Course Overview

Course objectives:

• Learn how crypto primitives work
• Learn how to use them correctly and reason about security
Some of The Communication Security You Enjoy Daily

- Paying by credit card at a store
- Cell phone usage (calls, SMS, etc.)
- Email, chat, online calls (skype, google voice, etc)
- Secure browsing, shopping online
- Cloud storage and communication between your devices
- Software updates (computer, phones, tablets, etc.)
- Online Banking (huge deal!!!!)
- Car keys
- Electronic doors
- CIA, NSA, etc
- Etc, etc
Information Security

Integrity

data not tampered with or deleted
`no insertions of false data or replays of old data

Availability

data not destroyed or rendered unusable
Privacy or confidentiality

- keeping information secret from all but those who are authorized to see it.

Entity Authentication or Identification

-- corroboration of the identity of an entity
-- data and sender are as claimed

Non-Repudiation

- sender cannot disavow

(e.g., a person, a computer terminal, a credit card, etc.).
• **Message Authentication**
  - *corroborating the source of information; also known as data origin authentication.*

• **Signature**
  - *a means to bind information to an entity.*

• **Authorization**
  - *conveyance, to another entity, of official sanction to do or be something.*

• **Validation**
  - *a means to provide timeliness of authorization to use or manipulate information or resources.*
• Access Control
  - restricting access to resources to privileged entities.

• Certification
  - endorsement of information by a trusted entity.

• Timestamping
  - recording the time of creation or existence of information.

• Witnessing
  - verifying the creation or existence of information by an entity other than the creator.
• Receipt
  - acknowledgement that information has been received.

• Confirmation
  - acknowledgement that services have been provided.

• Ownership
  - a means to provide an entity with the legal right to use or transfer a resource to others.
**Definition**: Cryptography is the study of mathematical techniques related to aspects of information security such as confidentiality, data integrity, entity authentication, and data origin authentication.

Cryptography was once “the art of writing or solving codes.”

It has now become “the study of techniques for securing digital information, transactions, and distributed Computation.”
Secure communication

HTTPS

no eavesdropping
no tampering
Secure Sockets Layer / TLS

Two main parts:

1. Handshake Protocol: Establish shared secret key using public-key cryptography

2. Record Layer: Transmit data using shared secret key
   Ensure confidentiality and integrity
Protecting Files on Disk

1) Make sure no one else can access it (Confidentiality)
2) No one else can alter it (Integrity)
Building block

If the keys are the same: Symmetric encryption
Otherwise: Asymmetric encryption

E, D: cipher   k: secret key (e.g. 128 bits)
m, c: plaintext, ciphertext

Encryption algorithm is **publicly known**
  • Never use a proprietary cipher
Use Cases

Single use key: (one time key)
  • Key is only used to encrypt one message
  • encrypted email: new key generated for every email

Multi use key: (many time key)
  • Key used to encrypt multiple messages
    • encrypted files: same key used to encrypt many files
      Needs more work than for one-time key
Things to remember

Cryptography is:

• A powerful tool
• The basis for many security mechanisms

Cryptography is not:

• The solution to all security problems
• Reliable unless implemented and used properly
• Something you should try to invent yourself
  - many many examples of broken ad-hoc designs
Crypto core

Secret key establishment:

Secure communication:

confidentiality and integrity
But crypto can do much more

- Digital signatures

- Anonymous communication

- Anonymous digital cash
  - Can I spend a “digital coin” without anyone knowing who I am?
  - How to prevent double spending?
Protocols

- Elections
- Private auctions
Crypto magic

- Privately outsourcing computation
  
  ![Diagram of search query and results](image)

- Zero knowledge (proof of knowledge)
  
  ![Diagram of N = p \cdot q and proof π](image)
Some Number Theory!

**Definition:** if $a$, $b$ and $m$ are integers, we say that $a$ is congruent to $b$ modulo $m$ if $m \mid (a-b)$

$a$ is congruent to $b$ modulo $m$, we write

$$a \equiv b \pmod{m}.$$ 

**EX:**

$22 \equiv 4 \pmod{3}$ means that $3$ divides $(22-4)$ evenly!

