MSC 597C Cryptography

Lecture 3

Henry Brooks

June 9, 2020

Overview

- Review
  - Homework issues
  - Ciphers from last class
- Today
  - One-Time Pad (OTP)
  - Perfect Secrecy
  - Block Cipher vs. Stream Cipher

Historical Ciphers

- Caesar cipher (shift cipher)
  - Secret key is a **number**
  - Key space (number of secret keys) is 25
  - Brute force attack
- Monoalphabetic substitution cipher
  - Secret key is a **one-to-one function**
  - Key space is 26!
  - Frequency analysis (single letter, digrams, trigrams)
- Vigenere cipher (Polyalphabetic substitution cipher)
  - Secret key is a repeating **word** (e.g., CRYPTO $\rightarrow$ 2,17,24,15,19,14)
  - Substitution rules consist of 26 Caesar ciphers with shifts of 0 through 25
  - Strength is multiple ciphertext letters for each plaintext letter
  - Problem is frequency info can still be identified if the length of the keyword is determined

Symmetric Ciphers: Definition

$K \rightarrow$ Key space (i.e., set of all possible keys)

$M \rightarrow$ Message space (i.e., set of all possible messages)

$C \rightarrow$ Ciphertext space (i.e., set of all possible ciphertexts)

Definition

*A **cipher** defined over (K, M, C) is a pair of efficient algorithms (E, D) where $EK \times M \rightarrow C, D : K \times C \rightarrow M$ such that $\forall m \in M, k \in K$ it holds:*

$$E(k, m) = c \text{ and } D(k, c) = m$$

The last equation also known as **Correctness property** can be rewritten as:

$$D(k, E(k, m)) = m$$

Symmetric Ciphers:

The One Time Pad (OTP) (Vernam 1917)

The OTP cipher

$K = M = C = \{0,1\}^n$ i.e., the key, the message and the ciphertext have all the same length n

$$E : K \times M \to C, \quad E(k,m) = k \oplus m$$

$$D : K \times C \to M, \quad D(k,c) = k \oplus c$$

Correctness of the OTP

$D(k, E(k,m)) \stackrel{?}{=} m$

Indeed since: $D(k, E(k,m)) = k \oplus E(k,m) = k \oplus (k \oplus m) = m$

Example

m	0	1	1	0
k	1	0	1	0
C	1	1	0	0

Question

You are given a message m and its OTP encryption c Can you compute the OTP key from m and c?

A. No, I cannot compute the key.
B. Yes the key is $k = m \oplus c$.
C. I can only compute half the bits of the key
D. Yes, the key is $k = m \oplus m$.

Information Theoretic Security

Perfect Secrecy (Shannon 1949)

Intuition

The ciphertext should reveal no "info" about the plaintext

Definition

A cipher (E, D) *defined over* (K, M, C) *has* ***perfect secrecy*** *if:*

$$\forall\, m_0, m_1 \in M, \quad len(m_0) = len(m_1), \quad \text{and } \forall c \in C$$

$$P\left[E(k, m_0) = c\right] = P[E(k, m_1) = c]$$

where k *is chosen uniformly at random from* K *(i.e.,* $k \leftarrow K$*)*

(What does this mean?)

- The adversary that sees only the ciphertext c is not able to determine whether c is an encryption of m_0 or m_1.

What does 'uniformly at random' mean?

Definition

A probability distribution P over a finite set X is a function $P : X \to [0, 1]$ such that $\sum_{x \in X} P(x) = 1$

A probability distribution is called **uniform**, U, if its value $U(x)$ is the same for all x in X.

Example - Uniform distribution

- When flipping a coin the probability of getting head is $1/2$ and equal to the probability of getting tails where $X = [head, tails]$ and $U(head) = U(tails) = 1/2$
- When throwing a dice the probability to get any side of the dice $(1, 2, 3, 4, 5, 6)$ is equal to $1/6$.

Question

let $m \in M$ and $c \in C$. How many OTP keys $k \in K$ map m to c?

A. None
B. 1
C. 2
D. Depends on m

OTP - The Verona Project (1943-1980)

A program of the USA Signal Intelligence Service (i.e., NSA) run to decrypt messages encrypted by the Russian intelligence services (e.g., KGB) during **World War II**.

Encryption: The Russians used a code (ASCII) to convert letters to numbers and the OTP to encrypt the messages.

Reminder: The OTP when used correctly is **unbreakable**

But...

- The Russians generated the one-time pads (keys) with a very laborious process.
- A human was throwing a dice and collected the dice throws to generate the one time keys!
- **Mistake:** They used some one-time-pads (keys) twice!
- The NSA was able to intercept and decrypt about 3000 messages!

