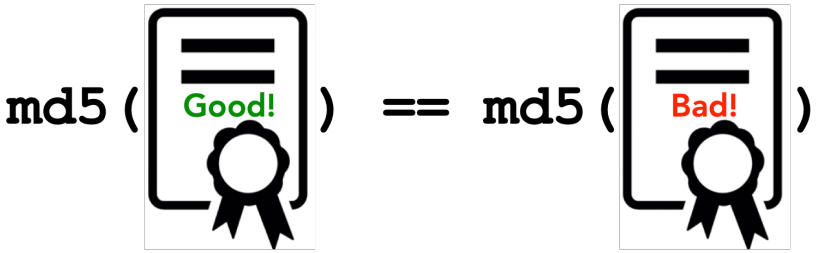
Hash Collision Attacks

- Security Impact of Collision Attacks
- Generating Two Different Files with the Same MD5 Hash
- Generating Two Programs with the Same MD5 Hash

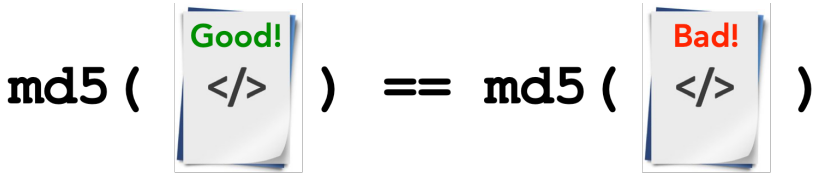
Security Impact of Collision Attacks

Forging public-key certificates



- Assume two certificate requests for www.example.com and www.attacker.com have same hash due to a collision
- CA signing of either request would be equivalent
- Attacker can get certificate signed for www.example.com without owning it!

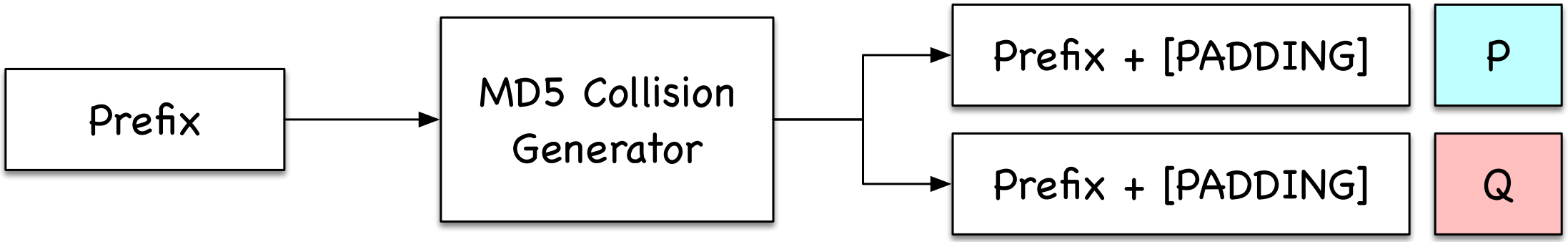
Integrity of Programs



- Ask CA to sign a legitimate program's hash
- Attacker creates a malicious program with same hash
- The certificate for legitimate program is also valid for malicious version

Generating Two Different Files w/ Same MD5 Hash

md5collgen tool generates two files with same prefix ("chosen prefix attack")



recall: md5 uses 64 byte (512 bit) blocks!

```
$ echo "Message prefix" > prefix.txt
$ md5collgen -p prefix.txt -o out1.bin out2.bin
...
```

