POLYALPHABETIC CIPHERS
AND THEIR MATHEMATICS

CIS 400/628 — Spring 2005
Introduction to Cryptography

This is based on Chapter 2 of Lewand.
A SIDE DISCUSSION ON FUNCTIONS: INJECTIVE, SURJECTIVE, AND BIJECTIVE

Consider the following:

\[ L = \{ a, b, c, \ldots, x, y, z \} \]

\[ U = \{ A, B, C, \ldots, X, Y, Z \} \]

\[ f : L \rightarrow U \]

Some examples that match the type of \( f \):

\[ f_1(l) = toupper(l) \]

\[ f_2(l) = \begin{cases} 
Z & \text{if } l = a \\
A & \text{otherwise}
\end{cases} \]

\[ f_3(l) = \begin{cases} 
B & \text{if } l = a \\
C & \text{if } l = b \\
\ldots \\
Z & \text{if } l = y \\
A & \text{if } l = z
\end{cases} \]
INJECTIVE FUNCTIONS

▶ An injective function is 1-to-1, i.e., each element in the output is mapped to by only one element in the input.

▶ For example, consider

\[ f : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+ , f(i) = i + 1 \]

▶ This is 1-to-1.

▶ But, does \( f^{-1}(0) \) exist?

▶ From the earlier definition of \( L, U, f_1, \) and \( f_2 \):

- \( f_1(l) = toupper(l) \) is injective.
- \( f_2(l) = \begin{cases} \mathbb{Z} & \text{if } l = a \\ A & \text{otherwise} \end{cases} \) is not!
SURJECTIVE FUNCTIONS

▶ A surjective function maps at least one value onto every member of the output set.

▶ This is sometimes called an onto mapping.

▶ For example, consider

\[ f : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+, \quad f(i) = i \text{ div } 2 + 1 \]

\[ f(1) = 1, \quad f(2) = 2, \quad f(3) = 2, \quad f(4) = 3, \quad f(5) = 3, \ldots \]

▶ But, is it injective?

▶ Again, from our earlier definitions,

- \( f_1(l) = \text{toupper}(l) \) is surjective.

- \( f_2(l) = \begin{cases} 
\mathbb{Z} \text{ if } l = a \\
A \text{ otherwise}
\end{cases} \) is not!
A bleubijection function is both 1-to-1 and onto.

I.e., it is both injective and surjective.

Again recalling our definitions:

- \( f_3(l) = \begin{cases} 
B & \text{if } l = a \\
C & \text{if } l = b \\
\ldots & \text{is bijective.} \\
Z & \text{if } l = y \\
A & \text{if } l = z \\
\end{cases} \)

- \( f_2(l) = \begin{cases} 
Z & \text{if } l = a \\
A & \text{otherwise} \\
\end{cases} \)

Recall that monoalphabetic ciphers are bijective.
SCAMBLING FREQUENCES

▶ To get around the weakness of monoalphabetic ciphers, we need to scramble letter frequencies somehow.

▶ A polyalphabetic substitution cipher is a cipher in which there is not a 1–1 map between plaintext and ciphertext letters.

▶ An example from Lewand:

- Let $S = \{00, 01, 02, \ldots, 99\} =$ two digit strings
- Define a map $a_i \mapsto S_i$, a subset of $S \ni \star S_0, \ldots, S_{25}$ are a partition of $S$.
  - $\star$ $S_0, \ldots, S_{25}$ are a partition of $S$.
  - $\star$ (freq. of $a_i) \approx \|S_i\| / \|S\|$.
- When encoding $a_i$ pick a random element of $S_i$.
- In the ciphertext, the freq. of all two digit seqs. is about the same.

▶ To analyze these schemes we need counting & probability.
# EXAMPLE

<table>
<thead>
<tr>
<th>Letter</th>
<th>Subset of $S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>15, 33, 37, 55, 57, 72, 91, 96</td>
</tr>
<tr>
<td>b</td>
<td>24</td>
</tr>
<tr>
<td>c</td>
<td>03, 39, 67</td>
</tr>
<tr>
<td>e</td>
<td>08, 12, 20, 46, 47, 59, 64, 79, 81, 85, 90, 94, 97</td>
</tr>
<tr>
<td>h</td>
<td>05, 16, 30, 42, 69, 99</td>
</tr>
<tr>
<td>k</td>
<td>77</td>
</tr>
<tr>
<td>r</td>
<td>21, 25, 65, 68, 92, 95</td>
</tr>
<tr>
<td>s</td>
<td>00, 28, 52, 63, 74, 78</td>
</tr>
<tr>
<td>t</td>
<td>07, 19, 23, 35, 38, 54, 70, 84, 89</td>
</tr>
<tr>
<td>u</td>
<td>09, 32</td>
</tr>
</tbody>
</table>

Starbucks at three $\Rightarrow$ 52 38 33 65 24 32 39 77 00
15 70 07 69 68 59 46
The Multiplication Principle

If task 1 can be done \( p_1 \) ways and task 2 can be done \( p_2 \) ways and ...