The Verona Project: How was the Russian's OTP broken?

Lets assume that the same key is used to encrypt two messages:

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

Then we have:

$$\begin{aligned} c_1 \oplus c_2 &= (m_1 \oplus k) \oplus (m_2 \oplus k) \\ &= (m_1 \oplus m_2) \oplus (k \oplus k) \\ &= m_1 \oplus m_2 \end{aligned}$$

If we have $m_1 \oplus m_2$ we can easily recover m_1 and m_2.

Why?

- Redundancy in English language
- Not all combinations of letters are likely or possible

Remember: Never use the same key in OTP!

Example

Let language

$$L = \{00100, 10011, 11100, 10100\}$$

If we have $m_1 \oplus m_2 = 10111$ then we can deduce that:

$m_1 = 00100$ and $m_2 = 10011$, since only by XORing m_1 and m_2 can we get $m_1 \oplus m_2$

How can we define the security of a cipher?

Step 1: Decide what the attacker can do (power) and what he wants to achieve (goal).

Power

- See ciphertexts?
- Encrypt some chosen messages?
- Decrypt some chosen messages?

Goal

- Decrypt messages
- Encrypt messages
- Find the secret key

How can we define the security of a cipher?

Step 2: Select the approach to secure the cipher

Information theoretic security

Even if the attacker had infinite time and computational power to try all possible keys he can determine the key only with probability $\frac{1}{N}$, where $N = \#$ of possible keys.
Example: One Time Pad

Complexity-based security

Used often in crypto. An attacker that can perform an attack can also break a very complex/hard problem.

What is a secure cipher?

Attacker's power

- Can get one ciphertext

Attacker's goal

- Break the cipher!

How to define a secure cipher?

Attempt 1

The attacker should not be able to recover the key!

$$E(k, m) = m$$

Attempt 2

The attacker should not be able to recover the whole plaintext!

$$E(k, m_0||m_1) = m_0||(k \oplus m_1)$$

Security requirement satisfied! Encryption algorithm not secure!

Block vs Stream Ciphers

Block Ciphers

- Must be given a minimum amount of data
- Typical symmetric cipher blocks: 64 or 128 bits
- If not enough data to fill a block, must either
  - Wait for more data, or
  - Pad the block with extra bits

Stream Ciphers

- Work in small units - bits or bytes
- Bit-oriented stream cipher: one bit in, one bit out
- Consider interactive terminal session...

Steganography

Dear George,
Greetings to all at Oxford. Many thanks for your letter and for the Summer examination package. All Entry Forms and Fee Forms should be ready for final despatch to the Syndicate by Friday 20th or at the very latest, I'm told, by the 21st. Admin has improved here, though there's room for improvement still; just give us all two or three more years and we'll really show you! Please don't let these wretched 16+ proposals destroy your basic O and A pattern. Certainly this sort of change, if implemented immediately, would bring choas.
Sincerely Yours.

Figure 2.9 A Puzzle for Inspector Morse

(from *The Silent World of Nicholas Quinn*, Colin Dexter)

Steganography

Dear George,
Greetings to all at Oxford. Many thanks for your
letter and for the Summer examination package.
All Entry Forms and Fee Forms should be ready
for final despatch to the Syndicate by Friday
20th or at the very latest, I'm told, by the 21st.
Admin has improved here, though there's room
for improvement still; just give us all two or three
more years and we'll really show you! Please
don't let these wretched 16+ proposals destroy
your basic O and A pattern. Certainly this
sort of change, if implemented immediately,
would bring chaos.
Sincerely Yours.

Figure 2.9 A Puzzle for Inspector Morse

(from *The Silent World of Nicholas Quinn*, Colin Dexter)

Other Steganography Techniques

- Character marking
  - Selected letters of printed or typewritten text are over-written in pencil
  - The marks are ordinarily not visible unless the paper is held at an angle to bright light
- Invisible ink
  - A number of substances can be used for writing but leave no visible trace until heat or some chemical is applied to the paper
- Pin punctures
  - Small pin punctures on selected letters are ordinarily not visible unless the paper is held up in front of a light
- Typewriter correction ribbon
  - Used between lines typed with a black ribbon, the results of typing with the correction tape are visible only under a strong light

Steganography vs. Encryption

- Steganography has a number of drawbacks when compared to encryption
  - It requires a lot of overhead to hide a relatively few bits of information
  - Once the system is discovered, it becomes virtually worthless
- The advantage of steganography
  - It can be employed by parties who have something to lose should the fact of their secret communication (not necessarily the content) be discovered
- Encryption flags traffic as important or secret or may identify the sender or receiver as someone with something to hide