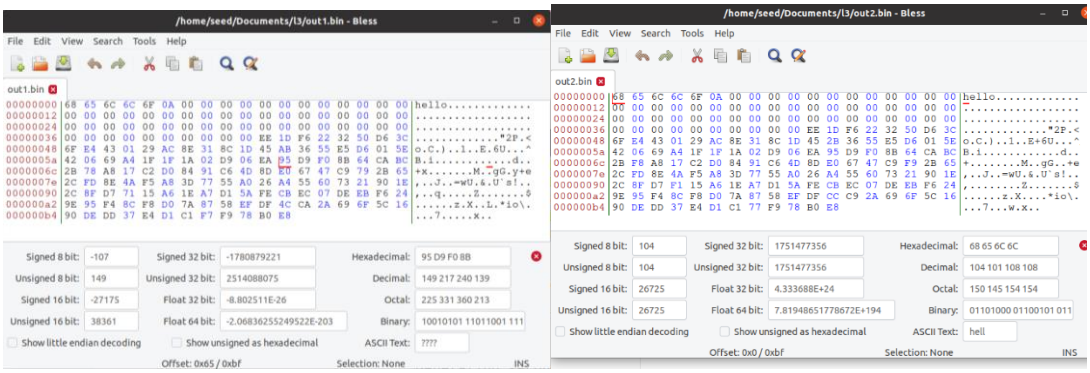
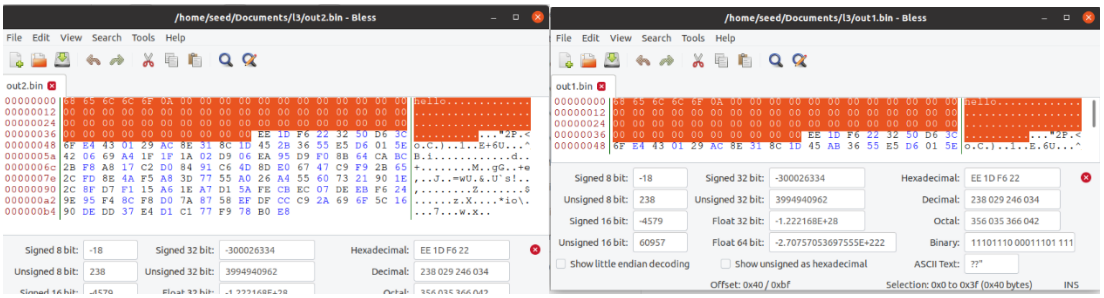
Task1

```
seed@VM: ~/.../l3
[10/24/22]seed@VM:~/.../l3$ python3 -c "print('hello')" > prefix.txt
[10/24/22]seed@VM:~/.../l3$ ls -ld prefix.txt
-rw-rw-r-- 1 seed seed 6 Oct 24 18:51 prefix.txt
[10/24/22]seed@VM:~/.../l3$ 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: 91e1e88548a4bef54d09a7c23bd03682

Generating first block: ..........
Generating second block: S01.
Running time: 7.372 s
[10/24/22]seed@VM:~/.../l3$ diff out1.bin out2.bin
Binary files out1.bin and out2.bin differ
[10/24/22]seed@VM:~/.../l3$ md5sum out1.bin out2.bin
78356f7607c8219082002005c97a54fd  out1.bin
78356f7607c8219082002005c97a54fd  out2.bin
[10/24/22]seed@VM:~/.../l3$
```

First, I follow the instruction to build to two files who have the same MD5 hash value and save them into out1.bin and out2.bin. Using “diff out1.bin out2.bin” command, the output shows that they have the different content. I use binary editor to view them, and it can be noticed that the output is different. At the same time, the output of md5sum is same, which means they have the same hash value.



