02 Introduction to Cryptography and Stream Cipher

2021 Spring
Information Security

Teacher: Po-Wen Chi
neokent@gapps.ntnu.edu.tw

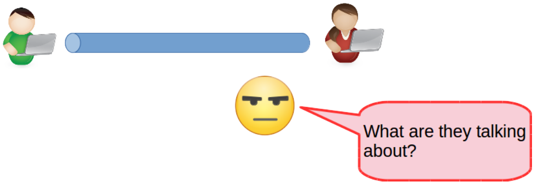
March 9, 2021

Department of Computer Science and Information Engineering,
National Taiwan Normal University

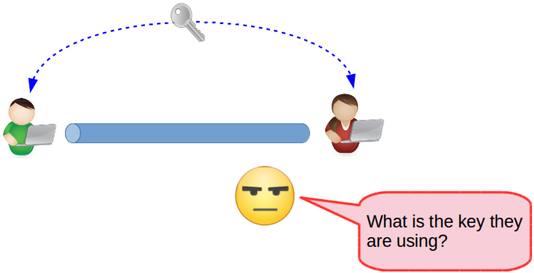
What is Cryptography

The Crypto Core

- Secure Communication.



- Key Establishment.



Cryptography Can Do More

- Digital Signature.
- Digital Currency.
- e-Voting.
- Digital Watermark.
- Secure computation.

In this class, we will not cover all these topics.

Cryptography is a Rigorous Science

1. Define the scenario and the threat model.
2. Propose a construction.
3. Prove the construction:
   - **A hard problem assumption**.
   - **Show that if the crypto construction is broken, the hard problem will be solved**.

Brute Force Attack

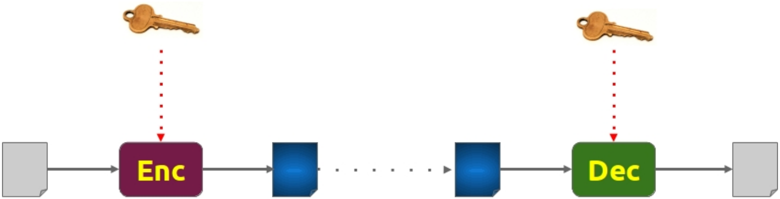
Try every possible key.

It always works but we will not claim that we break a crypto system through the brute force attack.

Suppose the key length is 256bit, and you can use 50 supercomputers that could check a billion billion (10^{18}) keys per second, you will use 3×10^{51} years to try all possibilities.

History

Symmetric Key Encryption



$$C = E_k(M)$$

$$M = D_k(C)$$

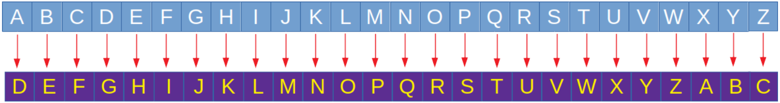
Encryption key and decryption key is the same key.

Definition

- **Encryption** should be **computationally efficient**.
- **Decryption**:
  - It should be **computationally efficient** with the correct key.
  - Without key, the probability of finding the correct message is **negligible**.

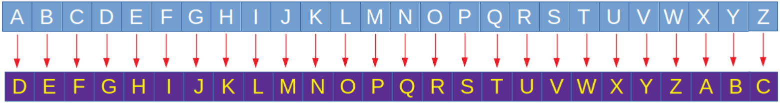
How will you design?

Caesar Cipher



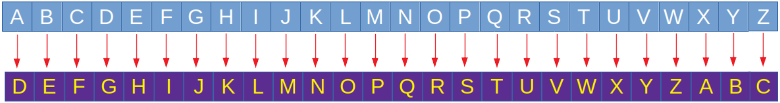
- In this case, key is shift 3.
- HELLO will be encrypted to

Caesar Cipher



- In this case, key is shift 3.
- HELLO will be encrypted to
  - KHOOR.

Caesar Cipher



- In this case, key is shift 3.
- HELLO will be encrypted to
  - KHOOR.
- ZHOFRPH will be decrypted to

Caesar Cipher

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

- In this case, key is shift 3.
- HELLO will be encrypted to
  - KHOOR.
- ZHOFRPH will be decrypted to
  - WELCOME.

Quiz

- How much time do you need to crack this crypto system?

Quiz

- How much time do you need to crack this crypto system?
- Answer: You only need to try 25 times to crack this system.