... task \( k \) can be done \( p_k \) ways, then the total number of ways of doing all \( k \) tasks is \( p_1 \times p_2 \times \ldots \times p_k \)

Examples  How many are there of:

- License Plates with three letters followed by four digits
- License Plates as before, but no repeated chars.
- Monoalphabetic ciphers
A permutation is an ordering of a set of objects.

EXAMPLES

a. How many perms. of \{ a, b, c \} are there?

b. Four spies: A, B, C, D. We choose one a pilot and another as co-pilot. Q: How many ways are there of doing this?

c. There are five spies. Choose one to go to Miami and another to go to Watertown. Q: How many ways can we do this?

d. Same as above, but choose a 3rd to go to Jersey City.

e. Q: How many perms. are there of \( r \) objects selected from a set of size \( n \). (Notation: \( P(r, n) \).)
COMBINATIONS

▶ A combination is a selection of $r$ objects from a set of size $n$. (We don’t worry about order.)

▶ The number of ways of selecting (choosing) $r$ objects from a set of size $n$ is:

$$C(n, r) = \frac{n!}{r!(n-r)!} = \frac{1}{r!} \cdot P(n, r).$$

We also write this \( \binom{n}{r} \).

▶ EXAMPLE

Suppose a lottery ticket contains 6 numbers from \{ 0, \ldots, 39 \}.

Q: How many tickets are possible when order matters?
Q: How many when order doesn’t matter?
TERMINOLOGY

- **Sample space**: the possible outcomes of an experiment
- **Event**: a subset of a sample space

In this course, sample spaces are usually finite.

**To Determine the prob. of an event in a finite sample space**

1. Determine the elements of $S$, the sample space
2. Assign a weight to each element of $S \ni$ each weight is $\geq 0$
   - the weights sum to 1.
3. Probability of $E = \sum_{a \in E} \text{weight}(a)$. 
BASIC PROPERTIES OF PROBABILITY

- \( \neg E = \{ x \in S : x \notin E \} \).
- \( p(\neg E) = 1 - p(E) \).
- For all \( E \), \( 0 \leq p(E) \leq 1 \).
- If \( E \) and \( F \) are events of \( S \), then
  \[
  p(E \cup F) = p(E) + p(F) - p(E \cap F).
  \]
- Computing \( p(E \cap F) \) can be tricky.

EXAMPLE: Roll a 6 sided die.

- \( p(\text{rolling an odd number}) = \frac{1}{2} \).
- \( p(\text{rolling a prime}) = \frac{1}{2} \).
- \( p(\text{rolling an odd prime}) = \frac{1}{3} \neq \frac{1}{2} \cdot \frac{1}{2} \).
INDEPENDENCE

DEFINITION  Suppose $E, F \subseteq S$.

$E$ and $F$ are independent iff $p(E \cap F) = p(E) \cdot p(F)$.

$E$ and $F$ are dependent iff $p(E \cap F) \neq p(E) \cdot p(F)$.

DEFINITION  

If an experiment is repeated in $n$ independent trials & if the probability of an event $E$ is $p$,
then the expected number of events $(\text{Exp}(E))$ is $p \cdot n$.

EXAMPLE:

Flipping a coin 5 times, $S = \{ \text{HHHHH, \ldots, TTTTT} \}$

$E(\text{no heads}) = 1/32$  $E(\text{1 head}) = 5/32$

$E(\text{2 heads}) = 10/32$  $E(\text{3 head}) = 10/32$

$E(\text{4 heads}) = 5/32$  $E(\text{5 heads}) = 1/32$

Exp. num. of heads = $\frac{1}{2} \cdot 5 = 2.5$.  ← how to interp.?
The problem with the monoalphabetic ciphers is that the frequency of characters is unchanged.

**Vigenère Cipher**
Plaintext = mollywillneverbreakthis. Key = chaos.

\[
\begin{array}{cccccccccccccccc}
m & o & l & l & y & w & i & l & l & n & e & v & e & r & n & e & v & e \\
+ & c & h & a & o & s & c & h & a & o & s & c & h & a & o & s & c & h & a \\
\hline 
- & c & h & a & o & s & c & h & a & o & s & c & h & a & o & s & c & h & a \\
\hline 
m & o & l & l & y & w & i & l & l & n & e & v & e & r & n & e & v & e & r & b & r & e & a & k & t & h & i & s &  \\
\end{array}
\]
THE VIGENÈRE TABLE

|   | 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 |
| a | 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 |
| b | 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 |
| c | 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 |
| d | 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 |
| e | 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 |
| f | 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 |
| g | 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 |
| h | 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 |
| i | 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 |
| j | 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 |
| k | 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 |
| l | 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 |
| m | 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 |
| n | 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 |
| o | 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 |
| p | 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 |
| q | 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 |
| ... |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| y | 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 |
| z | 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 |
CRYPTANALYSIS OF THE VIGENÈRE CIPHER

▶ Look at repeated strings & the distance between them (why?)

\[
\begin{align*}
\text{the light brown fox jumps} & \quad \text{hound hound hound hound hound hound} \\
\text{A V Y Y L N V N O U V K H S R E X O Z S Z} & \\
\text{over the lazy dog} & \quad \text{hound hound hound hound hound hound hound} \\
\text{C P R U A V Y Y D G M X B J}
\end{align*}
\]

▶ The Kasiski test: the g.c.d. of these distances is likely to be a multiple of the keyword length. (why?)