You say: $22$ is congruent $4$ modulo $3$!!!
**Congruent Mod n**

**Proposition:** $a \equiv b \mod n$ if and only if $a - b = kn$

Where $k$ is a integer and $n$ is a non zero integer that is, $n$ evenly divides $(a - b)$ written $n \mid (a - b)$

**Examples:**

- $30 \equiv 9 \mod 7$
- $9 \equiv 16 \mod 7$
- $-2 \equiv 12 \mod 7$
Theorem: If \( a, b, c \) and \( n \) integers with \( n > 0 \) such that \( a \equiv b \mod n \), then

(i) \( a + c \equiv b + c \mod n \)
(ii) \( a - c \equiv b - c \mod n \)
(iii) \( a*c \equiv b*c \mod n \)

Proof: \( a \equiv b \mod n \iff a - b = k n \ (k \text{ integer}) \)

\[
\begin{align*}
a + c - (b + c) &= k n \\
etc......
\end{align*}
\]
Theorem: If \( a, b, c, d \) and \( n \) integers with \( n > 0 \) such that \( a \equiv b \mod n \) and \( c \equiv d \mod n \) then

(i) \( a + c \equiv b + d \ (mod \ n) \)
(ii) \( a - c \equiv b - d \ (mod \ n) \)
(iii) \( a*c \equiv b*d \ (mod \ n) \)

Proof: Home Work!
Theorem: If $a, b, k$ and $n$ integers with $n, k > 0$ such that $a \equiv b \mod n$ then $a^k \equiv b^k \mod n$.

Proof:
mod

10 Mod 7 = 3 (the remainder)

For any integer n (positive and negative!!!)

n mod 7 will be one of 0,1,2,3,4,5,6
Integers Mod n

Given a modulus $n$, numbers are in the range $[0, n-1]$. Called the complete set of residues.

  e.g., for mod 7, have 0,1,2,3,4,5,6

If the number is not in the range $[0, n-1]$ we continue to “reduced” mod $n$

Note if $x \mod n = y$ where $y$ in $[0, n-1]$ iff

  $x \equiv y \mod n$
Reduction Mod n

Divide by n and take remainder; if negative, add n
(not same as in programming languages)

Examples

7 mod 7 = 0
17 mod 7 = 3
5 mod 7 = 5
-2 mod 7 = 5
-17 mod 7 = -3 mod 7 = 4
Arithmetic Mod $n$

Give $(n > 0)$ and any integers $a, b$ and $\text{op} = +, -, \text{or} \,*$

Then

$$(a \mod n \; \text{op} \; b \mod n) \mod n = (a \; \text{op} \; b) \mod n$$

(we assume mod has higher priority)
Examples

\[(13 + 12) \mod 7 = 4 = (13 \mod 7 + 12 \mod 7) \mod 7\]

\[9 \times 8 \mod 7 = 2 = (9 \mod 7 \times 8 \mod 7) \mod 7\]

\[(15 - 26) \mod 7 = 3 = (15 \mod 7 - 26 \mod 7) \mod 7\]
Properties

Integers mod n with addition (& subtraction) and multiplication form a commutative ring

In addition we have

\[(a*(b+c)) \mod n = (((a*b) \mod n + ((a*c) \mod n)) \mod n)\]

i.e. you can pretend mod to be like *
Principle of Modular Arithmetic

Evaluating a formula in modular arithmetic (i.e., reducing the result after each step) gives the same result as evaluating it in ordinary arithmetic and reducing the result mod \( n \)

\[
(4 \times 6 + 5) \mod 7 = (4 \times 6) \mod 7 + 5 \mod 7 \\
= (3 + 5) \mod 7 = 1 \\
(4 \times 6 + 5) \mod 7 = 29 \mod 7 = 1
\]
**Classical Cryptography**

**Definition:** A *cryptosystem* is a five-tuple \((\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})\), where the following are satisfied:

1. \(\mathcal{P}\) is a finite set of possible *plaintexts*.
2. \(\mathcal{C}\) is a finite set of possible *ciphertexts*.
3. \(\mathcal{K}\) The *keyspace*, is a finite set of possible keys.
4. For each \(K \in \mathcal{K}\) there is a encryption rule \(E_K \in \mathcal{E}\) a corresponding \(D_K \in \mathcal{D}\). Each \(E_k : \mathcal{P} \to \mathcal{C}\) and \(D_k : \mathcal{C} \to \mathcal{P}\) are functions such that \(D_k(\ E_k(x)\ ) = x\) for every plaintext \(x \in \mathcal{P}\).
Kerckhos' principle.

The cipher method must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience.

Security depends solely on the secrecy of the key, and not on Encryption.

Easier to keep a short key secret than a long program. If key is exposed, it is easier to replace it than Encryption algorithm.

With many users, it is easier for all pairs to use just one Encryption, and have different keys.
Today Kerckhos' principle is muddied to require that the crypto algorithms be made public.

Public scrutiny is more likely find weaknesses.

It is better for flaws to be found by “ethical hackers" than the enemy.

If the algorithm is supposed to be secret, then reverse engineering can compromise it. Usually, the key is not subject to reverse engineering.

Public design allows for standards.
Shift Cipher

Figure 31 A U.S. Confederate cipher disk used in the American Civil War.
Shift Cipher

\( P = C = K = E = \mathbb{Z}_{26} \) For \( 0 \leq K \leq 25 \), define

\[
E_k(x) = (x + K) \mod 26 \quad \text{and} \quad D_k(y) = (y - K) \mod 26 \quad (x, y \in \mathbb{Z}_{26})
\]

Remark: \( K = 3 \) the cryptosystem is often called Caesar Cipher.

