Introduction: Classical Cryptography

Guevara Noubir
http://www.ccs.neu.edu/home/noubir/Courses/CSG252/F04

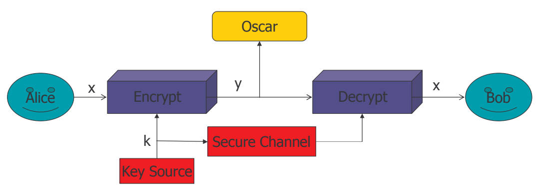
Textbook: "Cryptography: Theory and Applications",

Douglas Stinson, Chapman & Hall/CRC Press, 2002

Reading: Chapter 1

Simple Cryptosystems

- Goal:
  - Allow two entities (e.g., Alice, and Bob) to communicate over an insecure channel, such that an opponent (e.g., Oscar) cannot understand what is being communicated



Definition of Cryptosystem

- Definition:
  - A cryptosystem is a five-tuple (*P, C, K, E, D*) s.t. the following conditions are satisfied:
    1. *P* is a finite set of possible *plaintexts*
    2. *C* is a finite set of possible *ciphertexts*
    3. *K*, the keyspace, is a finite set of possible *keys*
    4. For each key k, there exists an encryption rule $e_k \in E$, and decryption rule $d_k \in E$ s.t. $d_k(e_k)$ = Identity
- Encoding a message:
  - $x = x_1 x_2 \ldots x_n \rightarrow y = x_1 x_2 \ldots x_n = e_k(x_1) e_k(x_2) \ldots e_k(x_n)$
- Note:
  - Each encryption function has to be injective

Review of Basics of Modular Arithmetic

- Congruence:
  - a, b: integers; m: positive integer
  - $a \equiv b \bmod m$ iff m divides $a-b$
  - a is said to be congruent to b mod m
  - Example: $101 \equiv 3 \bmod 7$
- Arithmetic modulo m:
  - $Z_m=\{0, 1, \ldots, m-1\}$; +, x operations
  1. Addition is closed
  2. Addition is commutative
  3. Addition is associative
  4. 0 is an additive identity
  5. Additive inverse of a is $m-a$
  6. Multiplication is closed
  7. Multiplication is commutative
  8. Multiplication is associative
  9. 1 is a multiplicative identity
  10. The distributive property is satisfied
- 1-5 $\Rightarrow Z_m$ is an abelian group
- 1-10 $\Rightarrow Z_m$ is a ring
- Other examples of rings: …

Shift Cipher

- Definition:
  - $P = C = K = Z_{26}$
  - $e_k(x) = (x+k) \bmod 26$
  - $d_k(x) = (x-k) \bmod 26$
- Example:
  - $k = 3$ is often called *Caesar Cipher*
- Alphabet encoding:

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Desired Properties of Cryptosystems

- Encryption and Decryption function can be efficiently computed
- Given a ciphertext y, it should be "difficult" for an opponent to identify the encryption key k, and the plaintext x
- How about the security of the shift cipher?
- Example:
- Average time to identify the encryption key?
- Conclusion about the key space?

Substitution Cipher

- Definition:
  - $P = C = Z_{26}$
  - K: set of all possible permutations of the P
  - $e_\pi(x) = \pi(x)$
  - $d_\pi(y) = ?$
- Example:

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

- Key Space:
  - $|K| = ?$

Affine Cipher

- Encryption function of the form:
  - $e(x) = (ax + b) \bmod 26$
- Conditions on (a, b)?
- Examples:
  - (a, b) = (2, 5)
  - (a, b) = (3, 5)

Affine Cipher

- Theorem:
  - The congruence $ax \equiv b \pmod{m}$ has a unique solution $x \in Z_m$ iff gcd(a, m) = 1
- Definition:
  - For a>1, $m \geq 2$, if gcd(a, m) = 1 then a and m are said to be relatively prime (co-prime).
  - The number of integers in Z_m that are relatively prime to m is denoted by $\phi(m)$: Euler phi-function (a.k.a totient function).