Q1:

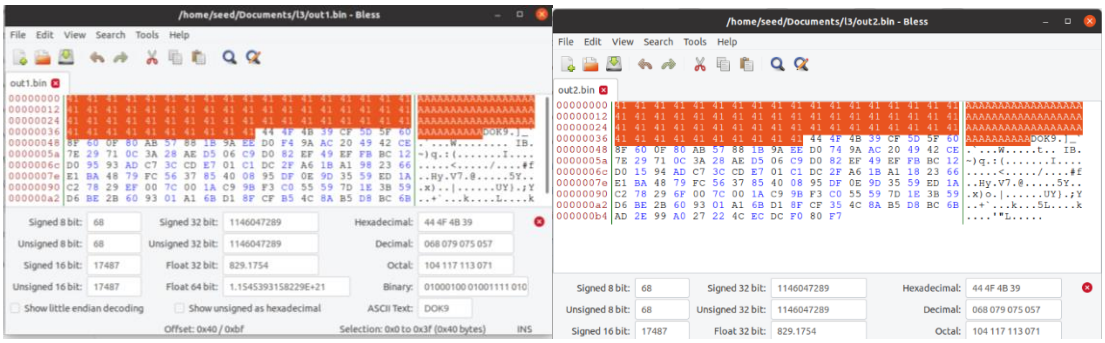


If the length of your prefix file is not multiple of 64, the space will be padded by 0 to make sure the prefix file size are the multiple of 64. It can be observed at the pictures as above.

Q2:

```
seed@VM: ~
[10/24/22]seed@VM:~$ python3 -c "print('A'*64,end='')" > prefix.txt
[10/24/22]seed@VM:~$ 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: b217e7185a63fe5f643fe6a6d401bf59

Generating first block: ..............................
Generating second block: S00..............
Running time: 15.9975 s
[10/24/22]seed@VM:~$ diff out1.bin out2.bin
Binary files out1.bin and out2.bin differ
[10/24/22]seed@VM:~$
```

Build two files with 64 bytes prefix file. And the outputs shown as follows. It can be noticed that prefix file is not padded because the file size is 64 bytes.

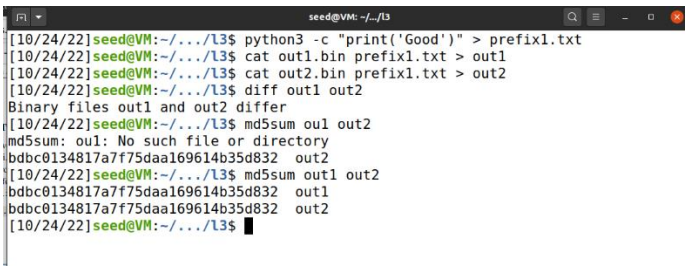


Q3:

```
seed@VM: ~/.../l3
[10/24/22]seed@VM:~/.../l3$ hexdump out1.bin >> out1
[10/24/22]seed@VM:~/.../l3$ hexdump out2.bin >> out2
[10/24/22]seed@VM:~/.../l3$ diff out1 out2
4,6c4,6
< 0000050 ee9a f4d0 ac9a 4920 ce42 297e 0c71 283a
< 0000060 d5ae c906 82d0 49ef fbef 12bc 95d0 ad93
< 0000070 3cc7 e7cd c101 2fdc 1ba6 98a1 6623 bae1
---
> 0000050 ee9a 74d0 ac9a 4920 ce42 297e 0c71 283a
> 0000060 d5ae c906 82d0 49ef fbef 12bc 15d0 ad94
> 0000070 3cc7 e7cd c101 2fdc 1ba6 18a1 6623 bae1
8,10c8,10
< 0000090 78c2 ef29 7c00 1a00 9bc9 c0f3 5955 1e7d
< 00000a0 593b bed6 602b 0193 6ba1 8fd1 b5cf 8a4c
< 00000b0 d8b5 6bbc 2ead a099 2227 6c4c f0dc f780
---
> 0000090 78c2 6f29 7c00 1a00 9bc9 c0f3 5955 1e7d
> 00000a0 593b bed6 602b 0193 6ba1 8fd1 35cf 8a4c
> 00000b0 d8b5 6bbc 2ead a099 2227 ec4c f0dc f780
15,17c15,17
< 0000050 ee9a f4d0 ac9a 4920 ce42 297e 0c71 283a
< 0000060 d5ae c906 82d0 49ef fbef 12bc 95d0 ad93
< 0000070 3cc7 e7cd c101 2fdc 1ba6 98a1 6623 bae1
---
> 0000050 ee9a 74d0 ac9a 4920 ce42 297e 0c71 283a
> 0000060 d5ae c906 82d0 49ef fbef 12bc 15d0 ad94
> 0000070 3cc7 e7cd c101 2fdc 1ba6 18a1 6623 bae1
19,21c19,21
< 0000090 78c2 ef29 7c00 1a00 9bc9 c0f3 5955 1e7d
< 00000a0 593b bed6 602b 0193 6ba1 8fd1 b5cf 8a4c
< 00000b0 d8b5 6bbc 2ead a099 2227 6c4c f0dc f780
---
> 0000090 78c2 6f29 7c00 1a00 9bc9 c0f3 5955 1e7d
> 00000a0 593b bed6 602b 0193 6ba1 8fd1 35cf 8a4c
> 00000b0 d8b5 6bbc 2ead a099 2227 ec4c f0dc f780
[10/24/22]seed@VM:~/.../l3$
```