**EX:**

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td>J</td>
<td>K</td>
<td>L</td>
<td>M</td>
<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>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td>J</td>
<td>K</td>
<td>L</td>
<td>M</td>
<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>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
</tbody>
</table>

\( P = \text{COVID, GO AWAY!!!} \rightarrow \text{COVIDGOAWAY} \)

\( C = \text{FRYLG JR DZDB} \rightarrow \text{FRYLGJRDZDB} \)
• Generalize shift cipher to any group:

DEF: A **commutative group** is a tuple \((G,+)\) which satisfies

1. Closed: \( \forall x, y \in G, \text{ then } x+y \in G \)

2. Associative: \( \forall x, y, z \in G \quad (x + y) + z = x + (y + z) \)

3. Identity/zero: \( 0 \in G, \forall x \in G, \quad x + 0 = 0 + x = x \)

4. Inverses: \( \forall x \in G \exists y \in G \text{ such that } x + y = y + x = 0 \)

5. **Commutative** (or **abelian**):

\[ \forall x, y \in G \quad x + y = y + x \]
Substitution Cipher

\[ \mathcal{P} = \mathcal{C} = \mathcal{K} = \mathcal{E} = \mathbb{Z}_{26} \]. \( \mathcal{K} \) consists all possible Permutations of the 26 symbols 0,1,..,25. For each Permutation. \( \pi \in \mathcal{K} \), define

\[ E_k(x) = \pi(x) \]

and define

\[ D_k(y) = \pi^{-1}(y) \]

Where \( \pi^{-1} \) is the inverse permutation of \( \pi \)
Example:

Suppose $m = 6$ and the key is the following permutation $\pi$:

\[
\begin{array}{c|cccccc}
  x & 1 & 2 & 3 & 4 & 5 & 6 \\
  \pi(x) & 3 & 5 & 1 & 6 & 4 & 2 \\
\end{array}
\]

\[
\begin{array}{c|cccccc}
  x & 1 & 2 & 3 & 4 & 5 & 6 \\
  \pi^{-1}(x) & 3 & 6 & 1 & 5 & 2 & 4 \\
\end{array}
\]

Plain Text: shesellsseashellsbytheseashore.
plaintext into groups of six letters: shesel lsseas hellsb ythese ashore
Cipher Text: EESLSH SALSES LSHBLE HSYEET HRAEOS
Observation: Key space for substitution cipher is $26!$

Observation: Key space for shift cipher is $25$.

Observation: Shift is a special case of substitution.

$26! = 4.0 \times 10^{26}$ is very large even for the computer.

Searching the key space is stupid!!

We will see later how to break substitution cipher easily.
Cryptanalysis

ciphertext-only attack: The opponent possesses a string of ciphertext, y.
    (worst kind of attack)

known plaintext attack: The opponent possesses a string of plaintext, x,
    and the corresponding ciphertext, y.

chosen plaintext attack: The opponent has obtained temporary access to
    the encryption machinery. Hence he can choose a plaintext string, x,
    and construct the corresponding ciphertext string, y.

chosen ciphertext attack: The opponent has obtained temporary access to
    the decryption machinery. Hence he can choose a ciphertext string, y,
    and construct the corresponding plaintext string, x
TABLE 2.1: Probabilities of occurrence of the 26 letters

<table>
<thead>
<tr>
<th>letter</th>
<th>probability</th>
<th>letter</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.082</td>
<td>N</td>
<td>.067</td>
</tr>
<tr>
<td>B</td>
<td>.015</td>
<td>O</td>
<td>.075</td>
</tr>
<tr>
<td>C</td>
<td>.028</td>
<td>P</td>
<td>.019</td>
</tr>
<tr>
<td>D</td>
<td>.043</td>
<td>Q</td>
<td>.001</td>
</tr>
<tr>
<td>E</td>
<td>.127</td>
<td>R</td>
<td>.060</td>
</tr>
<tr>
<td>F</td>
<td>.022</td>
<td>S</td>
<td>.063</td>
</tr>
<tr>
<td>G</td>
<td>.020</td>
<td>T</td>
<td>.091</td>
</tr>
<tr>
<td>H</td>
<td>.061</td>
<td>U</td>
<td>.028</td>
</tr>
<tr>
<td>I</td>
<td>.070</td>
<td>V</td>
<td>.010</td>
</tr>
<tr>
<td>J</td>
<td>.002</td>
<td>W</td>
<td>.023</td>
</tr>
<tr>
<td>K</td>
<td>.008</td>
<td>X</td>
<td>.001</td>
</tr>
<tr>
<td>L</td>
<td>.040</td>
<td>Y</td>
<td>.020</td>
</tr>
<tr>
<td>M</td>
<td>.024</td>
<td>Z</td>
<td>.001</td>
</tr>
</tbody>
</table>

1. E, having probability about 0.120
2. T, A, O, I, N, S, H, R, each having probability between 0.06 and 0.09
3. D, L, each having probability around 0.04
4. C, U, M, W, F, G, Y, P, B, each having probability between 0.015 and 0.028
5. V, K, J, X, Q, Z, each having probability less than 0.01.