Enhanced Version: Monoalphabetic Substitution Cipher

- Example:
  - Plaintext: ABCDEFGHIJKLMNOPQRSTUVWXYZ
  - Ciphertext: QWERTYUIOPASDFGHJKLZXCVBNM

If We Want to Break Substitution Cipher ...

How many keys do we need to try?

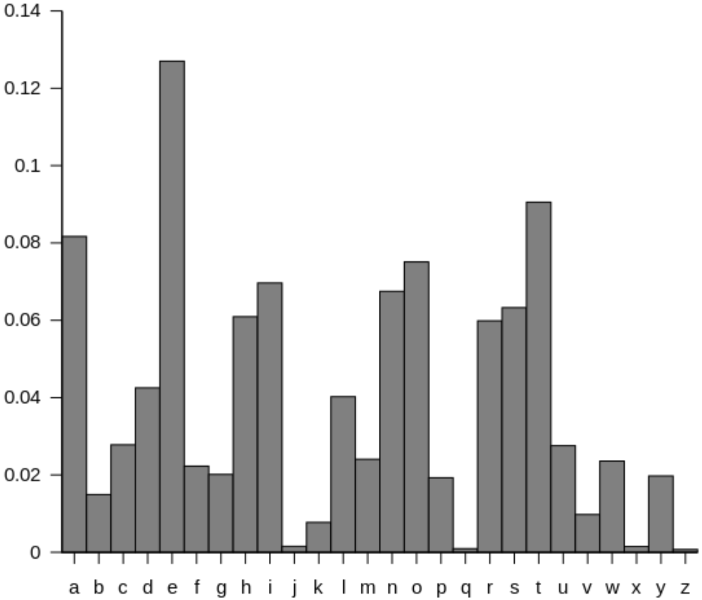
$$26! \approx 2^{88} \approx 2.56 \times 10^{26}$$

If it takes 10 seconds to try one possible password, how much time do we need to try all possible passwords?

$$2.56 \times 10^{27}\text{seconds} = 8.12 \times 10^{19}\text{years}$$

It looks safe, right?

Frequency Analysis

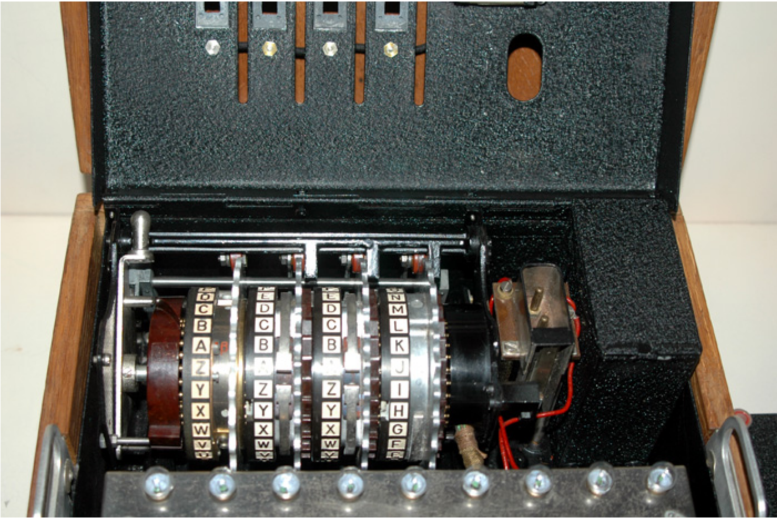


Vigenère Cipher

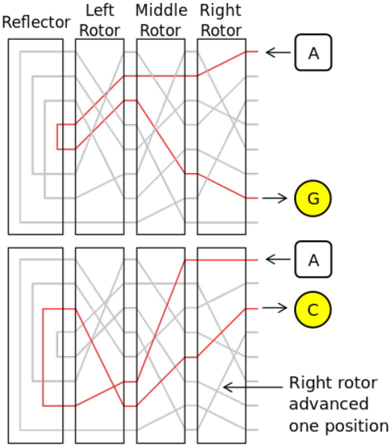
K=	C	R	Y	P	T	O	C	R	Y	P	T	O
M=	W	H	A	T	A	N	I	C	E	D	A	Y
C=	Z	Z	Z	J	U	C	L	U	D	T	U	N

How will you break this??

