Introduction to Cryptography
ECE 597XX/697XX

Part 1

Introduction

Israel Koren

Course Outline

♦ I. Introduction
♦ II. Stream Ciphers
♦ III. Block Ciphers: DES
♦ IV. AES
♦ V. Other Block Ciphers
♦ VI. Public Key Ciphers
♦ VII. RSA
♦ VIII. Discrete Logarithm
♦ IX. ECC
♦ X. Digital Signatures
♦ XI. Hash Functions
♦ XII. Message Authentication Codes (MACs)
♦ XIII. Key Establishment
Reference books

♦ Recommended books:

Further Reading and Information


History of Cryptography (great bedtime reading)

Software (excellent demonstration of many ancient and modern ciphers)
♦ Cryptool, http://www.cryptool.de
Administrative Details

♦ Instructor: Prof. Israel Koren
♦ KEB 309E, Tel. 545-2643
♦ Email: koren@ecs.umass.edu
♦ Office Hours: 4:00-5:00 pm, Tues. & Thur.
♦ Course web page: Lecture notes, exercises (OWL) and other details regarding the course available at: http://euler.ecs.umass.edu/ece597

Grading

♦ Midterm - 30%
  • Wed. Oct. 26, 2:30-4:30pm (tentative)
♦ Homework (OWL quizzes) - 15%
♦ Final Exam - 55%

Content of this part

♦ Overview on the field of cryptology
♦ Basics of symmetric cryptography
♦ Cryptanalysis
♦ Substitution Cipher
♦ Modular arithmetic
♦ Shift (or Caesar) Cipher and Affine Cipher
♦ Other historic ciphers, e.g., Enigma
Classification of the field of Cryptology

Some Historical Facts

- **Ancient Crypto**: Early signs of encryption in Egypt in 2000 BC. Letter-based encryption schemes (e.g., Caesar cipher) popular ever since.
- **Symmetric ciphers**: All encryption schemes from ancient times until 1976 were symmetric ones.
- **Asymmetric ciphers**: In 1976 public-key (or asymmetric) cryptography was proposed by Diffie, Hellman and Merkle.
- **Hybrid Schemes**: The majority of today's protocols are hybrid schemes, i.e., they use both
  - symmetric ciphers (e.g., for encryption and message authentication) and
  - asymmetric ciphers (e.g., for key exchange and digital signature).
Symmetric Cryptography

- Alternative names: private-key, single-key or secret-key cryptography.

Alice (good) \( x \) \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow Unsecure channel (e.g. Internet) \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow Bob (good)

- Problem Statement:
  1) Alice and Bob would like to communicate via an unsecure channel (e.g., WLAN or Internet).
  2) A malicious third party Oscar (the bad guy) has channel access but should not be able to understand the communication.

\[ x \] is the plaintext
\[ y \] is the ciphertext
\[ K \] is the key

Set of all keys \( \{K_1, K_2, \ldots, K_n\} \) is the key space

Encryption with symmetric cipher:

\[ e( ) \]

Decryption
\[ d( ) \]

Secure Channel

Oscar (bad guy)

Unsecure channel (e.g. Internet)
Symmetric Cryptography

- Encryption and decryption are inverse operations if the same key $K$ is used on both sides: $d_K(y) = d_K(e_K(x)) = x$

- Important: The key must be transmitted via a secure channel between Alice and Bob.
- The secure channel can be realized, e.g., by manually installing the key for the Wi-Fi Protected Access (WPA) protocol or a human courier. However, the system is only secure if an attacker does not learn the key $K$

$\Rightarrow$ The problem of secure communication is reduced to secure transmission and storage of the key $K$.

Why do we need Cryptanalysis?

- There is no mathematical proof of security for any practical cipher
- The only way to have assurance that a cipher is secure is to try to break it (and fail)

Kerckhoff’s Principle is paramount in modern cryptography:

A cryptosystem should be secure even if the attacker (Oscar) knows all details about the system, with the exception of the secret key.

- In order to achieve Kerckhoff’s Principle in practice:
  Only use widely known ciphers that have been cryptanalyzed for several years by good cryptographers!
- Remark: It is tempting to assume that a cipher is more secure if its details are kept secret. However, history has shown time and again that secret ciphers can almost always be broken once they have been reversed engineered. (Example: Content Scrambling System (CSS) for DVD content protection.)
Cryptanalysis: Attacking Cryptosystems

- **Classical Attacks**
  - Mathematical Analysis
  - Brute-Force Attack
- **Implementation Attack**: Try to extract key through reverse engineering or power measurement, e.g., for a banking smart card.
- **Social Engineering**: E.g., trick a user into giving up her password

Brute-Force Attack (or Exhaustive Key Search) against Symmetric Ciphers

- Treats the cipher as a black box
- Requires (at least) 1 plaintext-ciphertext pair \((x_0, y_0)\)
- Check all possible keys until condition is fulfilled:
  \[ d_k(y_0) = x_0 \]
- How many keys do we need?