It is also useful to consider sequences of two or three consecutive letters, called digrams and trigrams, respectively. The 30 most common digrams are (in decreasing order):

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.

The twelve most common trigrams are:

THE, ING, AND, HER, ERE, ENT,
THA, NTH, WAS, ETH, FOR, DTH.
Cryptanalysis of the Substitution Cipher

Example:

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ
NDIFEFCMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ
NZUCDRJXYYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZIJ
XZWGCHSMRNMHDHCMMFQCHZJXJZWEJYUCFWJNZDIR

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

\[ D_K(Z) = e \]
Next High frequency: C, D, F, J, M, R, Y. \( \Rightarrow \) \{t, a, o, i, n, s, h, r\}
But difficult to decide.

Look for digrams \(-Z\) or \(Z-\)

- DZ and ZW (four times each);
- NZ and ZU (three times each);
- RZ, HZ, XZ, FZ, ZR, ZV, ZC, ZD, Z J (twice each).

Observe: ZW occurs four times and WZ not at all, and W occurs less often than many other characters, we might guess that \(d_K(W) = d\).

ZRW occurring near the beginning of the ciphertext,

RW occurs again later on.

R occurs frequently in the ciphertext and \(nd\) is a common digram, we might try \(d_K(R) = n\) as the most likely possibility
Our next step might be to try $d_K(N) = h$, since NZ is a common digram andZN is not ne−ndhe

$\rightarrow$  $d_K(C) = a$
Now, we might consider $M$, the second most common ciphertext character. The ciphertext segment $RNM$, which we believe decrypts to $nh-$, suggests that $h-$ begins a word, so $M$ probably represents a vowel. We have already accounted for $a$ and $e$, so we expect that $d_K(M) = i$ or $o$. Since $ai$ is a much more likely digram than $ao$, the ciphertext digram $CM$ suggests that we try $d_K(M) = i$ first. Then we
Next, we might try to determine which letter is the encryption of o. Since o is a common plaintext character, we guess that the corresponding ciphertext character is one of D, F, J, Y. Y seems to be the most likely possibility; otherwise, we would get long strings of vowels, namely aoi from CFM or CJM. Hence, let's suppose $d_K(Y) = o$.

The three most frequent remaining ciphertext letters are D, F, J, which we conjecture could decrypt to r, s, t in some order. Two occurrences of the trigram NMD suggest that $d_K(D) = s$, giving the trigram his in the plaintext (this is consistent with our earlier hypothesis that $d_K(D) \in \{r, s, t\}$). The segment HNCMF could be an encryption of chair, which would give $d_K(F) = r$ (and $d_K(H) = c$) and so we would then have $d_K(J) = t$ by process of elimination. Now, we have:

```
 o-r-riend-ro--arise-a-inedhise--t---ass-it
 YIFQFMZRWQFYVECFMDZPCVMRZWMDZVEJBTXCDDUMJ

 hs-r-riseasi-e-a-orationhadta-en--ace-hi-e
 NDIFEFMDZCDMZXKCEYFCJMYRNCWJCSZREXCHZUNMXZ

 he-asnt-oo-in-i-o-redso-e-ore-ineandhesett
 NZUCDRJXYSMRTMEYIFZWDYZVYZVFZUMRZCRWNZDZJJ

 -ed-ac-inhischair-aceti-ted--to-ardsthes-n
 XZWGCHSMRNMDDHNCFMQCHZJMXJZWIEYUCFWDJNZDIR
```
Our friend from Paris examined his empty glass with surprise, as if evaporation had taken place while he wasn’t looking. I poured some more wine and he settled back in his chair, face tilted up towards the sun.
The Affine Cipher

This is another special case of substitution cipher:

The encryption function:

$$E(x) = (ax + b) \mod 26 \quad \text{where } a, b \in \mathbb{Z}_{26}$$

Now to compute the decryption function we need to solve for $x$:

$$ax + b \equiv y \pmod{26}$$

equivalent to

$$ax \equiv y \pmod{26}$$
Theorem: The congruence $a x \equiv b \pmod{m}$ has a unique solution $x \in \mathbb{Z}_m$ for $\forall y \in \mathbb{Z}_m$ if and only if $\gcd(a, m) = 1$.

