytes, and run the collision tool again, and see what happens.

```

Terminal File Edit View Search Terminal Help
[01/29/19]seed@VM:~/md5lab$ echo $(printf 'A%.0s' {1..64}) > prefix.txt
[01/29/19]seed@VM:~/md5lab$ md5collgen -p prefix.txt -o out1.bin out2.bin
MD5 collision generator v1.5
by Marc Stevens (http://www.win.tue.nl/hashclash/)

Using output filenames: 'out1.bin' and 'out2.bin'
Using prefixfile: 'prefix.txt'
Using initial value: e2c2e7921e22752b0725a7de05ac896e

Generating first block: .....
Generating second block: S00.....
Running time: 3.05636 s
[01/29/19]seed@VM:~/md5lab$ diff out1.bin out2.bin
Binary files out1.bin and out2.bin differ
[01/29/19]seed@VM:~/md5lab$ md5sum out1.bin
e65d2011fdf70a34765ba5b75b0c5863 out1.bin
[01/29/19]seed@VM:~/md5lab$ md5sum out2.bin
e65d2011fdf70a34765ba5b75b0c5863 out2.bin
[01/29/19]seed@VM:~/md5lab$

```

As shown in the screenshot, out1.bin is still different from out2.bin, but they have the same hash string.

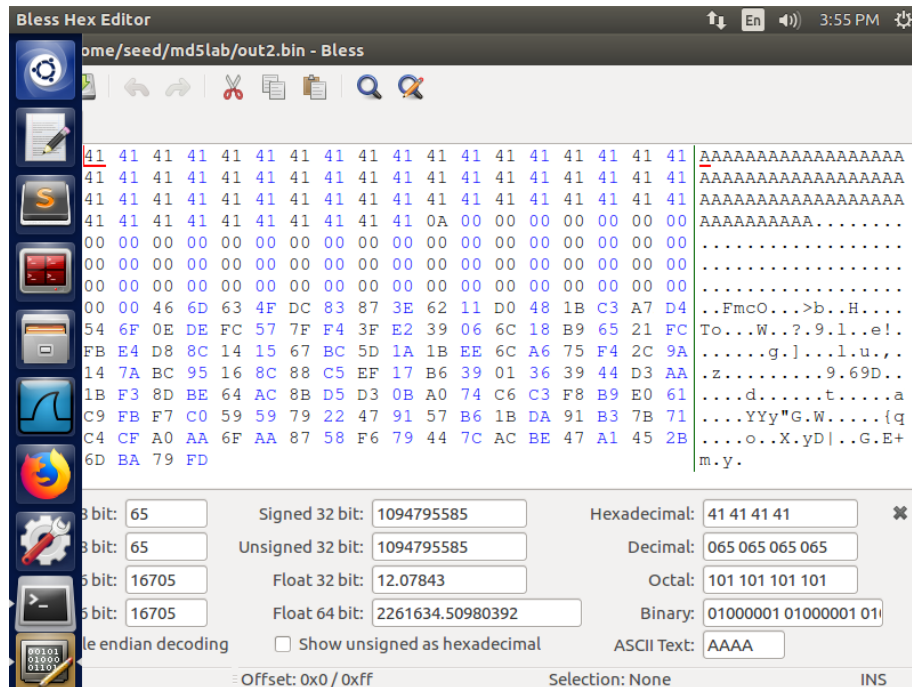
Question 3. Are the data (128 bytes) generated by md5collgen completely different for the two output files? Please identify all the bytes that are different.

The screenshot shows the Bless Hex Editor interface. The main window displays the hex dump of the file 'out1.bin'. The first 128 bytes are highlighted in red. The hex dump shows a series of '41' characters followed by a series of '00' characters, and then a series of characters that form the ASCII string 'AAAAA...Fmco...>b..H...To.^..W..?.9..l..e!..g..].n..l..u...z...E...9..69D...d...t..C...a...YYy"G..W....{q...*p...X..yD|..G..E..m..y.'

