姓名:蔡佩蓉 學號:109511286

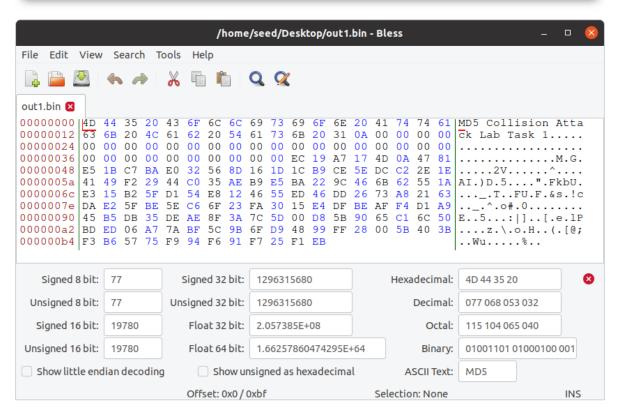
MD5 Collision Attack Lab (Lab4)

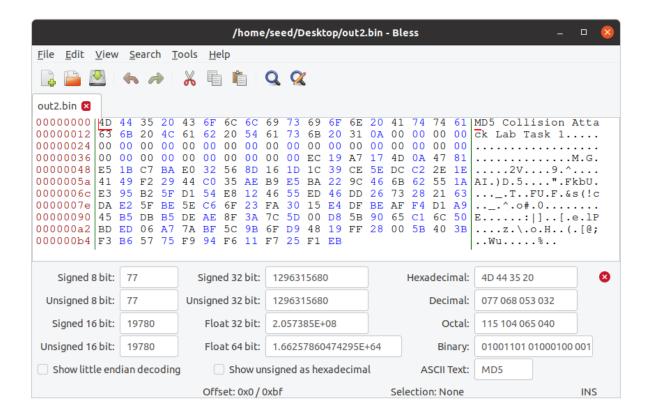
Task 1: Generating Two Different Files with the Same MD5 Hash

Questions:

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

seed@VM: -/Desktop Q ≡ _ □ 😣
[05/16/24]seed@VM:~/Desktop\$ echo "MD5 Collision Attack Lab Task 1" > prefix.txt [05/16/24]seed@VM:~/Desktop\$ md5collgen -p prefix.txt -o out1.bin out2.bin MD5 collision generator v1.5 by Marc Stevens (http://www.win.tue.nl/hashclash/)
by Marc Stevens (http://www.win.tue.ht/hashctash/)
Using output filenames: 'out1.bin' and 'out2.bin'
Using prefixfile: 'prefix.txt' Using initial value: 2ec4f2fabd5e01c34fc8ee7bcad4338a
Generating first block:
Generating second block: S11
Running time: 78.5462 s
[05/16/24]seed@VM:~/Desktop\$ diff outl.bin out2.bin
Binary files out1.bin and out2.bin differ
[05/16/24]seed@VM:~/Desktop\$ md5sum out1.bin
965c2e2bff496325a3e75f4ce3796158 out1.bin
[05/16/24]seed@VM:~/Desktop\$ md5sum out2.bin
965c2e2bff496325a3e75f4ce3796158 out2.bin
[05/16/24]seed@VM:~/Desktop\$