Rotor Machine: Enigma



Rotor Machine: Enigma



$$|K| = 26^4 \approx 2^{18}$$

One Time Pad

Exclusive Or $\oplus$

- $\oplus$ is a logical operation that outputs true only when inputs differ.

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0

- Some properties:
  - $m \oplus 0 = m$
  - $m \oplus m = 0$
  - $p \oplus q = q \oplus p$
  - $(a \oplus b) \oplus c = a \oplus (b \oplus c)$

One Time Pad (Vernam Cipher)

$$C = E_K(M) = M \oplus K$$

$$M = D_K(C) = C \oplus K$$

Example:

$$\begin{array}{rccccccc} M = & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ K = & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ \hline C = & 1 & 1 & 0 & 1 & 1 & 0 & 1 \end{array}$$

One Time Pad

- The key length should be equal to the message size.
- The key can only be used once.

Why?

One Time Pad

- The key length should be equal to the message size.
- The key can only be used once.

Why?

$$c_1 = E_k(m_1) = m_1 \oplus k$$

$$c_2 = E_k(m_2) = m_2 \oplus k$$

$$c_1 \oplus c_2 = m_1 \oplus k \oplus m_2 \oplus k = m_1 \oplus m_2$$

One Time Pad

- Advantage:
  Very fast. Only exclusive or operation.
- Drawback:
  The key size is equal to the message size. Are you kidding me?

Before we see how to solve this drawback, we want to know if One Time Pad is secure?

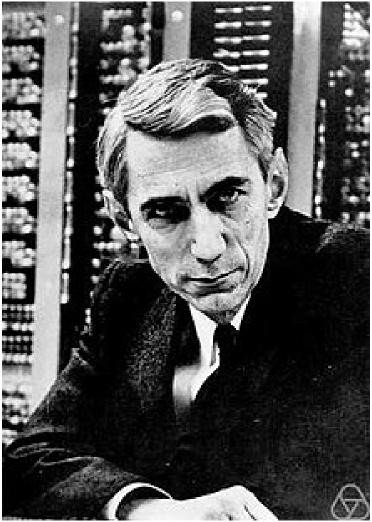
One Time Pad

- Advantage:
  Very fast. Only exclusive or operation.
- Drawback:
  The key size is equal to the message size. Are you kidding me?

Before we see how to solve this drawback, we want to know if One Time Pad is secure?

What is a secure cipher?

Shannon’ s idea:
A ciphertext should **reveal no information** about its plaintext.



Definition

A cipher (E, D) over $(\mathcal{M}, \mathcal{K}, \mathcal{C})$ is **perfectly secure** if

$$\forall m_0, m_1 \in \mathcal{M}, |m_0| = |m_1|, c \in \mathcal{C},$$

$$P[E_k(m_0) = c] = P[E_k(m_1) = c] \text{ when } k \text{ is uniform in } \mathcal{K}.$$

Information Theoretic Security

- Given a ciphertext c, no one can tell if the message is m_0 or m_1.
- An adversary learns nothing about the message from the ciphertext.
- There is no ciphertext only attack.

Some Notes

From my experience, most people think cryptography is hard because of two reasons:

1. Lots of math backgrounds are required.
2. Mathematical expression is hard to understand.

Do not be afraid and try to use your words to explain it.

OTP Has Perfect Secrecy

Proof:

$$
\begin{aligned}
& k \oplus m = c \\
\Rightarrow \quad & k = c \oplus m \\
\Rightarrow \quad & \#\{k \in \mathcal{K}, E_k(m) = c\} = 1
\end{aligned}
$$

So OTP has perfect secrecy.

OTP Has Perfect Secrecy

Proof:

$$
\begin{aligned}
& k \oplus m = c \\
\Rightarrow \quad & k = c \oplus m \\
\Rightarrow \quad & \#\{k \in \mathcal{K}, E_k(m) = c\} = 1
\end{aligned}
$$

So OTP has perfect secrecy.

What are you talking about??

Unfortunately

- For perfect secrecy, $|\mathcal{K}| \geq |\mathcal{M}|$. This is **impractical**.
- You cannot use the same key twice.

PS: One time pad is different from one time password.

Probability and Ciphers

Notations

- Let $\mathcal{M}$ denote the set of possible messages.
- Let $\mathcal{K}$ denote the set of possible keys.
- Let $\mathcal{C}$ denote the set of ciphertexts.

