Symmetric Key Encryption I

Anders Västberg

vastberg@kth.se

Slides by M. Goodrich, R. Tamassia , Göran Andersson

Module 1 Cryptography

Lecture 2 Symmetric Cryptography

Lecture 3 Symmetric C. cont. and Public Key Cryptography

Lecture 4 Cryptographic Hash and Message Digest

Lecture 5 Digital Signatures

Exercises during lectures

Home Assignment 1: Cryptography and Hashing

Hill Cipher assignment

One-way Hash Project

Cryptographic Concepts

Short Review

Encryption and Decryption

- The message M is called the **plaintext.**
- Alice encrypts M using an algorithm E that outputs a **ciphertext C for M.**
- Bob decrypts C using an algorithm D that outputs the plaintext M.

Encryption and Decryption

As equations:

$$C = E(M)$$

$$M = D(C)$$

Ancient Cryptography: Caesar Cipher

Gaius Julius Caesar (100-44 B.C.)

Used during the Gallic Wars 58-49 B.C.

The cipher does a cyclic shift by three of every letter in the alphabet

Letter	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
Code	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

Any shift between 1 to 25 can be used

Example Caesar Cipher

As equations:

$$C_i = E(M_i) = \text{symbol}(M_i) + 3 \pmod{26}$$

$$M_i = D(C_i) = \text{symbol}(C_i) - 3 \pmod{26}$$

Symmetric Cryptosystems

- Alice and Bob share a secret key, which is used for both encryption and decryption.

Symmetric Key Distribution

- Requires each pair of communicating parties to share a (separate) secret key.

Complete graph

$$\binom{n}{2} \quad \text{keys}$$

$$\binom{n}{2} = \frac{n!}{2!\,(n-2)!} = \frac{n\,(n-1)}{2}$$

Public-Key Cryptography

- Bob has two keys: a **private key,** S_B, which Bob keeps **s**ecret, and a **public key,** P_B, which Bob broadcasts widely.
- Alice encrypts using Bob's public key, P_B,
- $C = E_{P_B}(M)$
- Bob then uses his private key to decrypt the message
- $M = D_{S_B}(C)$.

Public-Key Cryptography

- Separate keys are used for encryption and decryption.

Public Key Distribution

- Only one key pair is needed for each participant

$$2n \quad \text{keys}$$

$$n > 5 \Rightarrow 2n < \binom{n}{2}$$

Symmetric Key Encryption I

Symmetric Cryptography

- Symmetric Cryptography Issues
  - Possible Attacks
  - Brute force attacks and unicity distance
- Substitution ciphers
- S-boxes
- Historic cryptographic methods
- One-time pads
  - Random number generators
- Stream Ciphers

Symmetric Cryptosystem

- Scenario
  - Alice wants to send a message (plaintext M) to Bob.
  - The communication channel is insecure and can be eavesdropped
  - If Alice and Bob have previously agreed on a symmetric encryption scheme and a secret key K, the message can be sent encrypted (ciphertext C)
- Issues
  - What is a good symmetric encryption scheme?
  - What is the complexity of encrypting/decrypting?
  - What is the size of the ciphertext, relative to the plaintext?

Basics

- Notations
  - Secret key K
  - Encryption function $E_K(M)$
  - Decryption function $D_K(C)$
- Form
  - Plaintext length typically the same as ciphertext length
  - Encryption and decryption are permutation functions (bijections) on the set of all n-bit arrays
- Efficiency
  - functions E_K and D_K should have efficient algorithms
- Consistency
  - Decrypting the ciphertext yields the plaintext
  - $D_K(E_K(M)) = M$

Attacks

- Attacker may have
  a) collection of ciphertexts (ciphertext only attack)
  b) collection of plaintext/ciphertext pairs (known plaintext attack)
  c) collection of plaintext/ciphertext pairs for plaintexts selected by the attacker (chosen plaintext attack)
  d) collection of plaintext/ciphertext pairs for ciphertexts selected by the attacker (chosen ciphertext attack)

Brute-Force Attack

- Try all possible keys K and determine if $D_K(C)$ is a likely plaintext
  - Requires some knowledge of the structure of the plaintext (e.g., PDF file or email message)
- Key should be a sufficiently long random value to make exhaustive search attacks unfeasible

Image by Michael Cote from http://commons.wikimedia.org/wiki/File:Bingo_cards.jpg

Encrypting English Text

- English text typically represented with 8-bit ASCII encoding
- A message with *t* characters corresponds to an *n*-bit array, with $n = 8t$

- Redundancy due to repeated words and patterns
  - E.g., “th”, “ing”
- English plaintexts are a very small subset of all n-bit arrays