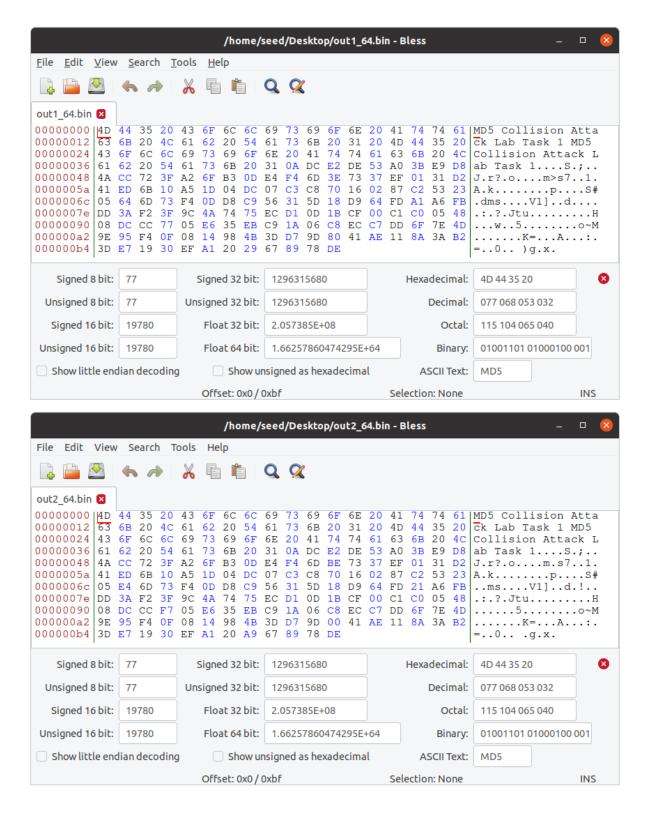




If the length of the prefix file is not a multiple of 64 bytes, the md5collgen tool will add padding to the prefix (padded with zeros). MD5 operates on 512-bit (64-byte) blocks, so the tool ensures that the prefix (plus padding) aligns to a 64-byte boundary before appending the collision blocks.

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

```
seed@VM: ~/Desktop
                                                                                                     Q =
[05/16/24]seed@VM:~/Desktop$ echo "MD5 Collision Attack Lab Task 1 MD5 Collision Attack Lab Task 1" > prefix64.txt
[05/16/24]seed@VM:~/Desktop$ md5collgen -p prefix64.txt -o out1_64.bin out2_64.bin
MD5 collision generator v1.5
by Marc Stevens (http://www.win.tue.nl/hashclash/)
Using output filenames: 'out1_64.bin' and 'out2_64.bin'
Using prefixfile: 'prefix64.txt
Using initial value: 90a77ba13e4ba412473710293110f4fe
Generating first block:
Generating second block: S01..
Running time: 2.39742 s
[05/16/24]seed@VM:~/Desktop$ diff out1_64.bin out2_64.bin
Binary files out1_64.bin and out2_64.bin differ
[05/16/24]seed@VM:~/Desktop$ md5sum out1 64.bin
a6372a4f720ade04909b33b368506118 out1_64.bin
[05/16/24]seed@VM:~/Desktop$ md5sum out2_64.bin
a6372a4f720ade04909b33b368506118 out2_64.bin
[05/16/24]seed@VM:~/Desktop$
```



The md5collgen tool will not need to add any padding before appending the collision blocks (no zero padding). The two output files will still share the same 64-byte prefix, followed by their respective collision blocks.

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



No, not all bytes are different (the different bytes are boxed above).

