CSCI 331:
Introduction to Computer Security

Lecture 11: Midterm Exam Review

Instructor: Dan Barowy

Williams

Announcements

- Next lab: meet in lobby of Jesup. Do not be late! We will leave promptly at the start of lab.
- CS Colloquium: Allison Koenecke @ MSR/Cornell
  Friday, 2:30pm in Wege Auditorium
  “Racial Disparities in Automatic Speech Recognition”

What topics?

Think about which topics you feel confused about. Take a few minutes and write them down on a piece of paper.

Everybody needs to tell me something.

Things we’ve covered

The C Programming Language

The C Programming Language

Basics

- Compilation using `gcc`.
- Warnings using `-Wall`
- Programs vs libraries
  - Build program with `-o` and specify name
  - Build library with `-c`

The C Programming Language

C Features

- The pointer as the basic unit of abstraction.
- `struct` as the basic unit of grouping.
- `typedef` as a way to give types useful names.
- Printing using `printf` and format specifiers.
- Memory as a resource that must be manually managed
  - Automatic (“local”) memory, allocated on the stack
  - Manual memory, allocated on the heap using `malloc`.

The C Programming Language

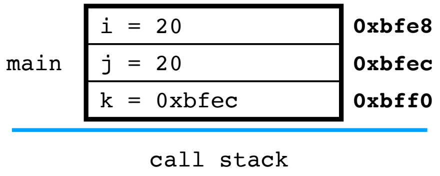
C Rules

0. Pointers are for **referring to** locations in **memory**.
1. When using a variable, **always** ask C to **reserve memory** for some **duration**.
2. **Always allocate** and **deallocate** long duration storage.
3. **Always initialize** variables.
4. **Watch out** for **off-by-one** errors.
5. **Always null-terminate** “C strings.”

The C Programming Language

State Diagrams

```
#include <stdio.h>

int main() {
  int i = 10, j = 0, *k;
  k = &i;
  *k = 20;
  k = &j;
  *k = i;
  printf("i = %d,
          j = %d,
         *k = %d\n",
         i, j, *k);
→ return 0;
}
```



(state **just before** the line indicated by the **arrow** is executed)

The C Programming Language

State Diagram Rules

The Rules

1. Initialize diagram with empty stack and heap.
2. When a function is called, put a box on the stack, and label it with the function's name.
3. Put global variables outside the box.
4. Put local (automatic) variables inside the box, including function parameters.
5. Manage allocated variables on the heap.
   (a) `malloc` adds objects.
   (b) `free` removes objects.
6. As the function runs, update values.
7. Returning from a function pops the stack frame and, if the function returns a value, assigns it to the storage awaiting the return value.

Makefiles

```
program: c.c b.o a.o
tab → gcc -o program c.c b.o a.o
```

target: dep_1 … dep_n
tab → command

command should produce target.

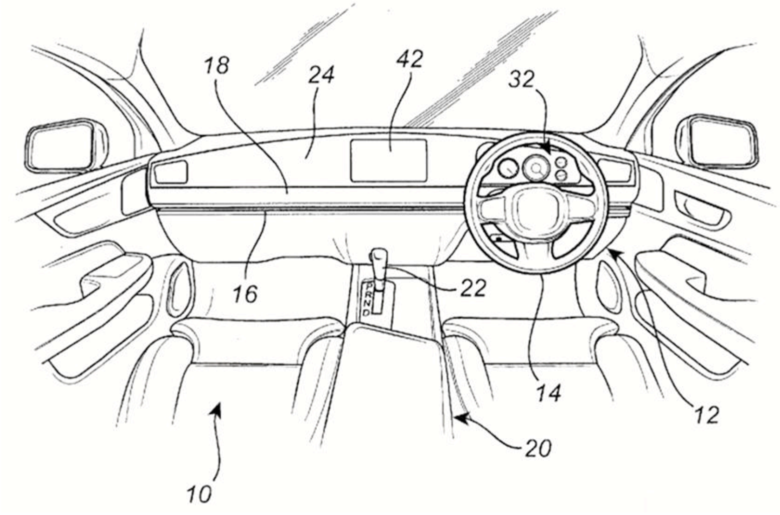
Makefiles

```
CFLAGS=-Wall -g

.PHONY: all
all: dictattack hashchain

database.o: database.h database.c
   gcc $(CFLAGS) -c database.c

crackutil.o: crackutil.h crackutil.c database.h
   gcc $(CFLAGS) -c crackutil.c

dictattack: crackutil.o database.o dictattack.c
   gcc $(CFLAGS) -o dictattack dictattack.c crackutil.o database.o -lmd
```