It is not complete different, use diff command to check the different parts. The different can be noticed as above.

Task 2

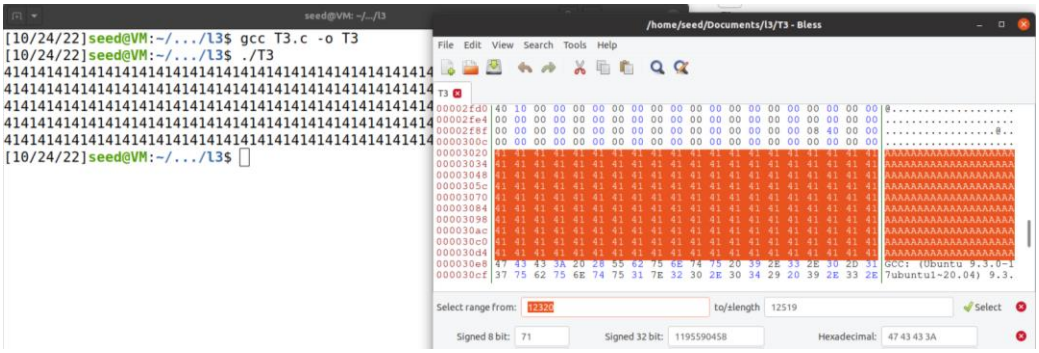


I built out1.bin and out2.bin in task 1, and MD5(out1.bin) == MD5(out2.bin). I build a new file and cat it with out1.bin and out2. bin. Then I observed that the MD5 value also same. It can be proofed that the hash value will also be same after two files have the same hash value concatenate same content.