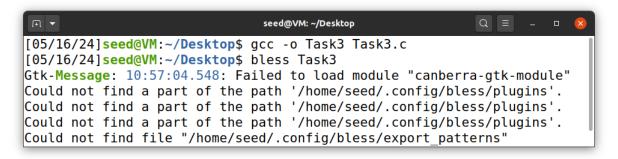
Task 2: Understanding MD5's Property

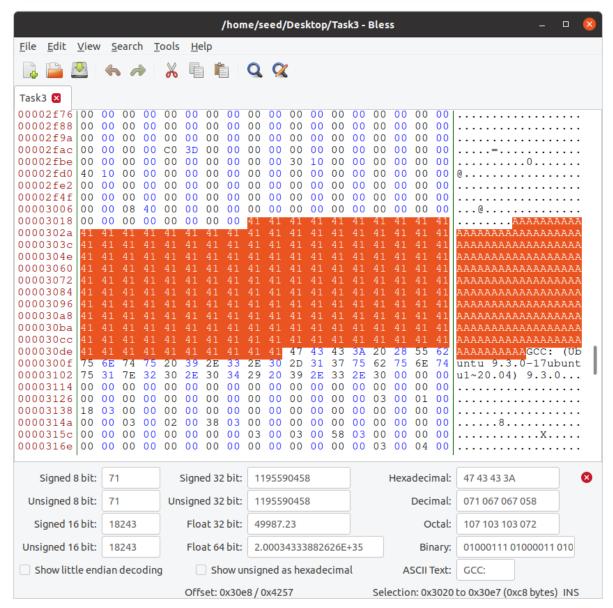
```
seed@VM: ~/Desktop
                                                   Q =
[05/16/24]seed@VM:~/Desktop$ md5collgen -p prefix.txt -o p1 p2
MD5 collision generator v1.5
by Marc Stevens (http://www.win.tue.nl/hashclash/)
Using output filenames: 'p1' and 'p2'
Using prefixfile: 'prefix.txt'
Using initial value: 2ec4f2fabd5e01c34fc8ee7bcad4338a
Generating first block: .............
Generating second block: S10...........
Running time: 16.6723 s
[05/16/24]seed@VM:~/Desktop$ diff p1 p2
Binary files p1 and p2 differ
[05/16/24]seed@VM:~/Desktop$ md5sum p1
480c66ab77e248c5b3940d36d46d628e p1
[05/16/24]seed@VM:~/Desktop$ md5sum p2
480c66ab77e248c5b3940d36d46d628e p2
[05/16/24]seed@VM:~/Desktop$ echo "abcde" > s
[05/16/24]seed@VM:~/Desktop$ cat p1 s > o1
[05/16/24] seed@VM:~/Desktop$ cat p2 s > o2
[05/16/24]seed@VM:~/Desktop$ md5sum o1
933f1188d55c50a2c3fac05a1d12c0b3 o1
[05/16/24]seed@VM:~/Desktop$ md5sum o2
933f1188d55c50a2c3fac05a1d12c0b3 o2
[05/16/24]seed@VM:~/Desktop$
```

Property of MD5: appending the same suffix to two inputs that have the same MD5 hash will result in the same hash for the concatenated outputs.

This is a fundamental property of many cryptographic hash functions and is crucial for understanding certain types of hash collisions and vulnerabilities.

Task 3: Generating Two Executable Files with the Same MD5 Hash





Length of prefix needs to be multiple of 64 bytes, 12320 (0x3020) is not multiple of 64, hence use 12352 (0x3040) [prefix = first 12352 bytes]. There is a 128-byte region, hence 12352 + 128 = 12480 (0x30C0) as suffix [from 12481st byte to the end].

```
seed@VM: ~/Desktop
                                                           Q ≡
[05/16/24]seed@VM:~/Desktop$ head -c 12352 Task3 > prefix
[05/16/24]seed@VM:~/Desktop$ md5collgen -p prefix -o a b
MD5 collision generator v1.5
by Marc Stevens (http://www.win.tue.nl/hashclash/)
Using output filenames: 'a' and 'b'
Using prefixfile: 'prefix'
Using initial value: aefe77439646539fe6db56554d46ae83
Generating first block: ...
Generating second block: S10.......
Running time: 1.58959 s
[05/16/24]seed@VM:~/Desktop$ tail -c +12481 Task3 > suffix
[05/16/24]seed@VM:~/Desktop$ cat a suffix > m
[05/16/24] seed@VM:~/Desktop$ cat b suffix > n
[05/16/24] seed@VM:~/Desktop$ chmod +x m n
[05/16/24]seed@VM:~/Desktop$ md5sum m
5487062a4869a5c7bc1523cb565eec78 m
[05/16/24] seed@VM:~/Desktop$ md5sum n
5487062a4869a5c7bc1523cb565eec78 n
[05/16/24]seed@VM:~/Desktop$
```