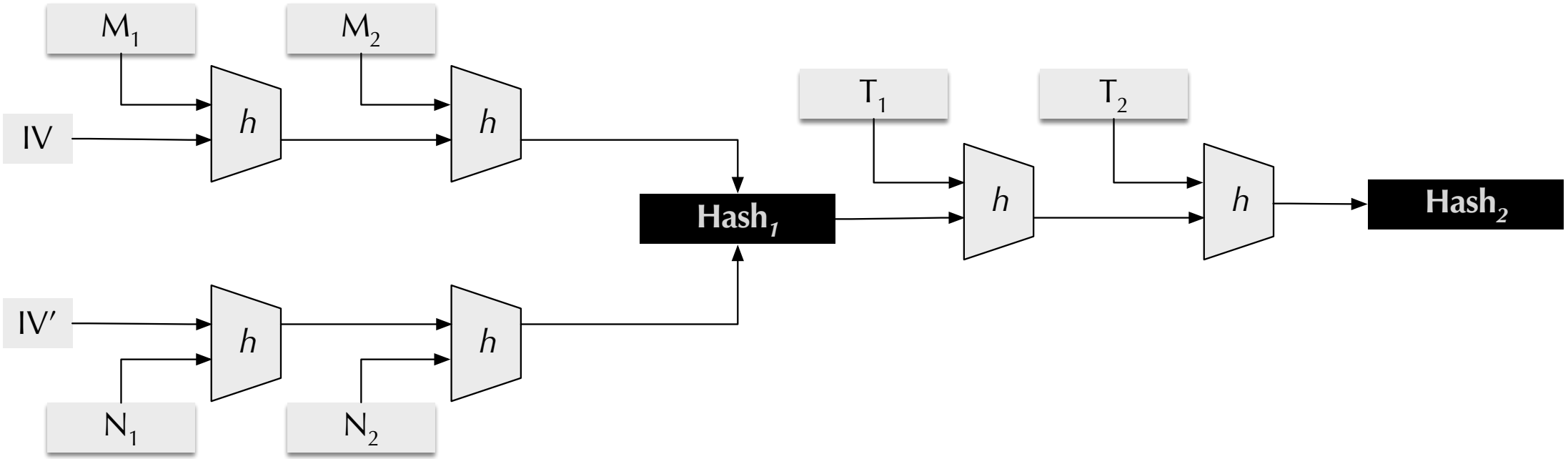
```
$ md5sum out1.bin
f53f8e097ffe4fd3710aad0fbac17123  out1.bin
$ md5sum out2.bin
f53f8e097ffe4fd3710aad0fbac17123  out2.bin
```

You can use a different hash function (e.g., SHA-256) to confirm that these files are different!

Length Extension

Generate two files with same prefix and same suffix

- Focus on MD5, SHA-1, SHA-2 using Merkle-Damgard construction
- If hash(M) = hash(N), then for any input T, hash(M || T) = hash(N || T)



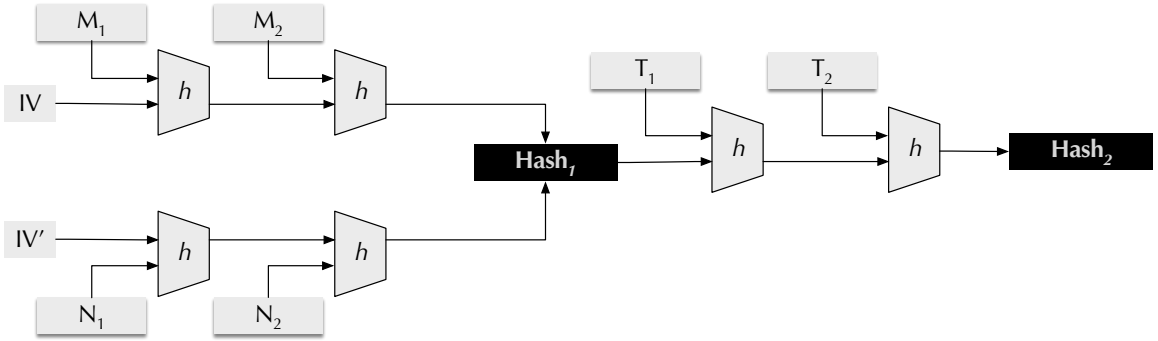
Length Extension

Example using out1.bin and out2.bin generated by md5collgen

```
$ echo "Message suffix" > suffix.txt
$ cat out1.bin suffix.txt > out1_long.bin
$ cat out2.bin suffix.txt > out2_long.bin
```

```
$ diff out1_long.bin out2_long.bin
Binary files out1_long.bin and out2_long.bin differ
```

```
$ md5sum out1_long.bin
0fbe0c2e0fc197a0f053b0640c7fd2d5  out1_long.bin
$ md5sum out2_long.bin
0fbe0c2e0fc197a0f053b0640c7fd2d5  out2_long.bin
```