Affine Cipher

- Theorem:
  - If $m = p_1^{e1} p_2^{e2} \cdots p_n^{en} \Rightarrow \phi(m) = (p_1^{e1} - p_1^{e1-1}) \ldots (p_n^{en} - p_n^{en-1})$ where p_i's are distinct primes and the e_i's are strictly positive integers
- Corollary:
  - The key space of affine ciphers is: $m\phi(m)$
- Definition:
  - For $a \in Z_m$, we denote by a^{-1} the multiplicative inverse of a s.t. $a^{-1} \in Z_m$ and $a\, a^{-1} \equiv a^{-1}a \equiv 1 \bmod m$
- Theorem:
  - a has an inverse iff gcd(a, m) = 1
  - If m is prime every element of Z_m has an inverse and Z_m is called a field

Affine Cipher

- Definition:
  - $P = C = Z_{26}$
  - $K = \{(a, b) \in Z_{26} x Z_{26} : gcd(a, 26) = 1\}$
  - For $k = (a, b) \in K$
    - $e_k(x) = (ax+b) \bmod m$
    - $d_k(y) = ?$
- Example:
  - k = (7, 3)

Vigenère Cipher

- Monoalphabetic cryptosystems:
  - For a given key: each alphabetic character is mapped to a *unique* alphabetic Character
  - E.g., shift cipher, substitution cipher, affine cipher
- Polyalphabetic crypotosystems
- Vigenere cipher
  - m: positive integer; $P = C = K = (Z_{26})^m$
  - For $k = (k_1, k_2, \ldots, k_m)$:
    - $e_k(x_1, \ldots, x_m) = (x_1+k_1, \ldots, x_m+k_m)$
    - $d_k(y_1, \ldots, y_m) = (y_1-k_1, \ldots, y_m-k_m)$
- Key space:

Hill Cipher

- $m \geq 2$ positive integer; P = C = $(Z_{26})^m$
- Idea: take m linear combinations of the m alphabetic characters of the plaintext
- Example:

$$k = \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix}$$

- Condition?

Hill Cipher

- Definition:
  - m: positive integer; $P = C = (Z_{26})^m$
  - $K = \{m \times m$ invertible matrices over $Z_{26}\}$
  - $e_k(x) = xk$
  - $d_k(y) = yk^{-1}$
- k^{-1} = ?
- det k = ?

Permutation Cipher

- Definition:
  - m: positive integer; P = C = $(Z_{26})^m$
  - K = {π: permutation of {1…m}}
  - $e_k(x_1, \ldots, x_m) = (x_{\pi(1)}, \ldots, x_{\pi(m)})$
  - $d_k(y_1, \ldots, y_m) = (y_{\pi^{-1}(1)}, \ldots, y_{\pi^{-1}(m)})$
- Example:
- Permutation matrix and it's inverse

Stream Ciphers

- Block Ciphers: $y = y_1 y_2 \ldots = e_k(x_1)\, e_k(x_2) \ldots$
- Stream Ciphers:
  - Generate a Keystream: $z = z_1 z_2 \ldots$
  - Encryption: $y = y_1 y_2 \ldots = e_{z1}(x_1)\, e_{z2}(x_2) \ldots$
- Synchronous Stream Cipher:
  - Keystream does not depend on the plaintext
- Definition of Synchronous Stream Cipher
  - A tuple (P, C, K, L, E, D), and a function g s.t.:
    - P (resp. C): finite set of possible plaintexts (resp. ciphertexts)
    - K: keyspace (finite set of possible keys)
    - L: finite set called keystream alphabet
    - g: keystream generator s.t. $g(k) = z_1 z_2 \ldots$ where $z_i \in L$
    - $\forall z \in L\ \exists e_z \in E,\ d_z \in D$ s.t. $d_z \circ e_z = Id$
- Example: Vigenere Cipher as a synchronous stream cipher

Stream Ciphers (Cont.)