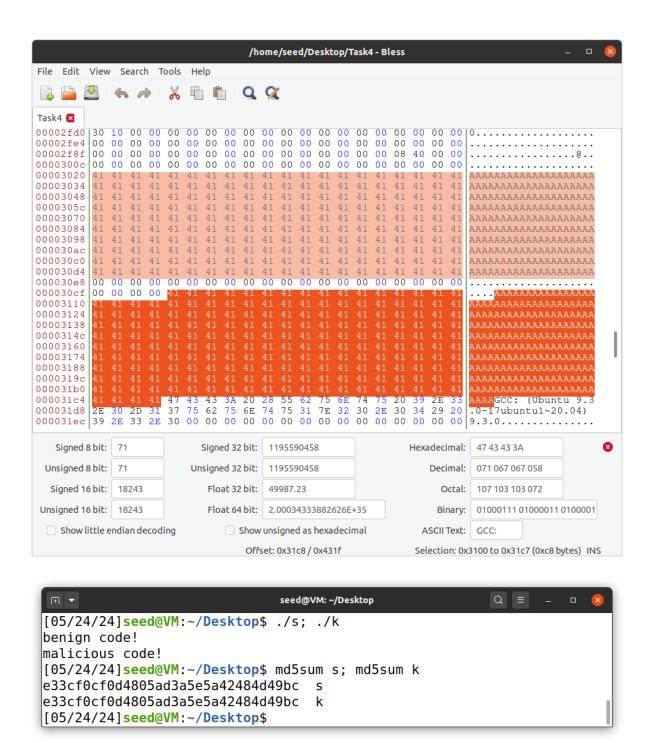
```
[05/24/24]seed@VM:~/Desktop$ echo $(./m) | md5sum;echo $(./n) | md5sum
4f3fdd9d18c1f9a01f6efd6c864c2e27 -
4a49680c53d1c507ec35c2a29b9c732f -
[05/24/24]seed@VM:~/Desktop$ md5sum m; md5sum n
5487062a4869a5c7bc1523cb565eec78 m
5487062a4869a5c7bc1523cb565eec78 n
[05/24/24]seed@VM:~/Desktop$
```

We observed that both have different outcomes (different content), but same hash value.

Task 4: Making the Two Programs Behave Differently

```
seed@VM: ~/Desktop
                                                         Q =
[05/24/24]seed@VM:~/Desktop$ gcc Task4.c -o Task4
[05/24/24]seed@VM:~/Desktop$ head -c 12352 Task4 > prefix
[05/24/24]seed@VM:~/Desktop$ md5collgen -p prefix -o p q
MD5 collision generator v1.5
by Marc Stevens (http://www.win.tue.nl/hashclash/)
Using output filenames: 'p' and 'q'
Using prefixfile: 'prefix'
Using initial value: 3287cc4fd721c212b60f5b11195c375d
Generating first block: .......
Generating second block: W.....
Running time: 2.91007 s
[05/24/24]seed@VM:~/Desktop$ tail -c +12745 Task4 > suffix
[05/24/24] seed@VM:~/Desktop$ tail -c 160 p > middle
[05/24/24] seed@VM:~/Desktop$ python3 -c "print('\x00'*40)" > tmp
[05/24/24]seed@VM:~/Desktop$ head -c 40 tmp > m40
[05/24/24] seed@VM:~/Desktop$ head -c 24 tmp > m24
[05/24/24]seed@VM:~/Desktop$ cat p m40 m24 middle m40 suffix > s
[05/24/24]seed@VM:~/Desktop$ cat q m40 m24 middle m40 suffix > k
[05/24/24] seed@VM:~/Desktop$ chmod +x s
[05/24/24] seed@VM:~/Desktop$ chmod +x k
```

Get the last 160 bytes [32 (char A) + 128 (md5 padding)] from p. The length of the array is 200, hence, 40 bytes (200-160) need to be filled in to the a and b arrays. However, the difference between the two arrays is 0x3100 - 0x3020 = 0xE0 (224bytes), not 200 bytes, hence insert another 24 bytes of zero in between.



Executing both codes (s and k) returns different output, s returns benign code, k returns malicious code. However, the hash value for s and k are the same.