T-79.5501
Cryptology

Notes from Lecture 2:

- Entropy of key
- Unicity Distance
- Design principles for symmetric ciphers
- Modular arithmetic

Key length

key length in bits = key entropy
if and only if the keys are chosen equiprobably

Example. Bluetooth PIN

Maximum length 128 bits.

Maximum entropy = 128 bits never achieved in practise.

Two reasons:

1) User selects PIN (in a hurry, to set up a connection)
2) Encoding of keypad characters. Each character takes 8 bits => PIN has at most 16 characters.

   Numeric PIN: max entropy ~ $16 \log_2 10 \sim 53$

   Alphanumeric PIN: max entropy = $16 \log_2 36 \sim 83$

Ciphertext only attack

How much ciphertext is needed to determine the key from ciphertext only? (assuming no bounds on the computations adversary needs to make)

Example: Exhaustive key search given a ciphertext. With each possible key candidate perform decryption, and see if the result makes sense. Works only if plaintext not completely random.

Shift cipher, ciphertext: WNAJW

d_5(WNAJW) = river; d_{22}(WNAJW) = arena

Key is not uniquely determined.

Using statistical characteristics of plaintext language we can determine how long plaintext must be, on the average, to determine the key uniquely.

Theorem 2.10

Let $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ be a cryptosystem. Then

$H(\mathbf{K}|\mathbf{C}) = H(\mathbf{K}) + H(\mathbf{P}) - H(\mathbf{C})$.

Proof: K and P independent =>

$H(\mathbf{K},\mathbf{P},\mathbf{C}) = H(\mathbf{K},\mathbf{P}) = H(\mathbf{K}) + H(\mathbf{P})$.

On the other hand,

$H(\mathbf{K},\mathbf{P},\mathbf{C}) = H(\mathbf{K},\mathbf{C}) = H(\mathbf{C}) + H(\mathbf{K}|\mathbf{C})$. □

Example 2.3 Continued

$H(\mathbf{P}) \approx 0.81$

$H(\mathbf{K}) \approx 1.5$

$H(\mathbf{C}) \approx 1.85$

Thm 2.10 tells $H(\mathbf{K}|\mathbf{C}) = 0.81 + 1.5 - 1.85 = 0.46$.

Can be computed also directly:

$H(\mathbf{K}|\mathbf{C}) = \sum_y \Pr[y]\, H(\mathbf{K}|y) =$

$1/8 \cdot H(\mathbf{K}|1) + 7/16 \cdot H(\mathbf{K}|2) + 1/4 \cdot H(\mathbf{K}|3) + 3/16 \cdot H(\mathbf{K}|4)$

where, e.g. $H(\mathbf{K}|3) = -\frac{3}{4}\log_2(\frac{3}{4}) - \frac{1}{4}\log_2(\frac{1}{4}) \approx 0.8$,

since $\Pr[K_1|3] = 0$, $\Pr[K_2|3] = \frac{3}{4}$ and $\Pr[K_3|3] = \frac{1}{4}$

Conclusion: The average uncertainty about the key is 0.46 bits if one ciphertext character is given.

Entropy of language

Definition 2.7: Suppose L is a language. The entropy of L is defined as

$$H_L = \lim_{n\to\infty} H(\mathbf{P}^n)/n.$$

Here **P** denotes the random variable of one character, $\mathbf{P}^2$ the random variable of two characters, …, $\mathbf{P}^n$ a word of n characters. Let $\mathscr{P}$ be the set of possible characters. Then it follows from Thm 2.6 and Cor 2.9 that

$$H(\mathbf{P}^n) \le n\, H(\mathbf{P}) \le n \log_2|\mathscr{P}|, \text{ for all } n,$$

with equalities if and only if the language is purely random. It follows that $H_L \le \log_2|\mathscr{P}|$.