<table>
<thead>
<tr>
<th>Key length in bits</th>
<th>Key space</th>
<th>Security life time (assuming brute-force as best possible attack)</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>(2^{64})</td>
<td>Short term (few days or less)</td>
</tr>
<tr>
<td>128</td>
<td>(2^{128})</td>
<td>Long-term (several decades in the absence of quantum computers)</td>
</tr>
<tr>
<td>256</td>
<td>(2^{256})</td>
<td>Long-term (also resistant against quantum computers - note: QC may never exist)</td>
</tr>
</tbody>
</table>

An adversary only needs to succeed with one attack. A long key space does not help if other attacks (e.g., social engineering) are possible.
Substitution Cipher (Historical cipher)

Great tool for understanding brute-force vs. analytical attacks
Encrypts letters rather than bits (like all ciphers until after WW II)

Idea: replace each plaintext letter by a fixed other letter.

<table>
<thead>
<tr>
<th>Plaintext</th>
<th>Ciphertext</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>k</td>
</tr>
<tr>
<td>B</td>
<td>d</td>
</tr>
<tr>
<td>C</td>
<td>w</td>
</tr>
</tbody>
</table>

for instance, ABBA would be encrypted as kddk

Example (ciphertext):

iq ifcc vqqr fb rdq vfl lcq na rdq cfjwhwz hr bnnb
hcc hwwhbsqvqbre hwq vhlq

How secure is the Substitution Cipher?

Attacks against the Substitution Cipher

1. Attack: Exhaustive Key Search (Brute-Force Attack)

Simply try every possible substitution table until an intelligent plaintext appears (note that each substitution table is a key).

How many substitution tables (= keys) are there?

\[ 26 \times 25 \times \ldots \times 3 \times 2 \times 1 = 26! \approx 2^{88} \]

Search through \(2^{88}\) keys is completely infeasible with today's computers

Q: Can we now conclude that the substitution cipher is secure since a brute-force attack is not feasible?
A: No - We have to protect against all possible attacks.
2. Attack: Letter Frequency Analysis  
(Brute-Force Attack)

- Letters have very different frequencies in English
- Moreover: frequency of plaintext letters is preserved in the ciphertext.

<table>
<thead>
<tr>
<th>Letters</th>
<th>Frequency in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>13.0000</td>
</tr>
<tr>
<td>T</td>
<td>2.7000</td>
</tr>
<tr>
<td>A</td>
<td>2.0000</td>
</tr>
<tr>
<td>O</td>
<td>7.6000</td>
</tr>
<tr>
<td>N</td>
<td>6.7000</td>
</tr>
<tr>
<td>S</td>
<td>6.0000</td>
</tr>
<tr>
<td>H</td>
<td>6.0000</td>
</tr>
<tr>
<td>R</td>
<td>5.7000</td>
</tr>
<tr>
<td>D</td>
<td>4.3000</td>
</tr>
<tr>
<td>L</td>
<td>3.2000</td>
</tr>
<tr>
<td>C</td>
<td>2.8000</td>
</tr>
<tr>
<td>U</td>
<td>2.7000</td>
</tr>
<tr>
<td>M</td>
<td>2.1000</td>
</tr>
<tr>
<td>W</td>
<td>1.8000</td>
</tr>
<tr>
<td>F</td>
<td>1.7000</td>
</tr>
<tr>
<td>G</td>
<td>1.3000</td>
</tr>
<tr>
<td>Y</td>
<td>1.0000</td>
</tr>
<tr>
<td>P</td>
<td>0.9000</td>
</tr>
<tr>
<td>B</td>
<td>0.6000</td>
</tr>
<tr>
<td>V</td>
<td>0.4000</td>
</tr>
<tr>
<td>K</td>
<td>0.7000</td>
</tr>
<tr>
<td>J</td>
<td>0.2000</td>
</tr>
</tbody>
</table>

- "e" is the most common letter in English: (almost 13%)
- The next most common one is "t" with about 9%.