The bottom panel shows the conversion of the selected hex data to various formats:

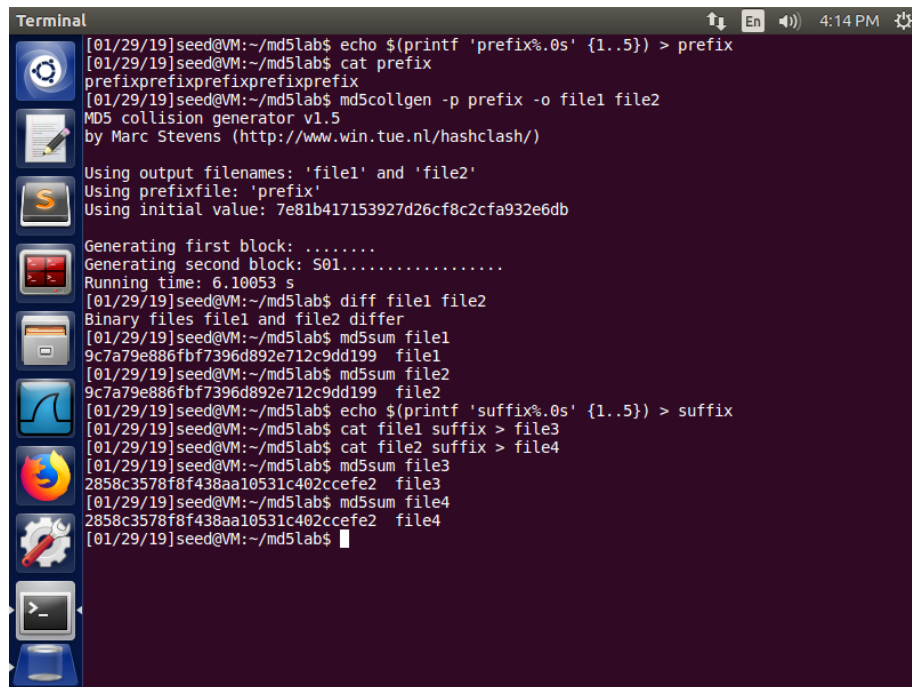
- 8 bit: 65
- Signed 32 bit: 1094795585
- Hexadecimal: 41 41 41 41
- 8 bit: 65
- Unsigned 32 bit: 1094795585
- Decimal: 065 065 065 065
- 6 bit: 16705
- Float 32 bit: 12.07843
- Octal: 101 101 101 101
- 6 bit: 16705
- Float 64 bit: 2261634.50980392
- Binary: 01000001 01000001 01
- ASCII Text: AAAA

Additional options include 'little endian decoding', 'Show unsigned as hexadecimal', 'Offset: 0x0 / 0xff', 'Selection: None', and 'INS'.



As shown in `bless`, different bytes are located 93, AD, BB, F7, F8, FB.

Task 2: Understanding MD5's Property



As shown in the screenshot, we generate two different files with the same MD5 hash string and concatenate with same suffix. The new files, file3 and file4, also have the same MD5 hash string.

Task 3 : Generating two executable files with the same MD5 hash

Sample code from guide:



```

#include <stdio.h>

unsigned char xyz[200] = {
    0x41, 0x41, 0x41, 0x41, 0x41, 0x41,
    ... (omitted) ...
    0x41, 0x41, 0x41, 0x41, 0x41, 0x41,
};

int main()
{
    int i;
    for (i=0; i<200; i++){
        printf("%x", xyz[i]);
    }
    printf("\n");
}