Proof: $\Rightarrow$ Suppose unique solution and $\gcd(a, m) = d$

$ax - b = my$ for some $y$; let $x_0, y_0$ be some particular solution

now $d \mid a$ and $d \mid m$; consider $x = x_0 + \frac{m}{d} t$ and $y = y_0 + \frac{a}{d} t$

note $t$ is a parameter, you get different value of $x$ different value of $t$

Now $a x - my = a (x_0 + \frac{m}{d} t) - m (y_0 + \frac{a}{d} t) = a x_0 + a (\frac{m}{d} t) - m y_0 + m (\frac{a}{d} t)$

$= a x_0 - m y_0$

$\Leftarrow$ assume $\gcd(a, m) = 1$ and $\exists x_1 \neq x_2$

$sa + tm = 1$; $sab + btm = b$ hence $sb$ is a solution to $a x \equiv b \pmod{m}$

now suppose $a x_1 \equiv b \pmod{m}$ and $ax_2 \equiv b \pmod{m}$

$a x_1 \equiv ax_2 \pmod{m}$; $\gcd(a,m) = 1$

$x_1 \equiv x_2 \pmod{m}$;
Example 2.10 Ciphertext obtained from an Affine Cipher

FMXVEDKAPHFERBNDKRXRSREFMORUDSDKDVSHPVFEDK
APRKDLYEVLRHHRH
$e_K(4) = 17$ and $e_K(19) = 3$.

$4a + b = 17$
$19a + b = 3$.

$a = 6, b = 19; \quad \gcd(6,26) = 2$ no inverse

Next: $e_K(4) = 17$ and $e_K(19) = 4$.

$4a + b = 17$
$19a + b = 4; \quad a = 13; \quad \gcd(a,26) = 13$ no inverse

Next: $e_K(4) = 17$ and $e_K(19) = 10$

$4a + b = 17$
$19a + b = 10$

$a = 3; b = 5 \quad \text{Good}$

Algorithms are quite general definitions of arithmetic processes
Why Modular Arithmetic?

Numbers are bounded in size, so encryption doesn’t cause expansion - allows use of huge numbers, hundreds of digits long

Can compute multiplicative inverses, which is needed for decryption in ciphers that use multiplication (or exponentiation)

Some operations are prohibitively time consuming for huge numbers, making cryptanalysis infeasible
Multiplicative Inverses

Given $a$, $x$ is an inverse of $a$ mod $n$ if

$$a \times x \mod n = 1$$

note that $a$ is also an inverse of $x$

also write $x = 1/a \mod n$ and $x = a^{-1} \mod n$

examples of inverses

$$2 \times 4 \mod 7 = 1$$
$$2 \times 6 \mod 11 = 1$$
Inverse Exists if Relatively Prime

a has an inverse mod n; i.e. there is a solution x to
a \times x \mod n = 1 \text{ if and only if } \gcd(a, n) = 1
(i.e. a and n are relatively prime)
Proof Sketch

Let $a$ be a number where $0 \leq a < n$; and $\gcd(a,n) = 1$

consider $a^0, a^1, ..., a^i, ..., a^j, ..., a^{n-1} \mod n$

assume for some $i$ and $j$: $a^i \mod n = a^j \mod n$

then $n \mid (a^j - a^i)$ - that is, $n \mid (a \cdot (j - i))$

$n$ can’t divide $a$ because $\gcd(a, n) = 1$

so $n \mid (j - i)$

but that’s impossible since $i, j < n$

so each $a^i \mod n$ must be unique

so $a^i \mod n = 1$ for some $i$
**Euler Totient Function**

\( \phi(n) = \) number of elements (residues) in 0, ..., n-1 that are relatively prime to n

\( \phi(p) = p - 1 \) for prime p

\( \phi(p^*q) = \phi(p) \phi(q) = (p - 1)(q-1) \) for primes p, q

Fermat’s theorem: \( a^{p-1} \mod p = 1 \) (gcd(a,p) = 1)

Euler’s generalization: \( a^{\phi(n)} \mod n = 1 \)

implies \( a \ast a^{\phi(n)-1} \mod n = 1 \)
**Proof Sketch**

Show That: $a^{\phi(n)} \mod n = 1$ where gcd$(a, n) = 1$

Proof: Let $r_i$ be a residue, [defn: gcd$(ri, n ) = 1$] then

$$a*r_i \mod n = r_j$$ for some residue $r_j$

(this is true since gcd$(a,n) = 1$ & gcd$(ri, n ) = 1$)

$$a*r_1 \mod n, ..., a*r_{\phi(n)} \mod n$$

is a permutation of the residues $r_1, ..., r_{\phi(n)}$

implies

$$(a*r_1 \mod n * ... * a*r_{\phi(n)} \mod n) = (r_1 * ... * r_{\phi(n)})$$

$$(a^{\phi(n)} \mod n)(r_1 * ... * r_{\phi(n)}) = (r_1 * ... * r_{\phi(n)})$$

Hence result.
Example

1) \(a = 5\)  \(n = 7\) find \(x = 1/a;\)

   since \(n\) is prime \(\phi(n) = 6\)

   so the inverse is

   \[x = 1/a = a^{\phi(n)-1} \mod n = 5^{6-1} \mod 7 = 3\]