- Periodic Stream Cipher with period *d* iff:
  - $\forall i \geq 1\ z_{i+d} = z_i$
- Example:
  - Vigenere Cipher with keyword length *m* is a periodic stream cipher with period *m*
- Stream ciphers usually have L = Z_2:
  - $e_z(x) = (x+z) \bmod 2$
  - $d_z(x) = ?$

Stream Ciphers: LFSR

- Linear Feedback Shift Register (LFSR) can generate a synchronous linear keystream:
  - $$z_{i+m} = \sum_{j=0}^{m-1} c_j z_{i+j}$$
  - $c_j \in Z_2$, and initializing the registers with k_1, k_2, k_m
- Properties:
  - Linearity (linear combination of previous terms)
  - Degree *m* (depends only on the previous *m* terms)
  - $c_0 = 1$
  - Key is: $(k_1, k_2, \ldots, k_m, c_0, c_1, \ldots, c_{m-1})$
  - $(k_1, k_2, \ldots, k_m)$ should be different from (0, 0, …, 0)
  - If $(c_0, c_1, \ldots, c_{m-1})$ is carefully chosen and $(k_1, k_2, \ldots, k_m) \neq \mathbf{0}$ then the period of the keystream is 2^m-1
- Example: m=4, $(c_0, c_1, c_2, c_3) = (1, 1, 0, 0)$
- Advantages of LFSR: easy to implement in HW,

Non-Synchronous Stream Cipher

- Example: Autokey Cipher
  - $P = C = K = L = Z_{26}$
  - $z_1 = k$; $z_i = x_{i-1}$ (for all $i > 1$)
  - $e_z(x) = (x+z) \bmod 26$
  - $d_z(y) = (y-z) \bmod 26$
- Drawback?

Cryptanalysis

- Kerckhoffs' Principle:
  - The opponent knows the cryptosystem being used (no security through obscurity)
- Definition of attack models:
  - Ciphertext only attack
  - Known plaintext attack
  - Chosen plaintext attack
  - Chosen ciphertext attack
- Objective of the opponent:
  - Identify the secret key

Statistical Cryptanalysis

- Context:
  - Cipher-text only attack
  - Plaintext ordinary English (no punctuation, space)
  - Letters' probabilities (Beker and Piper):
    - A: 0.082, B: 0.015, C: 0.028, …
    - E: 0.120; T, A, O, I, N, S, H, R: [0.06, 0.09]; D, L: 0.04; C, U, M, W, F, G, Y, P, B: [0.015, 0.028]
    - V, K, J, X, Q, Z: < [0.01]
    - 30 most common digrams: TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, TI, IS, ET, IT, AR, TE, SE, HI, OF
    - 12 most common trigrams: THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH

Cryptanalysis of the Affine Cipher

- Example:
  - Ciphertext (57 characters)= FMXVEDKAPHFERBNDKRXRSREFMORUDSDKDVSHVUFEDKAPRK DLYEVLRHHRH
  - Occurences:
    - R:8; D:7; E, H, K:5, F, S, V:4
  - First guess: R: e; D: t
    - $4a + b = 17$; $19a + b = 3 \Rightarrow (a, b) = (6, 19)$ but $\gcd(a, 26) = 2 > 1$ illegal!
  - Second guess: R: e; E: t $\Rightarrow$ a=13 illegal!
  - Third guess: R: e; H: t $\Rightarrow$ illegal!
  - Fourth guess: R:e; K: t $\Rightarrow (a, b) = (3, 5)$
    - Results in plaintext = algorithmsarequitegeneraldefinitionsofarithmeticprocesses

Cryptanalysis of the Substitution Cipher

- Identify possible encryption of e (most common letter)
  - t, a, o, i, n, s, h, r: will probably be difficult to differentiate
- Identify possible digrams starting/finishing with e: -e and e-
- Use trigrams