Redundancy of language

Redundancy R_L of L is defined as

$$R_L = 1 - H_L / \log_2|\mathscr{P}|.$$

Example. L English, $\mathscr{P}$ alphabet of 26 characters,

$\log_2|\mathscr{P}| \approx 4,7$

$H(\mathbf{P}) \approx 4,15$

$H(\mathbf{P}^2)/2 \approx 3,62$

$H(\mathbf{P}^3)/3 \approx 3.22$ …

$H_L \approx 1,5$ (one estimate)

Unicity distance (Def 2.8)

Assume $|\mathcal{P}| = |\mathcal{C}|$. Then

$H(\mathbf{C}^n) - H(\mathbf{P}^n) \approx n\log_2|\mathcal{C}| - nH_L$

$\approx n\log_2|\mathcal{C}| - (n\log_2|\mathcal{P}| - nR_L \log_2|\mathcal{P}|) = n R_L\log_2|\mathcal{P}|$.

From Thm 2.10 we get

$H(\mathbf{K}|\mathbf{C}^n) \approx H(\mathbf{K}) - n R_L\log_2|\mathcal{P}|$

$= \log_2|\mathcal{K}| - n R_L\log_2|\mathcal{P}|$,

which gives an estimate of the entropy of the key given n characters of ciphertext. The key is uniquely determined exactly if $H(\mathbf{K}|\mathbf{C}^n) = 0$. This happens approximately for $n = n_0$, where

$$n_0 = \log_2|\mathcal{K}| / R_L\log_2|\mathcal{P}|$$

Example: see separate note.

Stream ciphers

Let ($\mathcal{P}$,$\mathcal{C}$,$\mathcal{K}$,$\mathcal{L}$,$\mathcal{E}$,$\mathcal{D}$, g) be a synchronous stream cipher (Definition 1.6)

$g(K,i) = z_i$ key-stream generation

$y_i = e_{zi}(x_i)$ encryption

$x_i = d_{zi}(y_i)$ decryption

Requirement:

Key-stream $\{z_i\}$ should be indistinguishable from one-time-pad

Block Ciphers

A block cipher is a cryptosystem ($\mathcal{P}$,$\mathcal{C}$,$\mathcal{K}$, $\mathcal{E}$,$\mathcal{D}$), for which it is typical that the same encryption operation e_K is applied to a number of consequent data blocks.

Even if $H(\mathbf{K}|\mathbf{C}^n)=0$ it should be computationally infeasible to solve for the key given ciphertext and any known plaintext features.

Shannon: Design the encryption operation for a block cipher as a composition of different transformations which produce diffusion and confusion.

Modular arithmetic

Given a positive integer m and any two integers a and b, we say that a is congruent to b modulo m, if m divides b – a. We then denote $a \equiv b \pmod{m}$.

When a is divided by m, there is a unique remainder, that is, an integer r, $0 \le r < m$, such that $a = km + r$, or what is equivalent, $a \equiv r \pmod{m}$. We also denote $r = a \bmod m$. We identify a with its remainder modulo m and compute with remainders modulo m.

Solving an equation mod m

Assume gcd(a,m) = 1. If $ax \equiv ay \pmod{m}$, then $x \equiv y \pmod{m}$. It follows that

$$\{ax \bmod m \mid x = 0,1,\ldots,m-1\} = \{0,1,\ldots,m-1\} = \mathbf{Z}_m,$$

which means that for all b in $\mathbf{Z}_m$, the equation

$$ax \equiv b \pmod{m} \qquad (1)$$

has a unique solution.

If gcd(a,m) = d, then the equation (1) has a solution if and only if d divides b. Then the number of solutions is d. To solve the equation (1), divide it first by d to get:

$$(a/d)x \equiv b/d \pmod{m/d}. \qquad (2)$$

Then gcd (a/d,m/d) = 1, and (2) has a unique solution x_0 modulo m/d. This gives d solutions mod m. The are:

$$x_0,\ x_1=x_0+m/d,\ x_2=x_0+2m/d,\ \ldots,x_{d-1}=x_0+(d-1)m/d.$$

Inverse mod m

It follows that equation

$$ax \equiv 1 \pmod{m}$$

has a solution if and only if $\gcd(a,m) = 1$. If a solution exists it is unique, and we denote it by $x = a^{-1} \bmod m$. It is the multiplicative inverse of element a modulo m.

Euclidean algorithm

see text-book 5.2.1

The Chinese Remainder Theorem

see text-book 5.2.2