Perfect Secrecy

A cryptosystem has perfect secrecy if

$$P[M = m | C = c] = P[M = m]$$

for all plaintexts $m \in \mathcal{M}$ and all ciphertexts $c \in \mathcal{C}$.

It means that seeing c reveals nothing about m.

Perfect Secrecy

A cryptosystem has perfect secrecy if

$$P[C = c | M = m] = P[C = c]$$

for all plaintexts $m \in \mathcal{M}$ and all ciphertexts $c \in \mathcal{C}$.

Proof:

$$P[M = m | C = c] = \frac{P[M = m] P[C = c | M = m]}{P[C = c]} = P[M = m].$$

Perfect Secrecy

Assume the cryptosystem is perfectly secure, then

$$\#\mathcal{K} \geq \#\mathcal{C} \geq \#\mathcal{M}.$$

Proof:

- For any encryption scheme, $\#\mathcal{C} \geq \#\mathcal{M}$.
- Since every ciphertext can occur, we have

$$Pr(C = c | M = m) = Pr(C = c) > 0.$$

  For all c, there must be a key k such that $E_k(m) = c$. So $\#\mathcal{K} \geq \#\mathcal{C}$.

Shannon's Definition

Let $(\mathcal{M}, \mathcal{C}, \mathcal{K}, E, D)$ denote a cryptosystem where $\#\mathcal{K} = \#\mathcal{C} = \#\mathcal{M}$. The cryptosystem provides perfect secrecy if and only if

- Every key is used with equal probability.
- For each $m \in \mathcal{M}$ and $c \in \mathcal{C}$, there is an unique key k that $E_k(m) = c$.

Entropy

Definition

Let X be a random variable witch takes on a finite set of values x_i, where $1 \leq i \leq n$ and has probability distribution $p_i = p(X = x_i)$. The entropy of X is defined to be

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i.$$

You can treat entropy as a kind of information.

Joint Entropy

$$H(X, Y) = -\sum_{i=1}^{n} \sum_{j=i}^{m} p_{i,j} \log_2 p_{i,j},$$

where $p_{i,j} = P(X = x_i, Y = y_j)$.

$$H(X, Y) = H(Y) + H(X|Y) \leq H(X) + H(Y).$$

Entropy in Cryptography

$$\begin{aligned} H(K,M,C) &= H(M,K) + H(C|M,K) \\ &= H(M,K) \\ &= H(K) + H(M) \end{aligned}$$

$$\begin{aligned} H(K,M,C) &= H(C,K) + H(M|C,K) \\ &= H(C,K) \end{aligned}$$

$$H(K,C) = H(K) + H(M)$$

$$H(K|C) = H(K,C) - H(C) = H(K) + H(M) - H(C)$$

Example

Example

$$\mathcal{M} = \{a, b, c, d\}, \mathcal{K} = \{k_1, k_2, k_3\}, \mathcal{C} = \{1, 2, 3, 4\}.$$

- $P(M = a) = 0.25, P(M = b) = 0.3, P(M = c) = 0.3, P(M = c) = 0.15.$
- $P(K = k_1) = 0.25, P(K = k_2) = 0.5, P(K = k_3) = 0.25.$
- $P(C = 1) = 0.2625, P(C = 2) = 0.2625, P(C = 3) = 0.2625, P(C = 4) = 0.2125.$

$$H(M) \approx 1.9525$$
$$H(K) \approx 1.5$$
$$H(C) \approx 1.9944$$

Continue

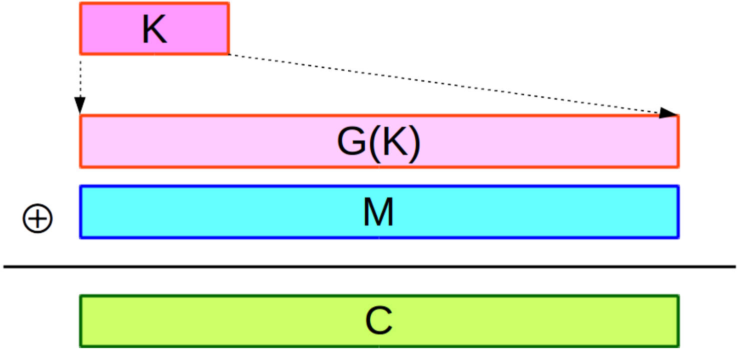
$$H(K|C) \approx 1.4583$$

So 0.042 bits of information about the key is revealed.