Check: \(5 \times 3 \mod 7 = 1\)

Given \(a = 4\) and \(n = 9\), find \(x = 1/a\)

\(\phi(n) = 6\)

\[x = 1/a = a^{\phi(n)-1} \mod n = 4^{5} \mod 9 = 7\]

check: \(4 \times 7 \mod 9 = 1\)
RING

- Often, addition too easily cracked
- Extra structure obfuscates: multiplication

DEF: A commutative ring is a 3-tuple \((R, +, \cdot)\) which satisfies:

1. \((R, +)\) is a commutative group
2. \(R\) is closed under “\(\cdot\)” -which is associative,
   commutative and has identity \(1 \in R\)
3. Distributive:
   \[\forall x, y, z \in R, \ x(y+z) = xy+xz \quad \text{and} \quad (x+y)z = xz+yz\]
Vigènere Cipher

Like Cesar cipher, but use a phrase as key

Example

Message THE BOY HAS THE BALL
Key VIG

Encipher using Cæsar cipher for each letter:

key VIGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIG
plain THEBOYHASTHEBALL
cipher OPKWWECIYOPKWKIRG
Vigènere Cipher

Let $m$ be a positive integer. Define

$$\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathcal{E} = \left( \mathbb{Z}_{26} \right)^m$$

For key $K = (k_1, k_2, \ldots, k_m)$, we define:

$$E_k (x_1, x_2, \ldots x_m) = (x_1 + k_1, x_2 + k_2, \ldots x_m + k_m)$$

And

$$D_k (y_1, y_2, \ldots y_m) = (y_1 - k_1, y_2 - k_2, \ldots y_m - k_m)$$

Where all operations are performed in $\mathbb{Z}_{26}$
Cryptanalysis of the Vigen`ere Cipher

Tools for Breaking

• **Index of Coincidence**
  Kasiski method

• **Known frequency distributions for single letters, digrams, and trigrams**
Index of Coincidence

• Measures the variation in the frequencies of letters in the ciphertext

• If period \( d = 1 \), then considerable variation since

• ciphertext letters corresponding to E, T, etc. will
  • occur more than those corresponding to J, X, etc.

• As \( d \) increases, variation becomes less

• Introduced in 1920s by William Friedman
Formula for IC

\[ IC = \sum_{i=0}^{n-1} \frac{(F_i \times (F_i - 1))}{(N \times (N - 1))} \]

\( F_i \) = frequency (# occurrences) of \( i \)th letter

\( N \) = length of ciphertext (# letters)

IC is the probability that two letters, chosen at random, are alike

For English, usually ranges from about \( .066 \) (for \( d = 1 \)) down to \( .038 \) for very large \( d \)

Statistical in nature, so does not give \( d \) exactly
# Expected IC for English

<table>
<thead>
<tr>
<th>d</th>
<th>IC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.066</td>
</tr>
<tr>
<td>2</td>
<td>0.052</td>
</tr>
<tr>
<td>3</td>
<td>0.047</td>
</tr>
<tr>
<td>4</td>
<td>0.045</td>
</tr>
<tr>
<td>5</td>
<td>0.044</td>
</tr>
<tr>
<td>10</td>
<td>0.041</td>
</tr>
<tr>
<td>large</td>
<td>0.038</td>
</tr>
</tbody>
</table>
Extreme Cases

\[ C = \sum_{i=0}^{n-1} \frac{(F_i \times (F_i - 1))}{(N \times (N - 1))} \]

approx: \[ \sum F_i^2 / N^2 \]

flat distribution (equally likely letters)

\[ F_i = N/n; \text{ hence the approximation} \]
\[ = n \times (N/n)^2 / N^2 = 1/n = 1/26 = 0.038 \]

totally skewed - all one letter

\[ F_i = N \text{ for one letter, } F_i = 0 \text{ for all others} \]

\[ IC = N^2 / N^2 = 1 \]
Use of IC

1. Compute IC across entire ciphertext. If high (near 0.066), then $d = 1$ is likely.

2. Estimate $d$ by other means. Compute IC for each of the $d$ sequences

   \[(i = 0, ..., d-1) \text{ of length } N/d \text{ (approx)}\]

   $c_i, c_{i+d}, c_{i+2d}, ...$

If individual IC values are high, then $d$ is probably accurate.
Kasiski Method

Introduced in 1863 by the Prussian military officer Friedrich W. Kasiski

Analyze repetitions in ciphertext to determine the period. Long repetitions in ciphertext suggest repetitions in key (and plaintext)
FIGURE 2.8 Sample ciphertext.

\[ d_1 = 51 \quad d_2 = 72 \quad \text{gcd} (51, 72) = 3 \]

- candidate period
FIGURE 2.9 Frequency distribution of ciphertext in Figure 2.8.