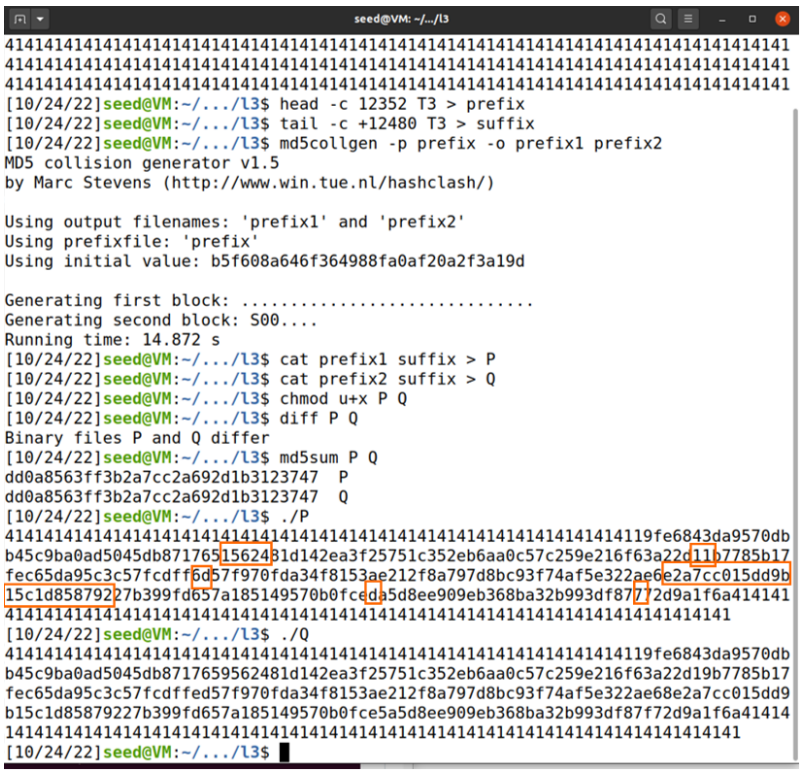
Task 3

```
#include <stdio.h>

unsigned char xyz[200] = {
   "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
   "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
   "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
   "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"};

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

I modify the C code as above, and gcc it gets T3. Open T3 file and can be find the array are saving at 0x3020 that the decimal is 12320.



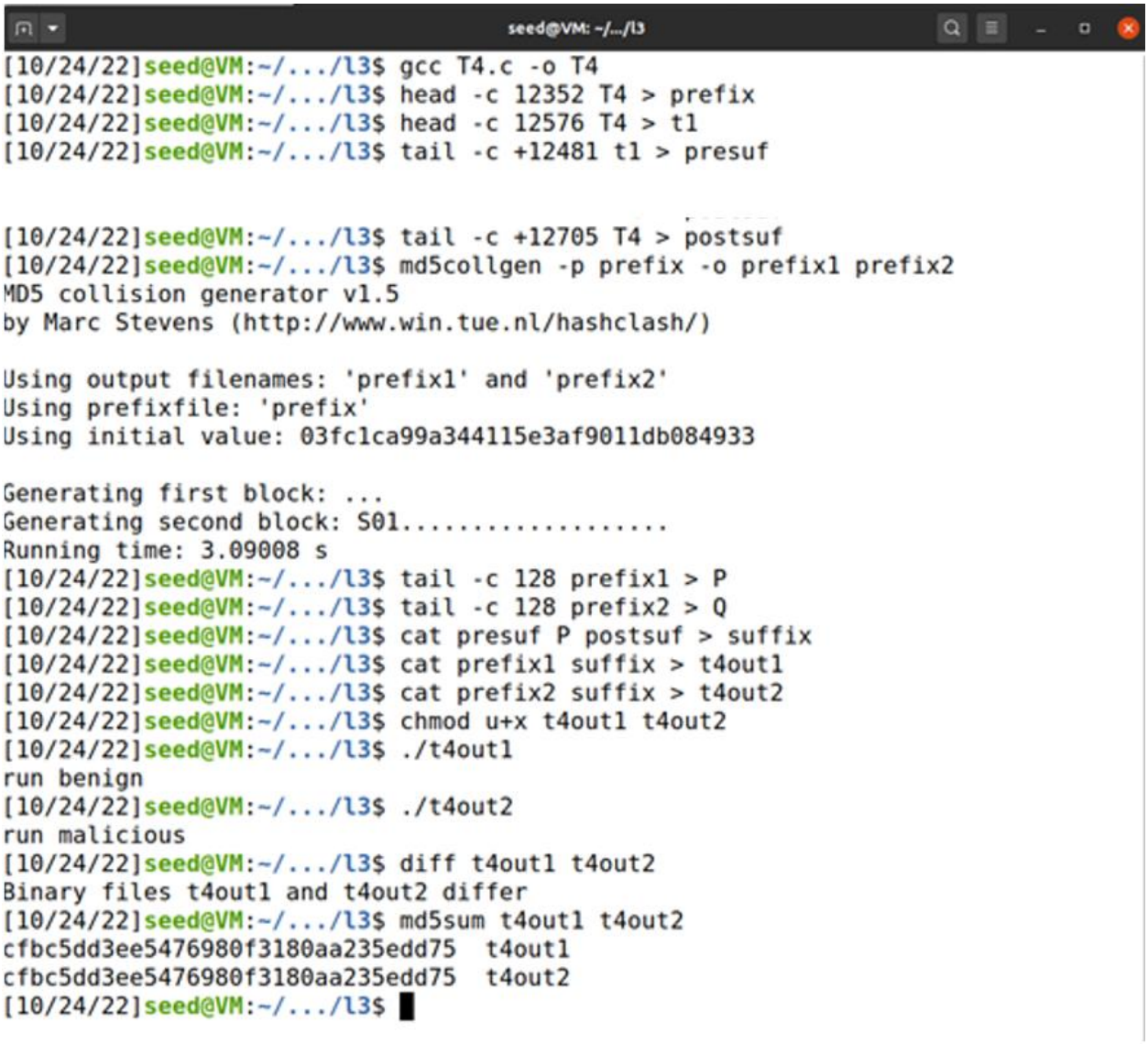