Breaking the Substitution Cipher with Letter Frequency Attack

- Let's return to our example and identify the most frequent letter:
  iq ifcc vqqr fb rdq vfllcq na rdq cfjwhwz hr
  bnnb hcc hwwbshvqbre hwq vhq
  • We replace the ciphertext letter q by E and obtain:
  iE ifcc vEEr fb rdE vfllcE na rdE cfjwhwz hr
  bnnb hcc hwwbshEvEbre hwE vhlE
  • By further guessing based on the frequency of the remaining letters we obtain the plaintext:
  WE WILL MEET IN THE MIDDLE OF THE LIBRARY AT
  NOON ALL ARRANGEMENTS ARE MADE
Breaking the Substitution Cipher with Letter Frequency Attack

- In practice, not only frequencies of individual letters can be used for an attack, but also the frequency of letter pairs (i.e., "th" is very common in English), letter triples, etc.
- cf. Problem 1.1 in *Understanding Cryptography* for a longer ciphertext you can try to break

**Important lesson:** Even though the substitution cipher has a sufficiently large key space of appr. $2^{88}$, it can easily be defeated with analytical methods. This is an excellent example that an encryption scheme must withstand all types of attacks.

Short Introduction to Modular Arithmetic

Why do we need to study modular arithmetic?
- Extremely important for asymmetric cryptography (RSA, elliptic curves etc.)
- Some historical ciphers can be elegantly described with modular arithmetic (cf. Caesar and affine cipher later on).

Most cryptosystems are based on **sets of numbers** that are
1. **discrete** (sets with integers are particularly useful)
2. **finite** (i.e., a finitely many numbers)

Even though the numbers are incremented every hour we never leave the set of integers
{1, 2, 3, ..., 11, 12}
Modular Arithmetic

- A system which allows to compute in finite sets of integers like the 12 integers we find on a clock (1, 2, 3, ..., 12).
- It is crucial to have an operation which "keeps the numbers within limits", i.e., after addition/multiplication they never leave the set (i.e., never larger than 12).

Definition: Modulus Operation
Let \( a, r, m \) be integers and \( m > 0 \). We write \( a \equiv r \mod m \)
if \((a-r)\) is divisible by \( m \).
- \("m"\) is called the modulus; \("r"\) is called the remainder

Examples for modular reduction.
- \( a=12 \) and \( m=9 \): \( 12 \equiv 3 \mod 9 \);
  \( a=34 \) and \( m=9 \): \( 34 \equiv 7 \mod 9 \)
- \( a=-7 \) and \( m=9 \): \( -7 \equiv 2 \mod 9 \)
(verify that the condition "\( m \) divides \((a-r)\)" holds in each case)

Properties of Modular Arithmetic (2)

- The remainder is not unique
  \( 12 \equiv 3 \mod 9 \);
  \( 12 \equiv 21 \mod 9 \);
  \( 12 \equiv -6 \mod 9 \)
By convention, we agree on the smallest positive integer \( r \) as remainder. This integer can be computed as
  \[ a = q \times m + r \quad \text{where} \quad 0 \leq r \leq m-1 \]
- Example: \( a=12 \) and \( m=9 \)
  \( 12 = 1 \times 9 + 3 \quad \rightarrow \quad r = 3 \)
Remark: This is just a convention. Algorithmically we are free to choose any other valid remainder to compute our crypto functions.

\[ \left( a+b \right) \mod m = \left( a \mod m + b \mod m \right) \mod m \]
\[ \left( a \times b \right) \mod m = \left( \left( a \mod m \right) \times \left( b \mod m \right) \right) \mod m \]
Properties of Modular Arithmetic (3)

- How do we perform modular division?
  
  Rather than divide, we prefer to multiply by the inverse. Ex:
  
  \[ \frac{b}{a} \equiv b \cdot a^{-1} \mod m \]

  The inverse \( a^{-1} \) of a number \( a \) is defined such that: \( a \cdot a^{-1} \equiv 1 \mod m \)

  Ex: What is \( \frac{5}{7} \mod 9 \)?
  
  The inverse of \( 7 \mod 9 \) is \( 4 \) since \( 7 \cdot 4 \equiv 28 \equiv 1 \mod 9 \), hence:
  
  \[ \frac{5}{7} \equiv 5 \cdot 4 = 20 \equiv 2 \mod 9 \]

- How is the inverse computed?
  