Entropy of Natural Language

- 8 bit per character, n-bit array plain text, k-bit decryption key
- Information content (entropy, H) of English: 1.25 bits per character (byte) n-bit arrays that are English text:

$$H(X) = 1.25\ n/8 \approx 0.156n = \alpha n$$

- For a natural language, constant $\alpha < 1$ such that there are $2^{\alpha n}$ messages among all n-bit arrays
- Redundancy

$$D = (1 - \alpha)$$

- Fraction (probability) of valid messages in English

$$2^{\alpha n}/2^n = 1/2^{(1-\alpha)n} \approx 0.56^n$$

- Brute-force decryption
  - Try all possible 2^k decryption keys
  - Stop when valid plaintext recognized
- Given a ciphertext, there are 2^k possible plaintexts
- Expected number of valid plaintexts

$$2^k/2^{(1-\alpha)n}$$

- Expected unique valid plaintext, (no spurious keys) achieved at unicity distance

$$1 = 2^k/2^{(1-\alpha)U} \Leftrightarrow U = k/(1-\alpha)$$

- This is minimum #bits to get the key by brute-force

Problem

- Find the unicity distance for a 256 bit key using English.
- $U = k/D = 256/(1 - 0.156) \approx 304$ bits

Substitution Ciphers

- Each letter is uniquely replaced by another.
- There are $26! \approx 4 \times 10^{26}$ possible substitution ciphers.

- One popular substitution "cipher" for some Internet posts is ROT13.

Public domain image from http://en.wikipedia.org/wiki/File:ROT13.png

Frequency Analysis

- Letters in a natural language, like English, are not uniformly distributed.
- Knowledge of letter frequencies, including pairs and triples can be used in cryptologic attacks against substitution ciphers.

a:	8.05%	b:	1.67%	c:	2.23%	d:	5.10%
e:	12.22%	f:	2.14%	g:	2.30%	h:	6.62%
i:	6.28%	j:	0.19%	k:	0.95%	l:	4.08%
m:	2.33%	n:	6.95%	o:	7.63%	p:	1.66%
q:	0.06%	r:	5.29%	s:	6.02%	t:	9.67%
u:	2.92%	v:	0.82%	w:	2.60%	x:	0.11%
y:	2.04%	z:	0.06%				

Letter frequencies in the book *The Adventures of Tom Sawyer*, by Twain.

Ancient Cryptography: Polybius Square

Polybius Greek Historian (c. 208- c. 125 B.C.)

Use numbers to represent letters

D => 14

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	IJ	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

Example:
Meet me at the olive tree => 3215154432151144442315343124511544421515

Substitution Boxes

- Substitution can also be done on binary numbers.
- Such substitutions are usually described by substitution boxes, or S-boxes.

	00	01	10	11
00	0011	0100	1111	0001
01	1010	0110	0101	1011
10	1110	1101	0100	0010
11	0111	0000	1001	1100

(a)

	0	1	2	3
0	3	8	15	1
1	10	6	5	11
2	14	13	4	2
3	7	0	9	12

(b)

Figure 8.3: A 4-bit S-box (a) An S-box in binary. (b) The same S-box in decimal.

Trithemius Progressive Key

German Abbot, Magician and Alchemist (1462-1516)

Example of a polyalfabetic cipher which counters frequency analysis

Principle first explain by Leon Battista Alberti (1404-1472)

Trithemius Progressive Key

0	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
1	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	A
2	C	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
3	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
4	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	D
5	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
6	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
7	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
8	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
9	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
10	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
11	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
12	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
13	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
14	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
15	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
16	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
17	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
18	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
19	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
20	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
21	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
22	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
23	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
24	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
25	Z	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

Vigenère Cipher, Published 1586

Encryption:

message = "ATTACKATDAWN"
key = "LEMON"

$\mathbb{m} = \{0, 19, 19, 0, 2, 10, 0, 19, 3, 0, 22, 13\}$

$\mathbb{k} = \{11, 4, 12, 14, 13, 11, 4, 12, 14, 13, 11, 4\}$

$\mathbb{c} = (\mathbb{k} + \mathbb{m}) \bmod 26$

$\{11, 23, 5, 14, 15, 21, 4, 5, 17, 13, 7, 17\}$

cipher = "LXFOPVEFRNHR"

Decryption:

$(\mathbb{c} - \mathbb{k}) \bmod 26$

Problem

- Find the Vigenère Cipher
  - message = ATTACKNOW
  - key = APPLE
- message = {0, 19, 19, 0, 2, 10, 13, 14, 22}
- key = {0, 15, 15, 11, 4, 0, 15, 15, 11}
- cipher = {0, 8, 8, 11, 6, 10, 2, 3, 7}
- cipher = AIILGKCDH

One-Time Pads

- There is one type of substitution cipher that is absolutely unbreakable.
  - The **one-time pad** was invented in 1917
  - We use a block of shift keys, $(k_1, k_2, \ldots, k_n)$, to encrypt a plaintext, M, of length n, with each shift key being chosen uniformly at random.
- Since each shift is random, every ciphertext is equally likely for any plaintext.