▶ In real text, there would be several repeated sequences.
THE FRIEDMAN TEST

The index of coincidence = the probability of selecting two random letters from a text and getting the same letter

\[ n = \] the number of letters in a text
\[ n_i = \] the number of \( a_i \) in a text

Prob. of selecting two random letters & getting two 'a's

\[
\begin{align*}
\frac{n_0(n_0 - 1)}{2} / \frac{n(n - 1)}{2} &= \frac{n_0(n_0 - 1)}{n(n - 1)} \\
\therefore \text{Prob. that two randomly chosen letters are the same}
\end{align*}
\]

\[ = \sum_{i=0}^{25} \frac{n_i(n_i - 1)}{n(n - 1)} = \text{IC}(T) \approx \sum_{i=1}^{25} \left( \frac{n_i}{n} \right)^2 \]

For English: IC \( \approx \) 0.065  
For random text: \( \text{IC} \approx 0.038 \)

So?
USING THE FRIEDMAN TEST

- For a monoalphabetic cipher, the $IC \approx 0.065$.

- So, if the IC for the ciphertext is closer to 0.038 than to 0.065, it’s probably not a monoalphabetic cipher.

- **Key observation:** for a keyword of length $r$, a Vigenère cipher is really $r$ monoalphabetic additive ciphers used in rotation!

- Therefore, every $r^{th}$ letter is from the same cipher $\Rightarrow$
  - the IC for pairs of letters from the same cipher will be $\approx 0.065$.
  - the IC for pairs not from the same cipher will be $\approx 0.038$.

- So each cipher is used on $\approx n/r$ letters.
CALCULATING THE LENGTH OF THE KEY

Number of ways to choose two chars. from a particular cipher:

\[
\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)}{2}
\]

But there are \( r \) ciphers, so the overall number of ways to choose 2 letters from the same cipher:

\[
r \cdot \frac{n(n-1)}{2} = \frac{n(n-1)}{2}
\]

And out of that, we’d expect to get

\[
0.065 \cdot \frac{n(n-1)}{2}
\]

Pairs of letters. (Why?)
How about the number of ways to choose letters not from a particular cipher?

\[
\left( \text{# of ways to choose 1 letter from the cipher} \right) \times \left( \text{# of ways to choose 1 letter not from the cipher} \right) \div 2 = \frac{n}{r}(n - \frac{n}{r}) \div 2
\]

Why divide by 2?

As before, there are \( r \) ciphers, so the overall number of ways to choose 2 letters not from the same cipher is:

\[
r \cdot \frac{n}{r}(n - \frac{n}{r}) \div 2 = n \cdot \frac{n}{r}(n - \frac{n}{r}) \div 2 \quad \text{and} \quad 0.038 \cdot \frac{n}{r}(n - \frac{n}{r}) \div 2
\]

is the expected number of pairs from this choosing.
CALCULATING THE LENGTH OF THE KEY, III

Given all that, the total number of expected pairs, $EN$, is thus:

$$EN = 0.065 \cdot \frac{n \cdot (\frac{n}{r} - 1)}{2} + 0.038 \cdot \frac{n \cdot (n - \frac{n}{r})}{2} =$$

$$0.065 \cdot \frac{n \cdot (n - r)}{2r} + 0.038 \cdot \frac{n^2 \cdot (r - 1)}{2r}.$$

So the probability of both pairs being the same is

$$IC \approx \frac{EN}{\binom{n}{2}} = \frac{2EN}{n \cdot (n - 1)} =$$

$$\frac{1}{r \cdot (n - 1)} \cdot (0.027n + r \cdot (0.038n - 0.065)).$$
PUTTING IT ALL TOGETHER

Solving for $r$ yields

$$r \approx \frac{0.027n}{IC \cdot (n - 1) - 0.038n + 0.065}$$

But $r$ is the key length! So, for a given ciphertext:

1. Calculate the IC for the ciphertext.
2. If it’s nearer to 0.038 than 0.065, try to solve for Vigenêre
3. Use the Friedman test to calculate $r$.
4. Use the Kasiski test to produce the g.c.d. of the expected key length.
5. Try keys whose length is the multiple of the Kasiski test result nearest the result of the Friedman test.
THE ONE TIME PAD

▶ This is the only provably secure cipher.
▶ Requires a key as long as the cleartext.
▶ Keys may never be reused.
▶ Keys must be randomly generated (but finding true randomness is hard).
▶ That means that every cleartext is as likely as any other.
▶ Use the mod 26 addition on a letter by letter basis:

\[
\begin{align*}
\text{now is the time for all} & \ldots \\
+ \text{FFIWM} & \text{LJZA} \text{NORVEKPB} \ldots \\
\hline \\
\text{ITEEEEQDTRZSWJVKA} & \text{M} \ldots \\
\end{align*}
\]