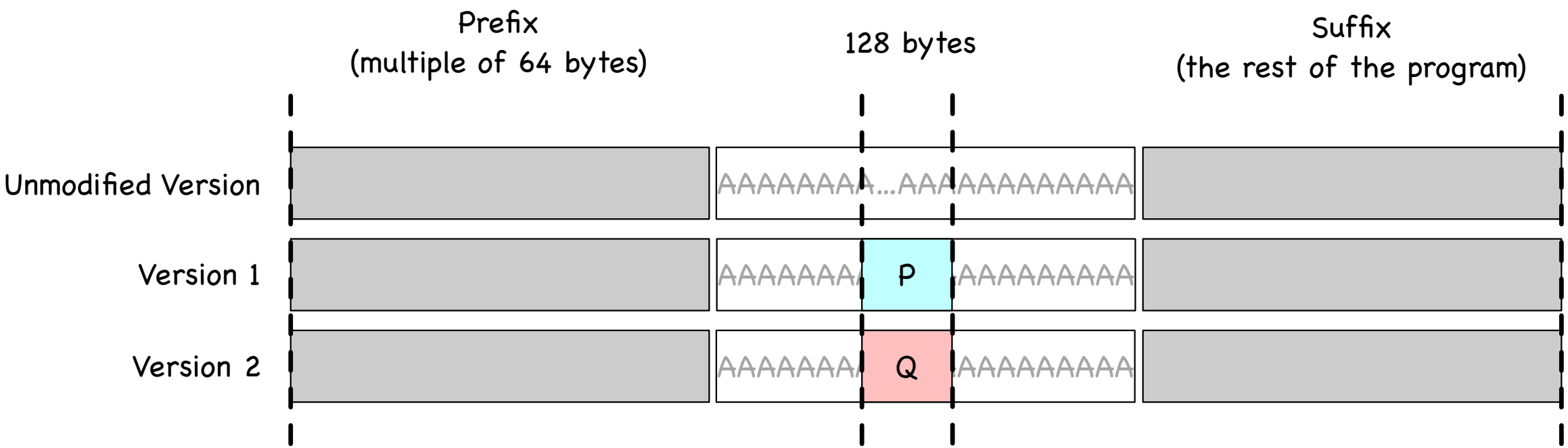
Generating Two Different Programs w/ Same MD5 Hash