Stream Cipher: Making OTP Practical

Stream Cipher: Making OTP Practical

Is it possible to use a short key instead?



$G(\cdot)$ is a **pseudo-random generation** function (PRG).

Quiz

Does Stream Cipher has perfect secrecy?

Quiz

Does Stream Cipher has perfect secrecy?

Definitely NO!

Since the key is shorter than the message.

PRG Should be Unpredictable

Suppose a PRG is **predictable** which means:

$$\exists i : G(k)|_{1,\ldots,i} \rightarrow G(k)|_{i+1,\ldots,n}$$

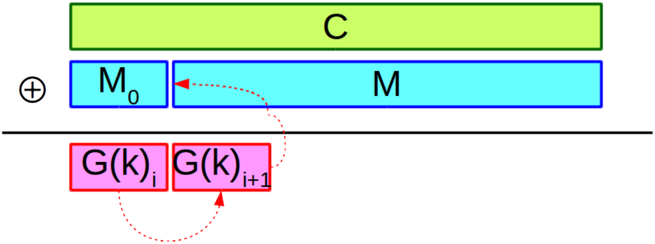
What will happen if PRG is predictable?

PRG Should be Unpredictable

Suppose a PRG is **predictable** which means:

$$\exists i : G(k)|_{1,\ldots,i} \rightarrow G(k)|_{i+1,\ldots,n}$$

What will happen if PRG is predictable?



Unpredictable PRG

predictable

We say that a PRG $G : K \to \{0,1\}^n$ is **predictable** if

$$\exists \text{ eff alg } A \text{ and } \exists 0 \le i \le n-1,$$

$$P[A(G(k)|_{1\ldots i}) = G(k)|_{i+1}] = \frac{1}{2} + \epsilon,$$

where ϵ is non-ignorable.

unpredictable

We say that a PRG is **unpredictable** if it is not predictable.

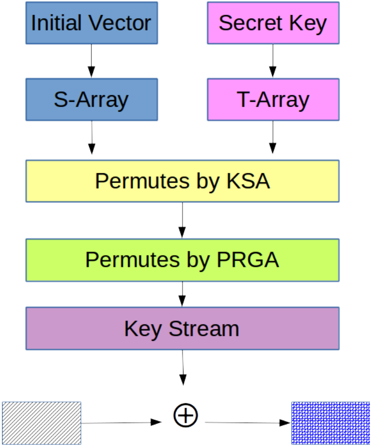
It is Easy, We can Use random() ⋯?

Never ever use random() in crypto.

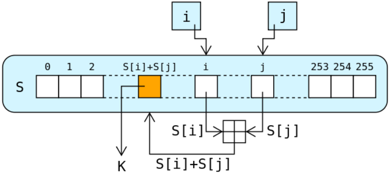
```
if (buf->rand_type == TYPE_0)
{
  int32_t val = state[0];
  val = ((state[0] * 1103515245) + 12345)
        & 0x7fffffff;
  state[0] = val;
  *result = val;
}
```

Stream Cipher in the Real World

Rivest Cipher 4 (RC4)

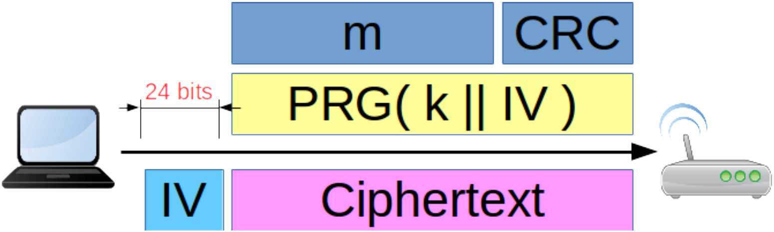



```
for i from 0 to 255
    S[i] := i
endfor
j := 0
for i from 0 to 255
    j := (j + S[i] + key[i mod keylength]) mod 256
    swap values of S[i] and S[j]
endfor
```



Wired Equivalent Privacy (WEP)

Based on RC4.



How will you attack??

Another Attack: Key Recovery

If the Initial Vector (24 bits) are as follows, what will happen??