<table>
<thead>
<tr>
<th>Char</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4.0</td>
</tr>
<tr>
<td>B</td>
<td>0.9</td>
</tr>
<tr>
<td>C</td>
<td>6.1</td>
</tr>
<tr>
<td>D</td>
<td>2.0</td>
</tr>
<tr>
<td>E</td>
<td>4.9</td>
</tr>
<tr>
<td>F</td>
<td>3.5</td>
</tr>
<tr>
<td>G</td>
<td>4.0</td>
</tr>
<tr>
<td>H</td>
<td>3.2</td>
</tr>
<tr>
<td>I</td>
<td>3.5</td>
</tr>
<tr>
<td>J</td>
<td>4.6</td>
</tr>
<tr>
<td>K</td>
<td>5.2</td>
</tr>
<tr>
<td>L</td>
<td>5.8</td>
</tr>
<tr>
<td>M</td>
<td>3.2</td>
</tr>
<tr>
<td>N</td>
<td>4.6</td>
</tr>
<tr>
<td>O</td>
<td>4.0</td>
</tr>
<tr>
<td>P</td>
<td>2.0</td>
</tr>
<tr>
<td>Q</td>
<td>3.8</td>
</tr>
<tr>
<td>R</td>
<td>8.7</td>
</tr>
<tr>
<td>S</td>
<td>4.3</td>
</tr>
<tr>
<td>T</td>
<td>2.0</td>
</tr>
<tr>
<td>U</td>
<td>3.5</td>
</tr>
<tr>
<td>V</td>
<td>4.0</td>
</tr>
<tr>
<td>W</td>
<td>1.7</td>
</tr>
<tr>
<td>X</td>
<td>0.6</td>
</tr>
<tr>
<td>Y</td>
<td>6.1</td>
</tr>
<tr>
<td>Z</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Number of Characters = 346
Index of Coincidence = .0434
**FIGURE 2.10** Frequency distributions for separate sequences.

<table>
<thead>
<tr>
<th>Char</th>
<th>Percent</th>
<th>Char</th>
<th>Percent</th>
<th>Char</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.0</td>
<td>A</td>
<td>9.6</td>
<td>A</td>
<td>2.6</td>
</tr>
<tr>
<td>B</td>
<td>0.0</td>
<td>B</td>
<td>0.9</td>
<td>B</td>
<td>1.7</td>
</tr>
<tr>
<td>C</td>
<td>6.0</td>
<td>C</td>
<td>1.7</td>
<td>C</td>
<td>10.4</td>
</tr>
<tr>
<td>D</td>
<td>1.7</td>
<td>D</td>
<td>2.6</td>
<td>D</td>
<td>1.7</td>
</tr>
<tr>
<td>E</td>
<td>4.3</td>
<td>E</td>
<td>10.4</td>
<td>E</td>
<td>0.0</td>
</tr>
<tr>
<td>F</td>
<td>5.2</td>
<td>F</td>
<td>2.6</td>
<td>F</td>
<td>2.6</td>
</tr>
<tr>
<td>G</td>
<td>2.6</td>
<td>G</td>
<td>2.6</td>
<td>G</td>
<td>7.0</td>
</tr>
<tr>
<td>H</td>
<td>0.9</td>
<td>H</td>
<td>8.7</td>
<td>H</td>
<td>0.0</td>
</tr>
<tr>
<td>I</td>
<td>5.2</td>
<td>I</td>
<td>5.2</td>
<td>I</td>
<td>0.0</td>
</tr>
<tr>
<td>J</td>
<td>8.6</td>
<td>J</td>
<td>0.0</td>
<td>J</td>
<td>5.2</td>
</tr>
<tr>
<td>K</td>
<td>13.8</td>
<td>K</td>
<td>0.9</td>
<td>K</td>
<td>0.9</td>
</tr>
<tr>
<td>L</td>
<td>1.7</td>
<td>L</td>
<td>1.7</td>
<td>L</td>
<td>13.9</td>
</tr>
<tr>
<td>M</td>
<td>0.9</td>
<td>M</td>
<td>1.7</td>
<td>M</td>
<td>7.0</td>
</tr>
<tr>
<td>N</td>
<td>0.9</td>
<td>N</td>
<td>9.6</td>
<td>N</td>
<td>3.5</td>
</tr>
<tr>
<td>O</td>
<td>0.0</td>
<td>O</td>
<td>12.2</td>
<td>O</td>
<td>0.0</td>
</tr>
<tr>
<td>P</td>
<td>2.6</td>
<td>P</td>
<td>1.7</td>
<td>P</td>
<td>1.7</td>
</tr>
<tr>
<td>Q</td>
<td>1.7</td>
<td>Q</td>
<td>0.0</td>
<td>Q</td>
<td>9.6</td>
</tr>
<tr>
<td>R</td>
<td>9.5</td>
<td>R</td>
<td>4.3</td>
<td>R</td>
<td>12.2</td>
</tr>
<tr>
<td>S</td>
<td>0.0</td>
<td>S</td>
<td>8.7</td>
<td>S</td>
<td>4.3</td>
</tr>
<tr>
<td>T</td>
<td>0.9</td>
<td>T</td>
<td>5.2</td>
<td>T</td>
<td>0.0</td>
</tr>
<tr>
<td>U</td>
<td>5.2</td>
<td>U</td>
<td>4.3</td>
<td>U</td>
<td>0.9</td>
</tr>
<tr>
<td>V</td>
<td>11.2</td>
<td>V</td>
<td>0.9</td>
<td>V</td>
<td>0.0</td>
</tr>
<tr>
<td>W</td>
<td>1.7</td>
<td>W</td>
<td>0.9</td>
<td>W</td>
<td>2.6</td>
</tr>
<tr>
<td>X</td>
<td>1.7</td>
<td>X</td>
<td>0.0</td>
<td>X</td>
<td>0.0</td>
</tr>
<tr>
<td>Y</td>
<td>4.3</td>
<td>Y</td>
<td>3.5</td>
<td>Y</td>
<td>10.4</td>
</tr>
<tr>
<td>Z</td>
<td>9.5</td>
<td>Z</td>
<td>0.0</td>
<td>Z</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Number of Characters in Sequence = 116  
Number of Characters in Sequence = 115  
Number of Characters in Sequence = 115  