Libraries: static vs shared

- Static library: compile with `-c`
- Shared library: link with `-l<whatever>`

.h files are **interfaces**



Building with libraries

```
CFLAGS=-Wall -g

.PHONY: all
all: dictattack

database.o: database.h database.c
   gcc $(CFLAGS) -c database.c

crackutil.o: crackutil.h crackutil.c database.o
   gcc $(CFLAGS) -c crackutil.c

dictattack: crackutil.o database.o dictattack.c
   gcc $(CFLAGS) -o dictattack dictattack.c crackutil.o database.o -lmd

.PHONY: clean
clean:
   rm -f *.o
   rm -f dictattack
   rm -rf *.dSYM
```

static library

shared library

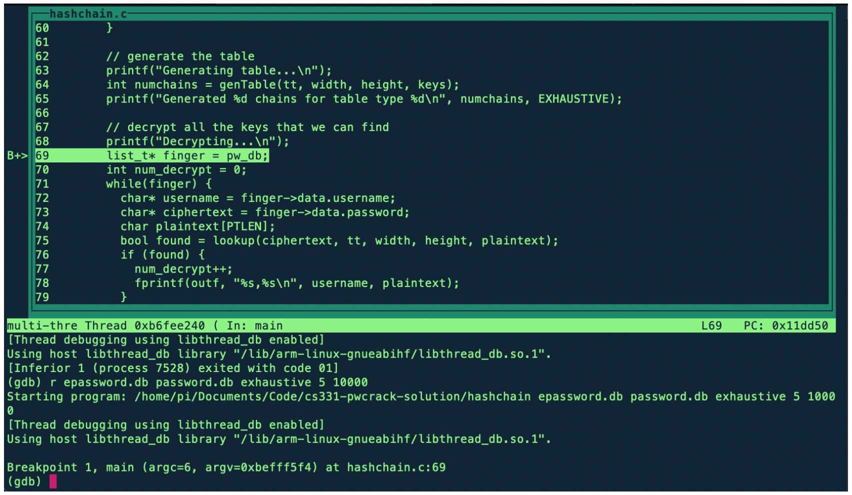
Finding memory errors with ASan

```
-fsanitize=address -static-libasan
```

Kinds of memory errors:

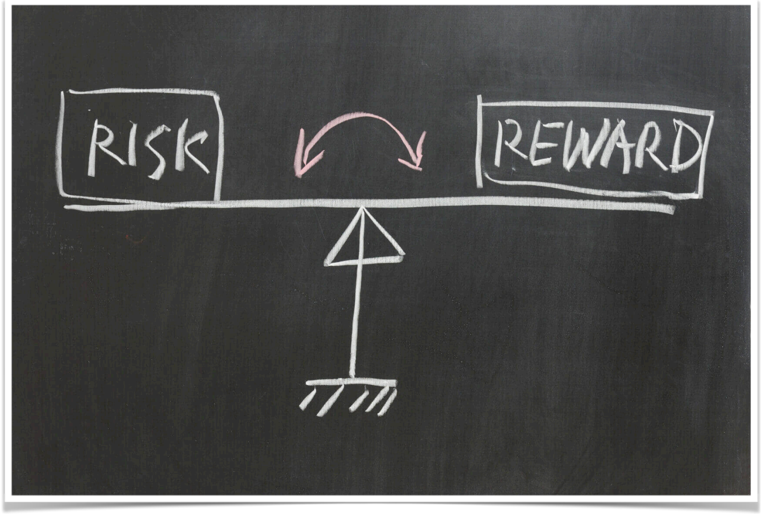
- Segmentation fault
- Memory leak
- Out-of-bounds read
- Buffer overflow (OOB write)
- Use-after-free
- Uninitialized read