3	255	

Key Recovery Attack

Before:

T

3	255	X	K1

S

0	1	2	3

After:

T

3	255	X	K1

S

3	1	2	0

```
i = 0, j = 0
j = j + S[i] + T[i] = 3
swap S[0] and S[3]
```

Key Recovery Attack

Before:

T

3	255	X	K1

S

3	1	2	0

After:

T

3	255	X	K1

S

3	0	2	1

```
i = 1, j = 3
j = j + S[i] + T[i] = 3
swap S[1] and S[3]
```

Key Recovery Attack

Before:

T

3	255	X	K1

S

3	0	2	1

After:

T

3	255	X	K1

S

3	0	5+X	1

```
i = 2, j = 3
j = j + S[i] + T[i] = 5 + X
swap S[2] and S[5+X]
```

Key Recovery Attack

Before:

T

3	255	X	K1

S

3	0	5+X	1

After:

T

3	255	X	K1

S

3	0	5+X	5+X+1+K1

```
i = 3,  j = 5+X
j = j + S[i] + T[i]
  = 5 + X + 1 + K1
swap S[3] and S[5+X+1+K1]
```

So What?

- After KSA, the probability that $S[5 + X + 1 + K1]$ remains at the same place is 5%.
- The first byte of WiFi frame is 0xAA. That is, you can undoubtedly derive the first byte of the keystream.
- The first byte of the keystream is S[S[0]+S[1]]=S[3]=5+X+1+K1.
- So you can get K1.

VAUDENAY, Serge; VUAGNOUX, Martin. Passive–only key recovery attacks on RC4. In: International Workshop on Selected Areas in Cryptography. Springer, Berlin, Heidelberg, 2007. p. 344-359.

Summary

- WEP is not secure.
  - IV is too short.
  - RC4 is not secure.

Semantic Security

Perfectly Security is Almost Impossible

- We know that OTP is perfectly secure but it is impractical.
- The practical solution, stream cipher, is not perfectly secure.
- We need **another security model** ...

Modify the Definition

Perfectly Secure

A cipher (E, D) over $(\mathcal{M}, \mathcal{K}, \mathcal{C})$ is **perfectly secure** if

$$\forall m_0, m_1 \in \mathcal{M}, |m_0| = |m_1|, c \in \mathcal{C},$$

$$P[E_k(m_0) = c] = P[E_k(m_1) = c], k \leftarrow \mathcal{K}.$$

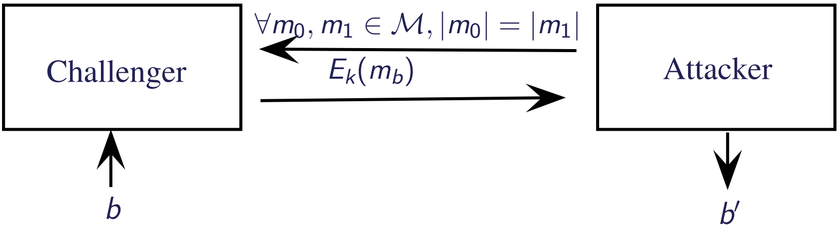
Semantically Secure

A cipher (E, D) over $(\mathcal{M}, \mathcal{K}, \mathcal{C})$ is **semantically secure** if

$$\forall m_0, m_1 \in \mathcal{M}, |m_0| = |m_1|, c \in \mathcal{C},$$

$$P[E_k(m_0) = c] \approx P[E_k(m_1) = c], k \leftarrow \mathcal{K}.$$

Semantic Security



- For $b \in \{0, 1\}$, W_b is the event that $b' = 1$ when the experiment choice is b.
- Advantage: $|P[W_0] - P[W_1]|$.
  - The probability that the attacker wins - 0.5.
- **Semantic Security**: if for all attackers, the advantage is negligible.

Quiz

Suppose efficient A can always deduce LSB of the message from the ciphertext.

Is the encryption system semantically secure??

Quiz

Suppose efficient A can always deduce LSB of the message from the ciphertext.

Is the encryption system semantically secure?? No

Quiz

Is OTP semantically secure?

Notes

- The attacker may collect lots of plaintext and ciphertext pairs before the challenge game.
- Undoubtedly, you cannot challenge known plaintexts.