Pseudo Random Number Generators

Desired properties for a PRNG

- Uniform distribution
- Independent numbers
- Very long period
- A simple PRNG is a linear congruential generator

$$x_{i+1} = a\,x_i + b \bmod m$$

- $a \in [1, m-1]$, $b \in [0, m-1]$

Linear Congruential Generator

$$x_{i+1} = a\,x_i + b \bmod m$$

- The period of this generator is at most $\phi(m)$
- Choosing a as a primitive root to m gives max period
- x_0 and b has to be chosen properly
- If m is a prime p then $\phi(p) = p - 1$ is the max period
- If $b = 0$ then $x_i = 0$ can be avoided and all values $[1,\ p - 1]$ are represented

Linear Congruential Generator

$$x_{i+1} = a\,x_i + b \bmod m$$

- If $m = p = 7$ then the primitive roots are $\{3, 5\}$
- If $a=3$ which is a primitive root of p, then the max period is $\phi(7) = 7 - 1 = 6$
- $b = 0$ and $x_0 \neq 0$ results in a uniform sequence $[1, 6]$
- I.e. we've got a die
- However this sequence is completely predictable
  {3, 2, 6, 4, 5, 1, 3, 2, 6, 4, 5, 1, 3, 2, 6, 4, 5, 1, 3, 2, 6, 4, 5, 1, 3}

- We have to use very large p and change a
- Then use $r_i = x_i / p$ that results in $r_i \in (0, 1)$
- Finally use $y_i = \text{floor}(6\,r_i + 1)$ for our die

Problem

- Make a primitive die in the interval [1,10]

- If $m = p = 11$ then the primitive roots are $\{2, 6, 7, 8\}$
- $a = 2$ gives max period $\phi(11) = 11 - 1 = 10$
- $b = 0$ and $x_0 \neq 0$ results in a uniform sequence $[1, 10]$
- I.e. we've got our predictable die
  {2, 4, 8, 5, 10, 9, 7, 3, 6, 1, 2, 4, 8, 5, 10, 9, 7, 3, 6, 1, 2, 4, 8, 5}

Stream Cipher

- Key stream
  - Pseudo-random sequence of bits $S = S[0], S[1], S[2], \ldots$
  - Can be generated on-line one bit (or byte) at the time
- Stream cipher
  - XOR the plaintext with the key stream $C[i] = S[i] \oplus M[i]$
  - Suitable for plaintext of arbitrary length generated on the fly, e.g., media stream
- Synchronous stream cipher
  - Key stream obtained only from the secret key K
  - Works for unreliable channels if plaintext has packets with sequence numbers
- Self-synchronizing stream cipher
  - Key stream obtained from the secret key and q previous ciphertexts
  - Lost packets cause a delay of q steps before decryption resumes

Stream Cipher (Binary Pad)

Encryption:

message = "This secret message cannot be revealed!"

$\mathbb{m}$ = {84, 104, 105, 115, 32, 115, 101, 99, 114, 101, 116, 32, 109,
101, 115, 115, 97, 103, 101, 32, 99, 97, 110, 110, 111, 116,
32, 98, 101, 32, 114, 101, 118, 101, 97, 108, 101, 100, 33}

$\mathbb{s}$ = {149, 134, 11, 33, 12, 64, 182, 249, 26, 136, 199, 132, 171,
64, 36, 54, 149, 200, 110, 145, 84, 25, 217, 179, 246, 145,
149, 210, 81, 13, 254, 55, 93, 6, 46, 23, 133, 78, 224}

$\mathbb{c} = \mathbb{m} \oplus \mathbb{s}$

{193, 238, 98, 82, 44, 51, 211, 154, 104, 237, 179, 164, 198, 37,
87, 69, 244, 175, 11, 177, 55, 120, 183, 221, 153, 229, 181, 176,
52, 45, 140, 82, 43, 99, 79, 123, 224, 42, 193}

cipher = "ÁîbR,3Ó hí³¤Æ%WEô¯ ±7x·Ý åµ°4– R+cO{à∗Á"

Decryption:

$\mathbb{r} = \mathbb{c} \oplus \mathbb{s}$

Key Stream Generation

- RC4
  - Designed in 1987 by Ron Rivest for RSA Security
  - Trade secret until 1994
  - Uses keys with up to 2048 bits
  - Simple algorithm
  - Do not fulfill the highest security standards

Summary Symmetric Cryptography I

- Possible Attacks
  - Brute force attacks and unicity distance
- One-time pads
- S-box
- Stream ciphers

That's all folks!