Debugging with gdbtui

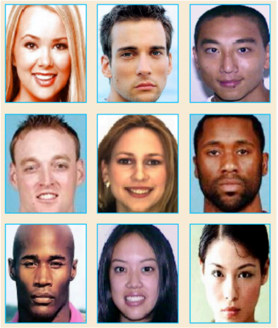


```
  ┌─hashchain.c─────────────────────────────────────────────────────────────────
  │60        }
  │61
  │62        // generate the table
  │63        printf("Generating table...\n");
  │64        int numchains = genTable(tt, width, height, keys);
  │65        printf("Generated %d chains for table type %d\n", numchains, EXHAUSTIVE);
  │66
  │67        // decrypt all the keys that we can find
  │68        printf("Decrypting...\n");
B+>│69        list_t* finger = pw_db;
  │70        int num_decrypt = 0;
  │71        while(finger) {
  │72          char* username = finger->data.username;
  │73          char* ciphertext = finger->data.password;
  │74          char plaintext[PTLEN];
  │75          bool found = lookup(ciphertext, tt, width, height, plaintext);
  │76          if (found) {
  │77            num_decrypt++;
  │78            fprintf(outf, "%s,%s\n", username, plaintext);
  │79          }
multi-thre Thread 0xb6fee240 ( In: main                                      L69   PC: 0x11dd50
[Thread debugging using libthread_db enabled]
Using host libthread_db library "/lib/arm-linux-gnueabihf/libthread_db.so.1".
[Inferior 1 (process 7528) exited with code 01]
(gdb) r epassword.db password.db exhaustive 5 10000
Starting program: /home/pi/Documents/Code/cs331-pwcrack-solution/hashchain epassword.db password.db exhaustive 5 1000
0
[Thread debugging using libthread_db enabled]
Using host libthread_db library "/lib/arm-linux-gnueabihf/libthread_db.so.1".

Breakpoint 1, main (argc=6, argv=0xbefff5f4) at hashchain.c:69
(gdb) 
```

Security as a tradeoff



Security as a tradeoff



e.g., memorability vs guessability

Security as a tradeoff

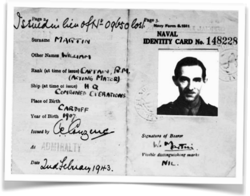
How to quantify risk-reward tradeoff

- Enumerate potential vulnerabilities
- Assign exploit probabilites
- Estimate cost of exploit
- Compute expected cost
- Rational expenses for mitigation do not exceed the expected cost of the exploit

Security properties

Confidentiality

Integrity


NAVAL IDENTITY CARD No. 148228
MARTIN

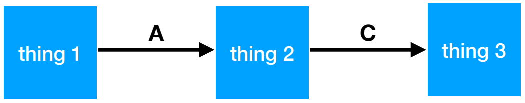
Authenticity

Availability

Security properties

Non-repudiation

CIAA graphical model



Crypto!

Encryption is the **process of encoding a message** so that it can be read only by the sender and the **intended recipient**.

- A **plaintext** p is the original, unobfuscated data. This is information you want to protect.
- A **ciphertext** c is encoded, or encrypted, data.
- A **cipher** f is an algorithm that converts **plaintext** to **cipertext**. We sometimes call this function an **encryption function**.
  - ✳More formally, a cipher is a function from plaintext to ciphertext, $f(p)=c$. The properties of this function determine what kind of encryption scheme is being used.
- A **sender** is the person (or entity) who enciphers or encrypts a message, i.e., the party that converts the plaintext into cipertext. $f(p)=c$
- A **receiver** is the person (or entity) who deciphers or decrypts a message, i.e., the party that converts the ciphertext back into plaintext. $f^{-1}(c)=p$

Cryptographic hash functions

Suppose we have:

$f(p)=c$, a **cipher** that maps **plaintexts** to **ciphertexts**; in this case, a **hash function**.

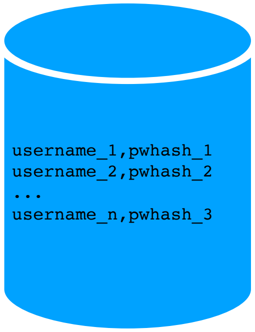
Because f is a hash function, there is **no inverse function** such that $f^{-1}(f(p))=p$.

A cryptographic hash function is **bitwise independent**, meaning that seeing one or more bits of output **does not help an attacker** predict the values of the remaining outputs.

Brute Force Password Attacks

Online, using a pseudoterminal.

Offline, using a password cracking algorithm.