```

a.out - GHex 5:10 PM

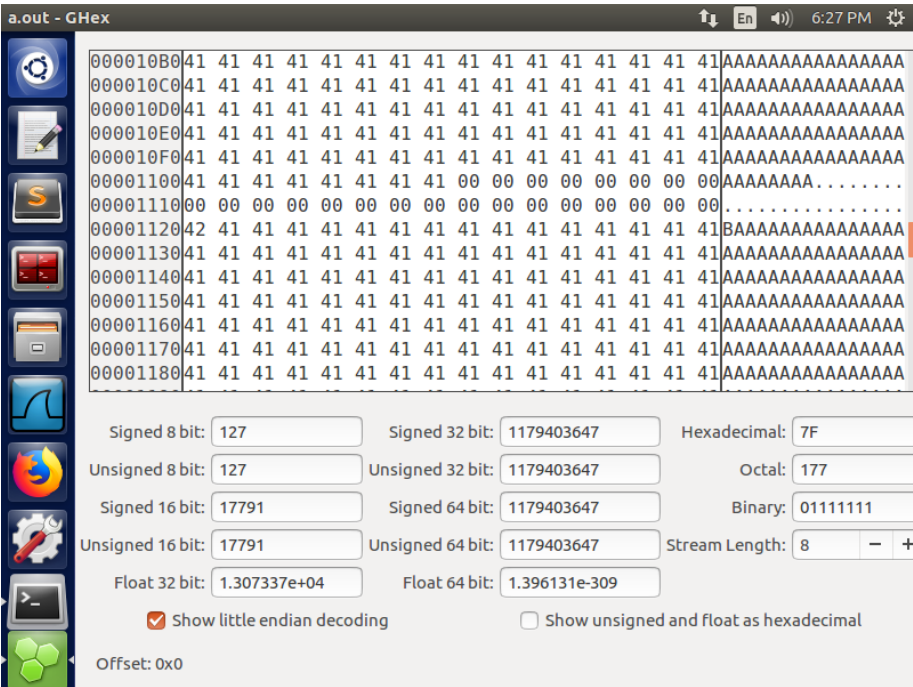
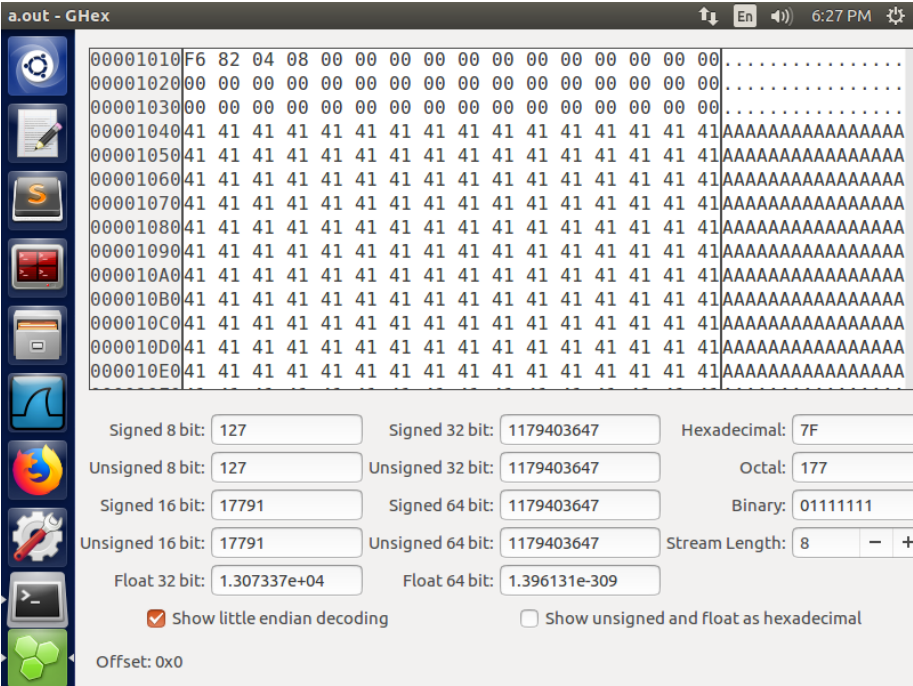
00001020	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00001030	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00001040	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
00001050	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
00001060	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
00001070	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
00001080	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
00001090	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
000010A0	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
000010B0	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
000010C0	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
000010D0	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
000010E0	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA
000010F0	41 41 41 41 41 41 41 41 41 41 41 41 41 41 41	AAAAAAAAAAAAAAAA