Stream Ciphers are Semantically Secure

Stream Ciphers are Semantically Secure

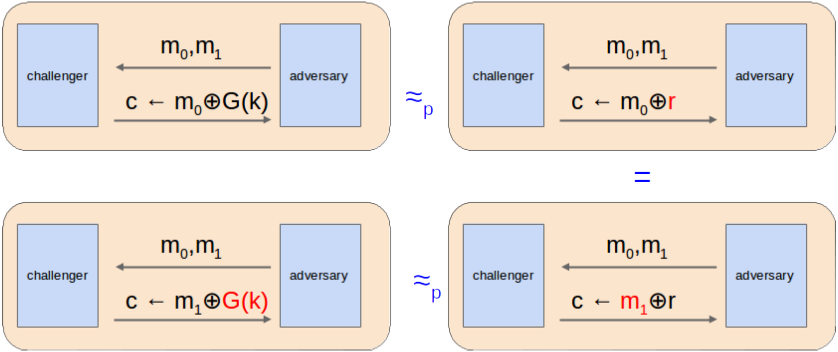
Theorem

If $G : k \rightarrow \{0,1\}^n$ is an unpredictable PRG, then the stream cipher is semantically secure. That is, $\forall$ semantically secure adversary A, $\exists$ a PRG adversary B such that

$$\mathsf{Adv}_{SS}[A, E] \leq 2 \times \mathsf{Adv}_{PRG}[B, G].$$

How to prove this?

Proof: Intuition



Appendix: Password

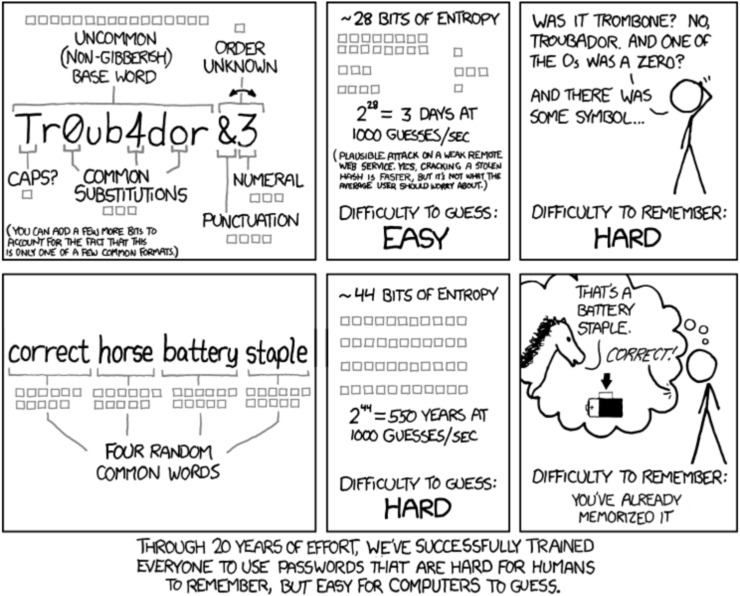
What is a Good Password?

- the use of both upper-case and lower-case letters (case sensitivity)
- inclusion of one or more numerical digits
- inclusion of special characters, such as @, #, $
- prohibition of words found in a password blacklist
- prohibition of words found in the user's personal information
- prohibition of use of company name or an abbreviation
- prohibition of passwords that match the format of calendar dates, license plate numbers, telephone numbers, or other common numbers

Password Duration

Some policies require users to change passwords periodically, often every 90 or 180 days.

However ...


UNCOMMON (NON-GIBBERISH) BASE WORD
ORDER UNKNOWN
Tr0ub4dor &3
CAPS?
COMMON SUBSTITUTIONS
NUMERAL
PUNCTUATION
(YOU CAN ADD A FEW MORE BITS TO ACCOUNT FOR THE FACT THAT THIS IS ONLY ONE OF A FEW COMMON FORMATS.)
~28 BITS OF ENTROPY
2^28 = 3 DAYS AT 1000 GUESSES/SEC
(PLAUSIBLE ATTACK ON A WEAK REMOTE WEB SERVICE. YES, CRACKING A STOLEN HASH IS FASTER, BUT IT'S NOT WHAT THE AVERAGE USER SHOULD WORRY ABOUT.)
DIFFICULTY TO GUESS:
EASY
WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE Os WAS A ZERO?
AND THERE WAS SOME SYMBOL...
DIFFICULTY TO REMEMBER:
HARD
correct horse battery staple
FOUR RANDOM COMMON WORDS
~44 BITS OF ENTROPY
2^44 = 550 YEARS AT 1000 GUESSES/SEC
DIFFICULTY TO GUESS:
HARD
THAT'S A BATTERY STAPLE.
CORRECT!
DIFFICULTY TO REMEMBER:
YOU'VE ALREADY MEMORIZED IT
THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.