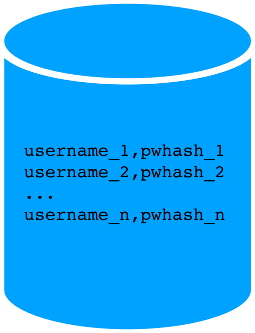


```
username_1,pwhash_1
username_2,pwhash_2
...
username_n,pwhash_3
```

Offline password database attacks

- Random guessing attack
- Enumeration attack
- Dictionary attack
- Precomputed hash chain attack
- Rainbow table attack

Random guessing: complexity (one pw)



```
username_1,pwhash_1
username_2,pwhash_2
...
username_n,pwhash_n
```

m = # of possible passwords

p = probability that random guess is correct

= 1/**m**

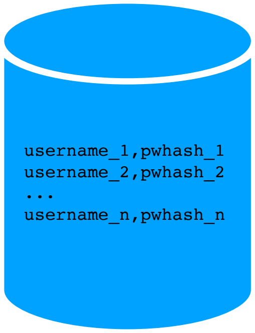
X = # guesses until success

$E[X] = (1-p)/p$ (geometric dist)

$= m - 1$

O(**m**) average **per pw** O(**mn**) average for **all pw**

Enumeration: complexity



```
username_1,pwhash_1
username_2,pwhash_2
...
username_n,pwhash_n
```

m = # of possible passwords

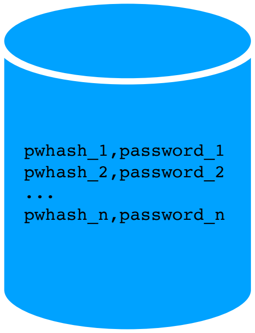
Average guesses to find **one pw**:

$O(m/2)$

Average guesses to find **all pw**:

$O(n \times m/2)$

Dictionary attack: complexity



```
pwhash_1,password_1
pwhash_2,password_2
...
pwhash_n,password_n
```

m = # of possible passwords

Time to compute dictionary:

$O(m)$

Time to lookup **one pw**:

$O(\log m)$

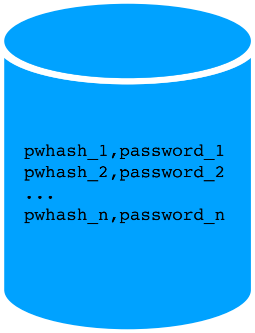
Time to lookup **all pws**:

$O(n \log m)$

Space needed:

$O(m)$

PCHC/rainbow attack: complexity



```
pwhash_1,password_1
pwhash_2,password_2
...
pwhash_n,password_n
```

m = # of possible passwords

Time to compute data structure:

$O(m)$

Time to lookup **one pw**:

$O(k)$

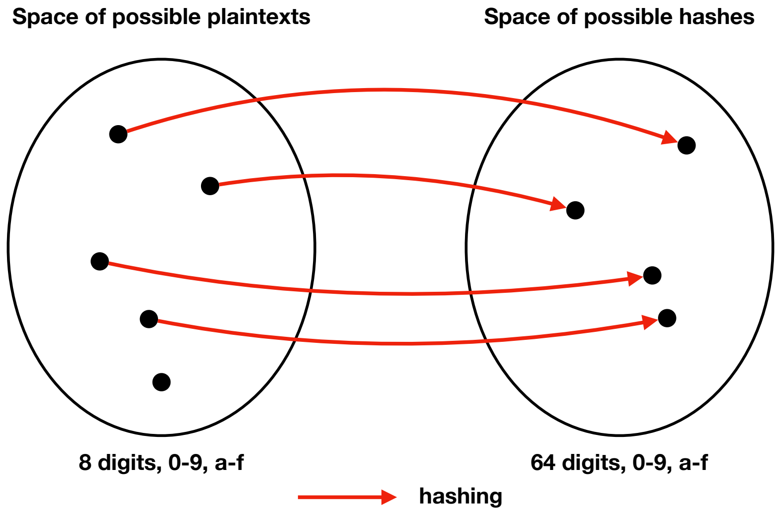
Time to lookup **all pws**:

$O(mk)$

Space needed:

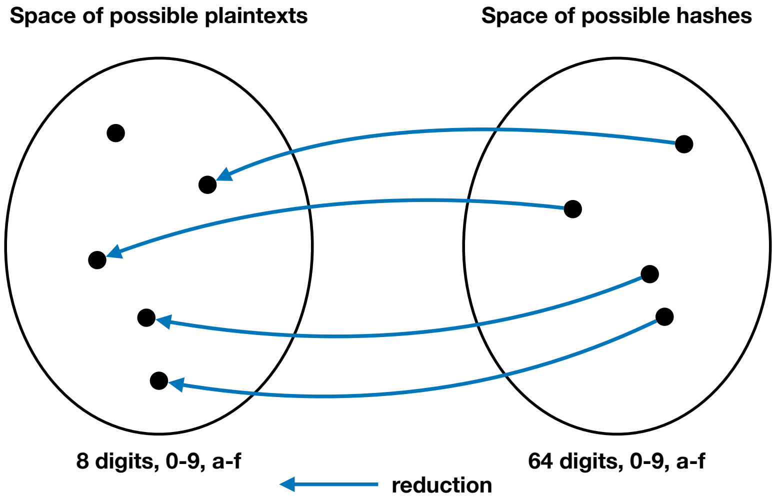
$O(m/k)$

Hash function



plaintext: “9a55302d” **ciphertext:** “4651f1799e5e36c878f3d980c59e94ae”

Reducer function

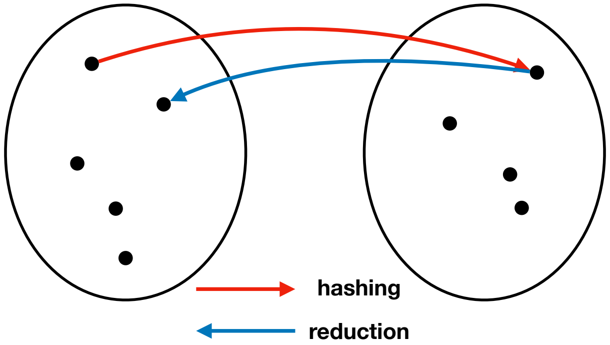


ciphertext: “`4651f1799e5e36c878f3d980c59e94ae`” **plaintext:** “`4651f179`”

Reducer function properties

A reducer $r(c)=p$ only needs to satisfy a couple properties.

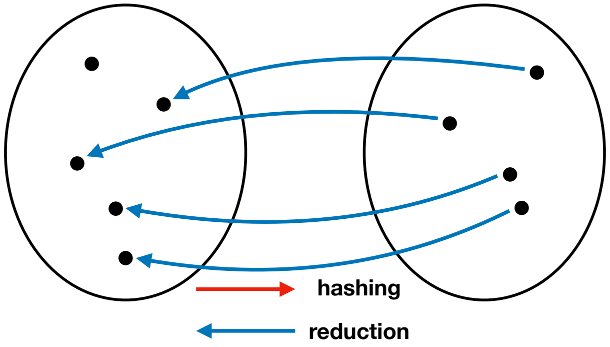
1. A reducer’s output, p, should map to the same domain as the *input* of the hash function, $f(p)=c$ (i.e,. plaintexts)



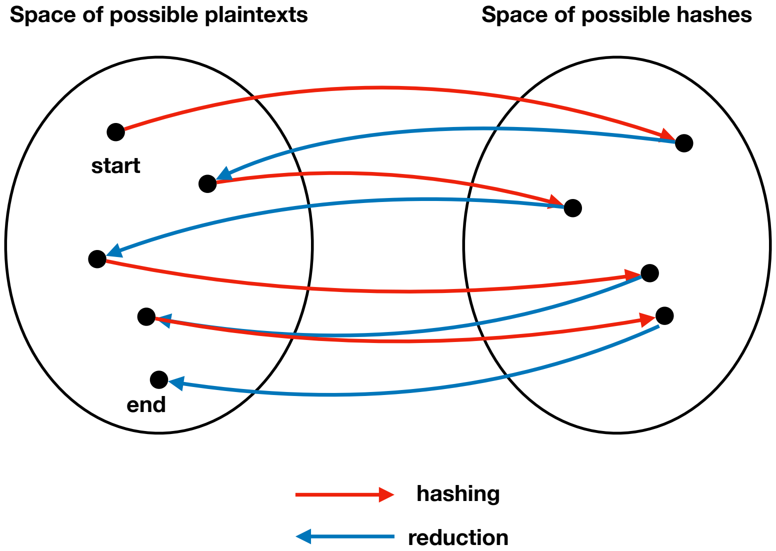
Reducer function properties

A reducer $r(c)=p$ only needs to satisfy a couple properties.