  The inverse of a number \( a \mod m \) only exists if and only if: \( \gcd(a, m) = 1 \)
  (above \( \gcd(5, 9) = 1 \), so that the inverse of 5 exists modulo 9)

  For now, the best way of computing the inverse is to use exhaustive search. The Euclidean Algorithm computes an inverse for a given number and modulus.

Properties of Modular Arithmetic (4)

- Modular reduction can be performed at any point during calculation
  
  Example: \( 3^8 \mod 7 \) (exponentiation is important in public-key cryptography).

  1. Approach: Exponentiation followed by modular reduction

  \[ 3^8 = 6561 \equiv 2 \mod 7 \]

  Generate intermediate result 6561 even though result can't be > 6.

  2. Approach: Exponentiation with intermediate modular reduction

  \[ 3^8 = 3^4 \cdot 3^4 = 81 \times 81 \]

  At this point we reduce the intermediate results 81 modulo 7:

  \[ 3^8 = 81 \times 81 \equiv 4 \times 4 \mod 7; \quad 4 \times 4 = 16 \equiv 2 \mod 7 \]

  Note that we can perform all these multiplications without a calculator, whereas mentally computing \( 3^8 = 6561 \) is challenging.

  General rule: For most algorithms it is advantageous to reduce intermediate results as soon as possible.
Modulo Arithmetic: The Ring $\mathbb{Z}_m$(1)

We can view modular arithmetic in terms of sets and operations in the set. By doing arithmetic modulo $m$ we obtain the integer ring $\mathbb{Z}_m$ with the following properties:

- **Closure**: Add and multiply any two numbers and the result is always in the ring.
- **Addition and multiplication are associative** - for all $a, b, c \in \mathbb{Z}_m$:
  \[ a + (b + c) = (a + b) + c; \quad a \times (b \times c) = (a \times b) \times c \]
  and **commutative**: $a + b = b + a; \quad a \times b = b \times a$

- The **distributive law** holds: $a \times (b + c) = (a \times b) + (a \times c)$

- There is the **neutral element** $0$ with respect to addition, i.e., for all $a \in \mathbb{Z}_m$: $a + 0 \equiv a \mod m$

- For all $a \in \mathbb{Z}_m$, there is always an **additive inverse element** $-a$ such that: $a + (-a) \equiv 0 \mod m$

- There is the **neutral element** $1$ with respect to multiplication, i.e., for all $a \in \mathbb{Z}_m$: $a \times 1 \equiv a \mod m$

- The **multiplicative inverse** $a^{-1}$ such that $a \times a^{-1} \equiv 1 \mod m$ exists only for some, but not for all, elements in $\mathbb{Z}_m$.

Modulo Arithmetic: The Ring $\mathbb{Z}_m$(2)

Roughly speaking, a ring is a structure in which we can always add, subtract and multiply, but we can only divide by certain elements (namely by those for which a multiplicative inverse exists).

- An element $a \in \mathbb{Z}_m$ has a multiplicative inverse only if $\gcd(a, m) = 1$
  We say that $a$ is **coprime** or **relatively prime** to $m$.

- Ex: Consider the ring $\mathbb{Z}_9 = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$
  The elements $0$, $3$, and $6$ do not have inverses since they are not coprime to $9$.

  The inverses of the other elements $1, 2, 4, 5, 7$, and $8$ are:
  
  $1^{-1} \equiv 1 \mod 9 \quad 2^{-1} \equiv 5 \mod 9 \quad 4^{-1} \equiv 7 \mod 9$
  $5^{-1} \equiv 2 \mod 9 \quad 7^{-1} \equiv 4 \mod 9 \quad 8^{-1} \equiv 8 \mod 9$
Shift (or Caesar) Cipher (1)

- Ancient cipher, allegedly used by Julius Caesar
- Replaces each plaintext letter by another one: simple rule: Take letter that follows after $k$ positions in the alphabet
  
  Needs mapping from letters $\rightarrow$ numbers:

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

- Example for $k = 7$; Plaintext = ATTACK = 0, 19, 19, 0, 2, 10
  Ciphertext = 7, 0, 0, 7, 9, 17 = haahjr

  Note that the letters “wrap around” at the end of the alphabet, which can be mathematically be expressed as reduction modulo 26, e.g.,
  $\quad 19 + 7 = 26 \equiv 0 \mod 26$

Shift (or Caesar) Cipher (2)

- Elegant mathematical description of the cipher.