We need the size of prefix file is multiply 64 bytes, and 12352 can be divided by 64. Therefore, I split the first 12352 bytes to prefix, and split from the 12352+128 bytes to the end of the file to suffix. I use prefix to build two files “prefix1” and “prefix2” that have the same MD5 hash value. Then cat prefix1 and suffix to P, cat prefix2 and suffix to Q. Finally, verify P and Q. The hash value of P and Q are same, and P and Q can be executable. However, their outputs are not same.



T4

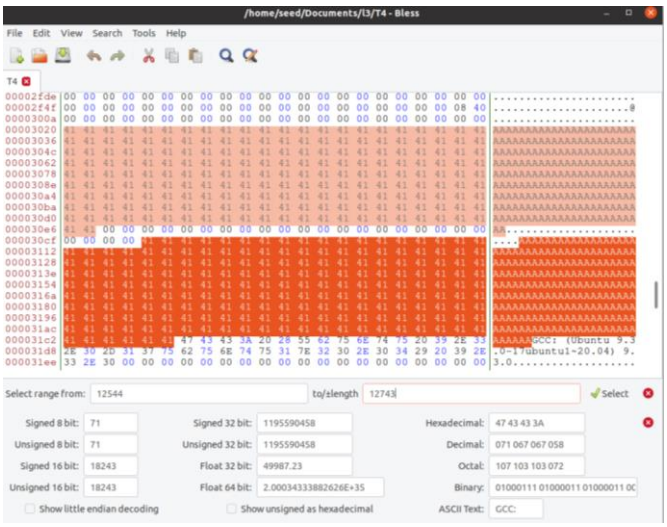
```
#include <stdio.h>