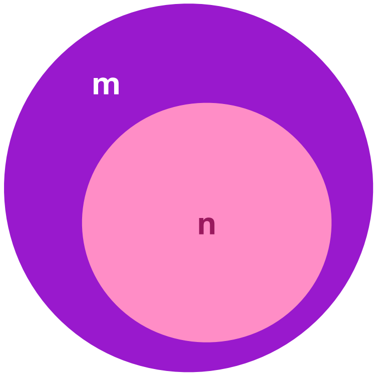
2. All plaintexts should be selected, given the space of ciphertexts, with equal probability.



Hash chain



Hashes are guaranteed to collide

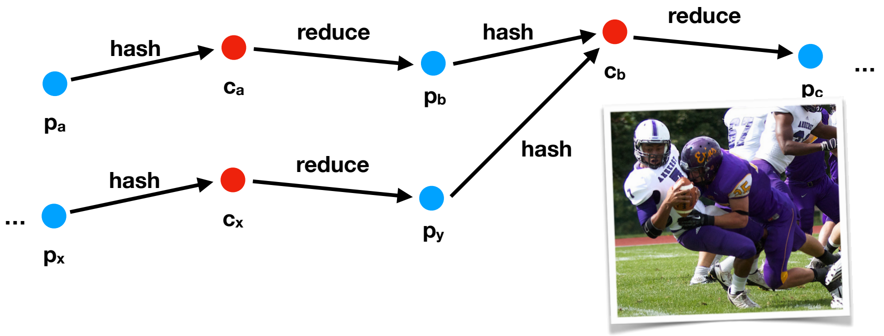


m: # of passwords n: # of hashes

If m > n, we know that **at least (m-n)/m** must collide.

“pigeonhole principle”

Collisions in a hash chain

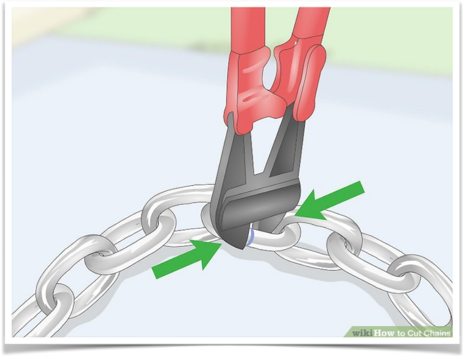


After the **collision**, the chain “**loops**.”

Collisions prevent us from enumerating the **entire space**!

Hash chain of length k

We are going to chop up our long chain into **smaller chains** of length **k**.



Store only **start** and **end**

start, end
p_m , p_{m-3}
…
p_5 , p_3
p_3 , p_1

Store it **backward**

end, start
p_{m-3} , p_m
…
p_3 , p_5
p_1 , p_3

Hash function lookup table:

plaintext	Hash of plaintext
♥♥♥♥	4A7D1ED414474E4033AC29CCB8653D9B
♥♥♥★	25BBDCD06C32D477F7FA1C3E4A91B032
♥♥★♥	FC1198178C3594BFDDA3CA2996EB65CB
♥♥★★	AE2BAC2E4B4DA805D01B2952D7E35BA4
♥★♥♥	DB2F40F24260BC41DB48D82D5E7ABF1D
♥★♥★	814F06AB7F40B2CFF77F2C7BDFFD3415
♥★★♥	2A66ACBC1C39026B5D70457BB71B142B
♥★★★	7D7C45B9A935CF9D845FC75679A41559
★♥♥♥	A9B7BA70783B617E9998DC4DD82EB3C5
★♥♥★	B8C37E33DEFDE51CF91E1E03E51657DA
★♥★♥	1E48C4420B7073BC11916C6C1DE226BB
★♥★★	7F975A56C761DB6506ECA0B37CE6EC87
★★♥♥	1E6E0A04D20F50967C64DAC2D639A577
★★♥★	C6BFF625BDB0393992C9D4DB0C6BBE45
★★★♥	2CBCA44843A864533EC05B321AE1F9D1
★★★★	B59C67BF196A4758191E42F76670CEBA

```
func reducer(c,i):

Convert the ith hexadecimal
digit of c into a plaintext
using the following table:
```

hex	plaintext
0	♥♥♥♥
1	♥♥♥★
2	♥♥★♥
3	♥♥★★
4	♥★♥♥
5	♥★♥★
6	♥★★♥
7	♥★★★
8	★♥♥♥
9	★♥♥★
A	★♥★♥
B	★♥★★
C	★★♥♥
D	★★♥★
E	★★★♥
F	★★★★

Find the first three rainbow chains of length 3.