  Let $k, x, y \in \{0, 1, \ldots, 25\}$
  - Encryption: $y = e_k(x) \equiv (x + k) \mod 26$
  - Decryption: $x = d_k(x) \equiv (y – k) \mod 26$

  - $k=3$ the “original” Caesar cipher; if $k=13 \Rightarrow$ encryption=decryption
  - Q: Is the shift cipher secure?
  - A: No! several attacks are possible, including:
    - Exhaustive key search (key space is only 26)
    - Letter frequency analysis, similar to attack against substitution cipher
Affine Cipher

- Extension of the shift cipher: rather than just adding a key to the plaintext, we also multiply
- We use for this a key consisting of two parts: $k = (a, b)$
  
  Let $k$, $x$, $y \in \{0, 1, \ldots, 25\}$
  - Encryption: $y = e_k(x) \equiv (a x + b) \mod 26$
  - Decryption: $x = d_k(y) \equiv a^{-1}(y - b) \mod 26$

- Since the inverse of $a$ is needed for inversion, we can only use values for $a$ for which: $\gcd(a, 26) = 1$
  
  There are 12 values for $a$ that satisfy this condition.
- The key space is only $12 \times 26 = 312$
- Again, several attacks are possible, including:
  - Exhaustive key search and letter frequency analysis, similar to the attack against the substitution cipher

Historic Ciphers

- Another variation of the substitution cipher - PigPen

- Poly-alphabetic cipher: Generate $L$ substitution tables - for the 1st letter use the 1st table, ..., for the $i$th letter use the $i$th table, up to $L$

- Encrypting "HELLO" results in "SHLJV"
- Key space is $(26!)^2$
**Vigenere Cipher**

- Used by the Confederate forces during the civil war 1860; Broken by Union cryptanalysts
- Use \( L \) cyclic shifts of the alphabet: Cipher key is \( L \) first letters of the shifts; Key space is \( 26^L \)
- Letter frequency analysis is less trivial
- Find first the value of \( L \) and then "break" each one separately
- Key observation: two identical segments of \( iL \) letters if the distance between the two is a multiple of \( L \)
  - Search for pairs of identical sequences of length \( \geq 3 \), record the distance between them - \( d_i \); \( L \) is the GCD of most of the \( d_i \)'s
- **Beale (book) Cipher**: Variant of Vigenere - use larger \( L \)
  - The "keyword" is the first few words in an agreed upon book
  - Simpler to memorize and has a larger key space

---

**Vernam Cipher (1919)**

- Telegraphic device using 5-bit ASCII characters
- Bitwise XOR of plaintext and key of the same length (provided on a paper tape): \( c = m \text{ xor } k \)
- If key consists of random bits - One-Time-Pad (OTP) cipher
  - OTP is a "perfectly secure cipher" (Shannon, 1945)
    - \( \text{Prob}(P=m|C=c) = \text{Prob}(P=m) \), where \( P= \text{plaintext}, C=\text{ciphertext} \)
  - Key management is difficult
    - Key must be random and used only once & still known to both sides
    - If same key used again an attacker can ask to encrypt a known message \( m \) and then \( k = m \text{ xor } c \); or
    - Some information is leaked - if \( c_1 = m_1 \text{ xor } k \) and \( c_2 = m_2 \text{ xor } k \), then \( c_1 \text{ xor } c_2 = m_1 \text{ xor } m_2 \)
  - Used in practice only when secrecy is paramount
Enigma Machine – used by Germany in WWII

- Invented in Germany in 1920 and then improved and used by the military in WWII
- Five rotors and the plugboard provided $1.6 \times 10^{20}$ settings
- Poly-alphabetic substitution cipher
- Initial setting of rotors and plugboard changed everyday
  - Product of permutations $PR, R, FR, R, R^2, R^3, R^{-1}, \ldots$
- The Enigma code was broken by a team headed by Alan Turing

Lessons Learned

- Never develop your own crypto algorithm unless you have a team of experienced cryptanalysts checking your design.
- Do not use unproven crypto algorithms or unproven protocols.
- Attackers always look for the weakest point of a cryptosystem, a large key space by itself is no guarantee for a cipher being secure; the cipher might still be vulnerable against analytical attacks.
- Key length for symmetric algorithms in order to thwart exhaustive key-search attacks:
  - 64 bit: insecure except for data with extremely short-term value
  - 128 bit: long-term security of several decades, unless quantum computers become available (quantum computers do not exist and perhaps never will)
  - 256 bit: as above, but probably secure against attacks by quantum computers.
- Modular arithmetic is a tool for expressing historical encryption schemes, such as the affine cipher, in a mathematically elegant way.