Index of Coincidence = .06747  
Index of Coincidence = .06499  
Index of Coincidence = .07597
I have been told by learned sources that when a truly great musician plays and his playing sounds so free and spontaneous that the listener has no idea of the amount of nonsensuous work, study, scholarship, analysis, and planning required to achieve these spontaneous effects. I shall not argue the point, only I wish to say that if it is true it leads me to the amazing realization that spontaneity does not come by itself.

Raymond Smullyan
The Hill Cipher

- For example, Let $m = 2$,
- we could write a plaintext element as $x = (x_1, x_2)$
- ciphertext element as $y = (y_1, y_2)$.

Here, $y_1$, $y_2$ would be a linear combination of $x_1$ and $x_2$,

For example  We might take
$$y_1 = (11x_1 + 3x_2) \mod 26$$
$$y_2 = (8x_1 + 7x_2) \mod 26.$$

Of course, this can be written more succinctly in matrix notation:
$$(y_1, y_2) = (x_1, x_2), \begin{bmatrix} 11 & 8 \\ 3 & 7 \end{bmatrix}.$$
**Theorem:** A real matrix has an inverse if and only if its Determinant is non zero.

**Theorem:** Matrix K has a inverse modulo 26 if and only If gcd (det (K), 26) = 1

Ex: Let Plain Text \( x = (2, 3, 4) \)  
\[ k = \begin{pmatrix} 10 & 5 & 12 \\ 3 & 14 & 21 \\ 8 & 9 & 11 \end{pmatrix} \]

\[ y = x K = (61 \mod 26, 88 \mod 26, 131 \mod 26) = (9, 10, 1) \]
\[ K^{-1} = \begin{bmatrix} 21 & 15 & 17 \\ 23 & 2 & 16 \\ 25 & 4 & 3 \end{bmatrix} \]

\[ x = y K^{-1} = (9, 10, 1) \begin{bmatrix} 21 & 15 & 17 \\ 23 & 2 & 16 \\ 25 & 4 & 3 \end{bmatrix} \]

\[ = (2, 3, 4) \]
Rotor Machines

• Implement polyalphabetic substitution ciphers with long periods using a bank of rotors

• Each rotor has 26 electrical contacts on both its front and rear faces. Each contact on front is wired to a contact on the rear to define a mapping. Shifting the rotor changes the mapping. The key is given by the wiring and the initial positions of the rotors
3. Rotor Machines  (1870-1943)

Early example: the Hebern machine (single rotor)
Rotor Machines (cont.)

Most famous: the Enigma (3-5 rotors)

# keys = $26^4 = 2^{18}$ (actually $2^{36}$ due to plugboard)
Enigma Machines

- rotor machines - many different models
- 5-rotor machine has period of 11,881,376
- invented by Arthur Scherbius
- used by all sides during Spanish Civil War, 1938-39
- used by Germans through WW II
- captured and broken by allies
- broken with machines called “bombes”
- said to have played a significant role in allied victory
- sent key, enciphered twice, using an agreed starting position
Enigma Machine
Enigma with cover opened