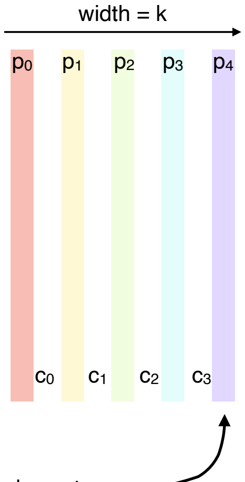
First three rainbow chains

♥♥♥♥ $\xrightarrow{h}$ 4A7D $\xrightarrow{r_0}$ ♥★♥♥ $\xrightarrow{h}$ DB2F $\xrightarrow{r_1}$ ★♥★★ $\xrightarrow{h}$ 7F97 $\xrightarrow{r_2}$ ★♥♥★

♥♥♥★ $\xrightarrow{h}$ 25BB $\xrightarrow{r_0}$ ♥♥★♥ $\xrightarrow{h}$ FC11 $\xrightarrow{r_1}$ ★★♥♥ $\xrightarrow{h}$ 1E6E $\xrightarrow{r_2}$ ♥★★♥

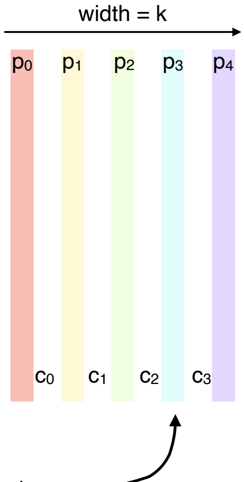
♥♥★♥ $\xrightarrow{h}$ FC11 $\xrightarrow{r_0}$ ★★★★ $\xrightarrow{h}$ B59C $\xrightarrow{r_1}$ ♥★♥★ $\xrightarrow{h}$ 814F $\xrightarrow{r_2}$ ♥★♥♥

end	start
★♥♥★	♥♥♥♥
♥★★♥	♥♥♥★
♥★♥♥	♥♥★♥



I hypothesize that c reduces to p_4

What reducer should I use? `reduce(c,3)`



I hypothesize that c reduces to p_{k-2}

What reducer should I use? `reduce(c,2)`

Then: `reduce(c,3)`

Rainbow table (for first 3 chains)

end	start
★♥♥★	♥♥♥♥
♥★★♥	♥♥♥★
♥★♥♥	♥♥★♥

Decrypt `FC11`.

Hypothesis: `FC11` is the third link in the chain.

FC11 $\xrightarrow{r_2}$ ♥♥♥★ Is ♥♥♥★ an **end**? **No.**

Hypothesis: `FC11` is the second link in the chain.

FC11 $\xrightarrow{r_1}$ ★★♥♥ $\xrightarrow{h}$ 1E6E $\xrightarrow{r_2}$ ♥★★♥ Is ♥★★♥ an **end**? **Yes.**

Decrypt from **start** ♥♥♥★:

plaintext

♥♥♥★ $\xrightarrow{h}$ 25BB $\xrightarrow{r_0}$ ♥♥★♥ $\xrightarrow{h}$ FC11

Countermeasures Against Cracking Attacks

- Password salts.
- Uniformly-distributed passwords.
- Two-factor authentication.
- Last-known IP address.
- Make hashing expensive.

Key Stretching

Key stretching is a technique used to make password decryption attacks **computationally expensive**. Unlike an ordinary user, an attacker must invoke a hash function many times. Key stretching **amplifies the cost of a hash function** using a **stretch factor s**.

$f^s(p) = c^s$ is an iterated hash function, where

$$f^1(p) = f(p) = c^1$$

$$f^2(p) = f(f(p)) = c^2$$

$$f^3(p) = f(f(f(p))) = c^3$$

$$\ldots$$

$$f^n(p) = c^n$$

Practice exam solutions

Q&A

Recap & Next Class

Today we learned:

Rainbow table generation

Rainbow table lookup

Sample buffer overflow exploit

Next class:

How to craft an exploit