Cryptanalysis of the Vigenère Cipher

- First step: identify the keyword length (m)
- Kasiski test [Kasiski 1863, Babbage 1854]:
  - Observation:
    - two *identical* segments of plaintext are encrypted to the *same* ciphertext if they are δ positions apart s.t. $\delta = 0 \bmod m$
  - Test:
    - Find all identical segments of length > 3 and record the distance between them: $\delta_1, \delta_2, \ldots$
    - m divides $\gcd(\delta_1, \delta_2, \ldots)$

Index of Coincidence to Find keyword Length

- Index of coincidence:
  - $x = x_1x_2 \dots x_n$; $I_c(x)$ is the probability that two random elements of x are identical
  - Let $f_0, f_1, \dots, f_{26}$ be the number of occurrences of A, B, …, Z in the string x
  - $$I_c(x) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)}$$
  - If x is a string of English text:
  - For a mono-alphabetic cipher $I_c(x)$ is unchanged $I_c(x) \approx \sum_{i=0}^{25} p_i^2 = 0.065$
- Try $m = 1, 2, \dots$
  - Decompose y in substrings: $y_1y_{m+1}y_{2m+1}\dots$; $y_2y_{m+2}y_{2m+2}\dots$; …
  - If for all substrings: I_c is close to 0.065 then m might be the length
- If wrong m, then $I_c \approx 26 / 26^2 = 0.038$

Cryptanalysis of the Vigenère Cipher (Cont.)

- Given the keyword length, each substring:
  - Length: $n' = n/m$
  - Encrypted by a shift: k
  - Probability distribution of letters: $f_0/n', f_1/n', \dots, f_{25}/n'$
- Therefore:
  - $f_k/n', f_{k+1}/n', \dots, f_{k+25}/n'$ should be close to $p_0, \dots, p_{25}$
  - Let: $M_g = \sum_{i=0}^{25} p_0 f_{g+i}$
  - If $g = k$, $M_g \approx 0.065$
  - If $g \neq k$, M_g significantly smaller then 0.065

Cryptanalysis of the Hill Cipher

- More difficult to break with cipher-text only
- Easy with known plaintext
- Goal: Find secret Matrix K
- Assumption:
  - Known: m
  - Known: m distinct plaintext-ciphertext pairs:
    - $(x_i, y_i = e(x_i))$
    - $x_i = (x_{1i}, \dots, x_{mi})$; $y_i = (y_{1i}, \dots, y_{mi})$: $y_i = x_i K$
- Define: Y s.t. rows are y_i (similarly X)
- Y = XK
- If X is invertible $\Rightarrow K = X^{-1}Y$
- What if X is not invertible?

Cryptanalysis of the Hill Cipher

- Example:
  - m = 2;
  - Plaintext: *friday*
  - *Ciphertext: PQCFKU*

$$X = \begin{pmatrix} 5 & 17 \\ 8 & 3 \end{pmatrix}; Y = \begin{pmatrix} 15 & 16 \\ 2 & 5 \end{pmatrix}$$

$$K = X^{-1}Y = \begin{pmatrix} 9 & 1 \\ 2 & 15 \end{pmatrix}\begin{pmatrix} 15 & 16 \\ 2 & 5 \end{pmatrix}$$

$$K = \begin{pmatrix} 7 & 19 \\ 8 & 3 \end{pmatrix}$$

  - Can be verified using the third plaintext-ciphertext pair

Cryptanalysis of the LFSR Stream Cipher

- Known-plaintext attack with known *m*
  - Given: $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$
- Need to compute $c_0, \ldots, c_{m-1}$
  - $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ allow us to compute $z_1, \ldots, z_n$
- If $n \geq 2m$ we can obtain *m* linear equations with *m* unknowns using:

$$z_{i+m} = \sum_{j=0}^{m-1} c_j z_{i+j}$$