Create two versions of a program with different values for the array xyz

```
$ cat print_array.c
#include <stdio.h>

unsigned char xyz[200] = { /* The contents of this array are set by you */ }

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

Generating Two **Different Programs** w/ **Same MD5 Hash**

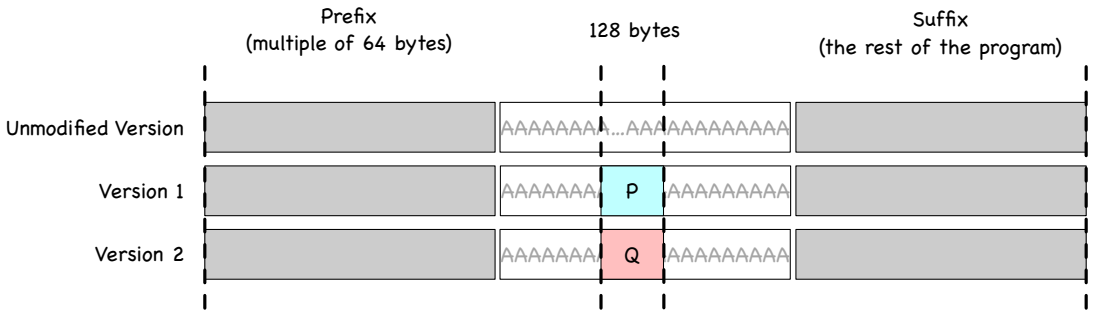
- Program will be compiled into binary (tip: fill xyz with fixed value)
- Portion of binary containing xyz will be divided into three parts



Generating Two **Different Programs** w/ **Same MD5 Hash**

Use md5collgen on prefix:

- generate two files with same hash
- last 128 bytes of each generated file is P and Q



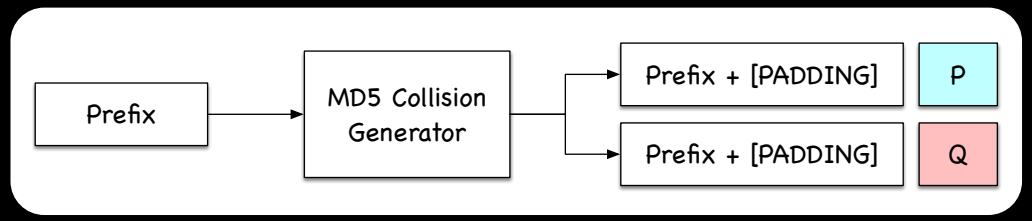
$$\text{md5}(\text{prefix} \mid\mid P) = \text{md5}(\text{prefix} \mid\mid Q)$$

$$\Downarrow$$

$$\text{md5}(\text{prefix} \mid\mid P \mid\mid \text{suffix}) = \text{md5}(\text{prefix} \mid\mid Q \mid\mid \text{suffix})$$

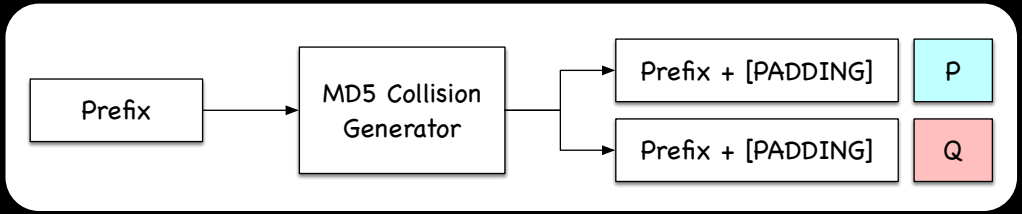
Generating Two Different Programs w/ Same MD5 Hash

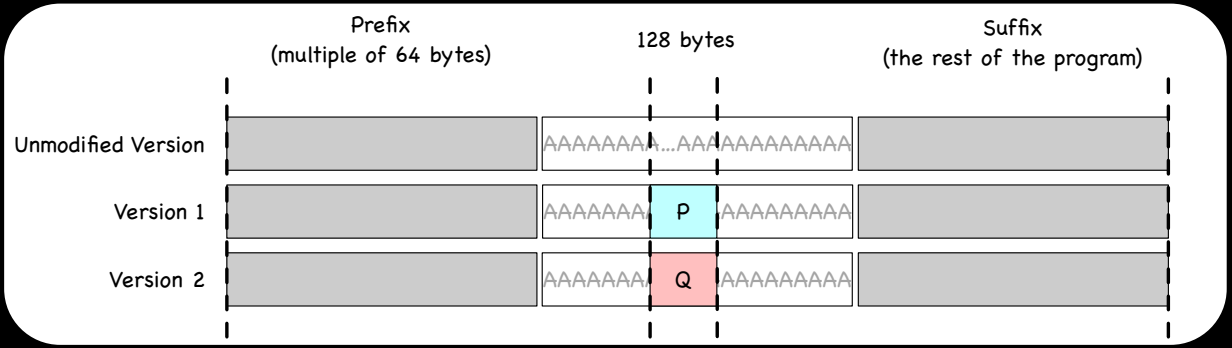
```
$ gcc print_array.c -o pa
$ ghex pa # confirm offset of start of array xyz — I see 4160
$ head -c  4160 pa > prefix
$ tail -c +4288 pa > suffix # 4160+128=4288

$ md5collgen -p prefix -o out1.bin out2.bin
$ tail -c 128 out1.bin > P
$ tail -c 128 out2.bin > Q

$ cat prefix P suffix > a1.out
$ cat prefix Q suffix > a2.out
$ chmod a+x a1.out a2.out

$ diff a1.out a2.out
Binary files a1.out and a2.out differ
$ md5sum a1.out
c09b82f44e37f7d3d32919fa7878d660  a1.out
$ md5sum a2.out
c09b82f44e37f7d3d32919fa7878d660  a2.out