unsigned char X[200] = {

   "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"

   "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
```

```
    "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
    "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
    };

unsigned char Y[200] ={
    "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
    "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
    "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
    "AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA"
    };

int main(){
    int t =1;
    for (int i = 0; i < 200; i++){
        if (X[i] != Y[i]){
            t = 0;
        }
    }
    if(t==0){
            printf("run malicious\n");
    }else{
            printf("run benign\n");
    }
    return 0;
}
```

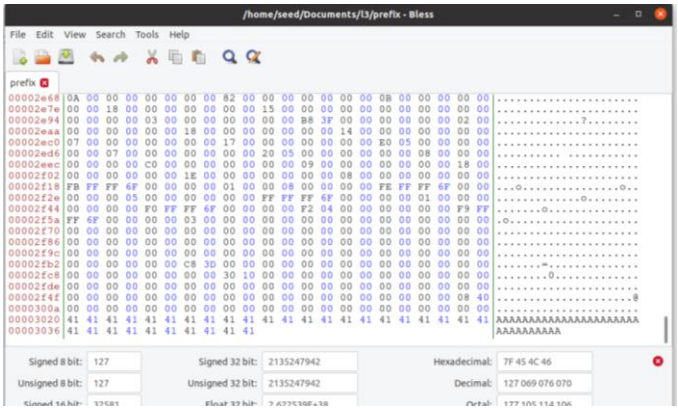


This is the whole process of Task4.

Compile T4.c and open T4. Find Array X and Array Y. It can be noticed that Array X starts at 12320 Bytes, Array Y starts at 12544 Bytes.

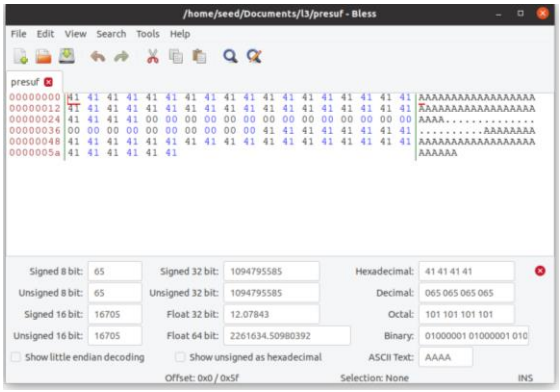


First, I split the prefix first, I choose the same size in Task 3, which include 32 bytes “A”.

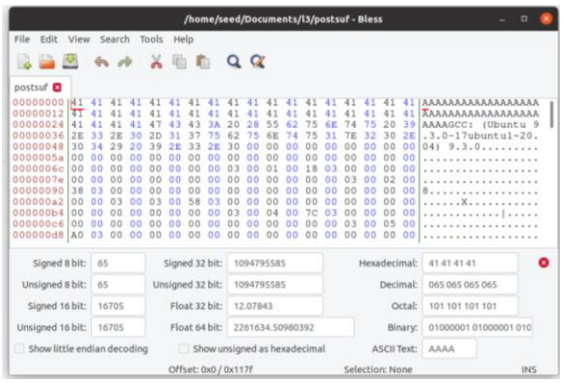


To make sure X==Y in the first version, array Y should be split into first 32 bytes and last (200-128-32 = 40) Bytes.

Therefore, I first split T4 at (12544+32 = 12576) bytes and save content in t1, and split t1 at (12352+128+1 = 12481) bytes and save the content into presuf.

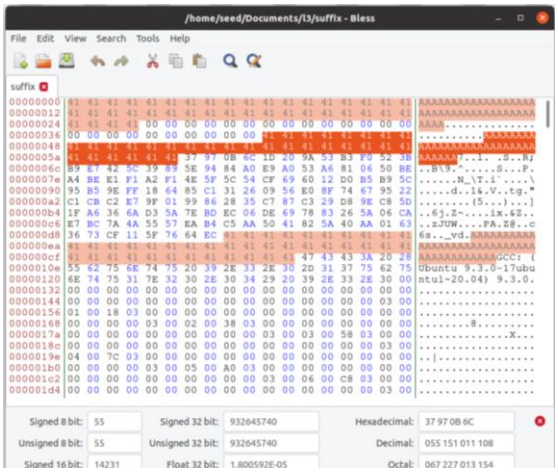


Then, I split T4 at (12576+128+1=12705)bytes and save the content in postsuf.



The next step is finding P and Q. I use prefix to get two new file “prefix1” and “prefix2” who have the same hash value. Based on the Task1, P and Q should be the last 128 bytes of “prefix1” and “prefix2”, because the prefix is the multiply of 64 bytes. Split last 128 bytes into “prefix1” and “prefix2” and get P and Q.

Then, I cat presuf, P, postsuf to get suffix.



Then, cat prefix1, suffix to get t4out1; cat prefix2, suffix to get t4out2.

Finally, verify t4out1, t4out2. The hash value of t4out1 and t4out2 are same, and t4out1 and t4out2 can be executable. However, their outputs are not same. T4out1 output that X and Y are same, and t4out2 output that X and Y are different.