Signed 8 bit:	127	Signed 32 bit:	1179403647	Hexadecimal:	7F
Unsigned 8 bit:	127	Unsigned 32 bit:	1179403647	Octal:	177
Signed 16 bit:	17791	Signed 64 bit:	1179403647	Binary:	01111111
Unsigned 16 bit:	17791	Unsigned 64 bit:	1179403647	Stream Length:	8 - +
Float 32 bit:	1.307337e+04	Float 64 bit:	1.396131e-309		

☒ Show little endian decoding
 ☐ Show unsigned and float as hexadecimal

Offset: 0x0


```
        break;
    }
}
if(flag)
    printf("run benign code\n");
else
    printf("run malicious code\n");
return 0;
}
```




```
Terminal
[01/29/19]seed@VM:~/md5lab$ head -c 8 oldsuffix > arrayend
[01/29/19]seed@VM:~/md5lab$ cat file1 arrayend > file1comp
[01/29/19]seed@VM:~/md5lab$ cat file2 arrayend > file2comp
[01/29/19]seed@VM:~/md5lab$ tail -c +9 oldsuffix > suffix
[01/29/19]seed@VM:~/md5lab$ head -c 24 suffix > intermediate
[01/29/19]seed@VM:~/md5lab$ tail -c +25 > array2start
^Z
[2]+  Stopped                  tail -c +25 > array2start
[01/29/19]seed@VM:~/md5lab$ ;c
bash: syntax error near unexpected token `;'
[01/29/19]seed@VM:~/md5lab$ lc
The program 'lc' is currently not installed. You can install it by typing:
sudo apt install mono-devel
[01/29/19]seed@VM:~/md5lab$ ls
a.out      arrayend  file1comp  file2comp  oldsuffix  suffix
array2start file1      file2      intermediate prefix      task4.c
[01/29/19]seed@VM:~/md5lab$ rm array2start
[01/29/19]seed@VM:~/md5lab$ tail -c +25 suffix > array2start
[01/29/19]seed@VM:~/md5lab$ cat file1comp intermediate > file1inter
[01/29/19]seed@VM:~/md5lab$ cat file2comp intermediate > file2inter
[01/29/19]seed@VM:~/md5lab$ tail -c +4161 file1comp > file1array
[01/29/19]seed@VM:~/md5lab$ tail -c +201 array2start > suffix
[01/29/19]seed@VM:~/md5lab$ cat file1inter file1array suffix > bin1
[01/29/19]seed@VM:~/md5lab$ cat file2inter file1array suffix > bin2
[01/29/19]seed@VM:~/md5lab$ chmod +x bin1
[01/29/19]seed@VM:~/md5lab$ chmod +x bin2
[01/29/19]seed@VM:~/md5lab$ ./bin1
run benign code
[01/29/19]seed@VM:~/md5lab$ ./bin2
run malicious code
[01/29/19]seed@VM:~/md5lab$ md5sum bin1
572842db18cbcc39f33a0c531a71b6e5  bin1
[01/29/19]seed@VM:~/md5lab$ md5sum bin2
572842db18cbcc39f33a0c531a71b6e5  bin2
[01/29/19]seed@VM:~/md5lab$
```

From the first screenshot, we notice that the array **a** start at offset 1040. So we first cut the file until the first 64 bytes of array **a** and generate 128 bytes using `md5collgen` command.

```
$ head -c 4224 a.out > prefix
$ md5collgen -p prefix -o file1 file2
```

Now we need to fill the rest part, we first complete array **a**:

```
$ tail -c +4353 a.out > oldsuffix
$ head -c 8 oldsuffix > arrayend
$ cat file1 arrayend > file1comp
$ cat file2 arrayend > file2comp
```

Then we fill the gap between array **a** and **b**:

```
$ tail -c +9 oldsuffix > suffix
$ head -c 24 suffix > intermediate
$ tail -c +25 suffix > array2start
$ cat file1comp intermediate > file1inter
$ cat file2comp intermediate > file2inter
```

Then we replace array **b** with the array in file1 and wrap up:

```
$ tail -c +4161 file1comp > file1array
$ tail -c +201 array2start > suffix
$ cat file1inter file1array suffix > bin1
$ cat file2inter file1array suffix > bin2
```

As shown in the last screenshot, `bin1` and `bin2` have the same value, but they execute different branches.