$ vimdiff <(./a1.out) <(./a2.out)# can you spot the difference?!
```



Generating Two Different Programs w/ Same MD5 Hash

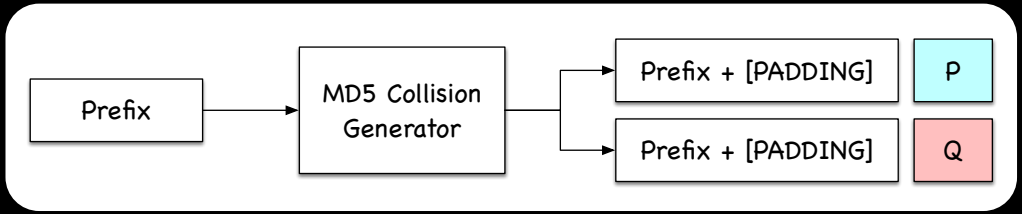
```
$ gcc print_array.c -o pa
$ ghex pa # confirm offset of start of array xyz — I see 4160
$ head -c  4160 pa > prefix
$ tail -c +4288 pa > suffix # 4160+128=4288

$ md5collgen -p prefix -o out1.bin out2.bin
$ tail -c 128 out1.bin > P
$ tail -c 128 out2.bin > Q

$ cat prefix P suffix > a1.out
$ cat prefix Q suffix > a2.out
$ chmod a+x a1.out a2.out

$ diff a1.out a2.out
Binary files a1.out and a2.out differ
$ md5sum a1.out
c09b82f44e37f7d3d32919fa7878d660  a1.out
$ md5sum a2.out
c09b82f44e37f7d3d32919fa7878d660  a2.out

$ vimdiff <(./a1.out) <(./a2.out)# can you spot the difference?!
```

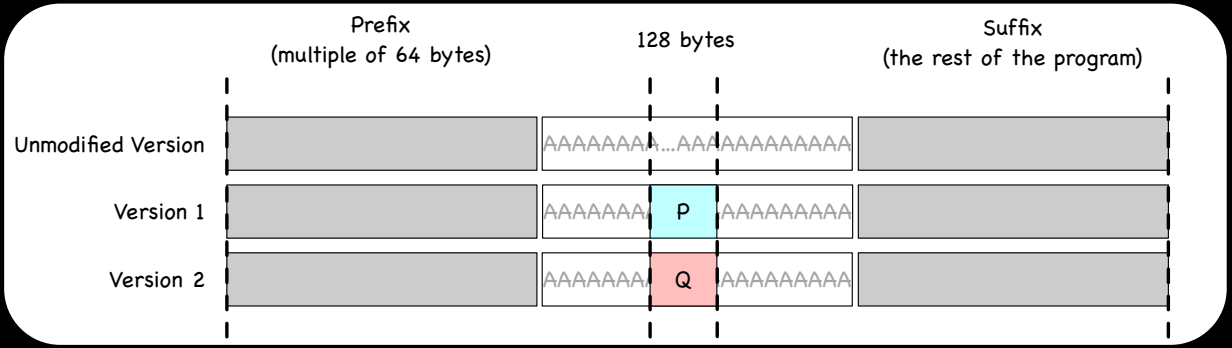




Generating Two Different Programs w/ Same MD5 Hash

```
$ gcc print_array.c -o pa
$ ghex pa # confirm offset of start of array xyz — I see 4160
$ head -c  4160 pa > prefix
$ tail -c +4288 pa > suffix # 4160+128=4288

$ md5collgen -p prefix -o out1.bin out2.bin
$ tail -c 128 out1.bin > P
$ tail -c 128 out2.bin > Q

$ cat prefix P suffix > a1.out
$ cat prefix Q suffix > a2.out
$ chmod a+x a1.out a2.out

$ diff a1.out a2.out
Binary files a1.out and a2.out differ
$ md5sum a1.out
c09b82f44e37f7d3d32919fa7878d660  a1.out
$ md5sum a2.out
c09b82f44e37f7d3d32919fa7878d660  a2.out

$ vimdiff <(./a1.out) <(./a2.out)# can you spot the difference?!
```





In the lab, you'll try this and even take it one step further :-)