Polyalphabetic cyphers

Let $E_1, E_2, \ldots, E_n$ be distinct substitution cyphers. To encrypt a plaintext message $P = p_1 p_2 p_3 \ldots$ apply the $E_i$ ($i = 1, \ldots, n$) cyclically to the plaintext characters. Thus $E(P) = E_1(p_1)E_2(p_2)E_3(p_3) \ldots E_n(p_n)E_1(p_{n+1})E_2(p_{n+2}) \ldots$

**block size** = $n$ characters
Special Case: Vignère Cipher

Let \{a, b, c, \ldots, z\} = \mathbb{Z}_{26}, \text{ i.e.}
\begin{align*}
a &\leftrightarrow 0, \quad b \leftrightarrow 1, \ldots, \quad z \leftrightarrow 25. 
\end{align*}

Key is a sequence \(k_1, k_2, \ldots, k_n \in \mathbb{Z}_{26}\)

Define \(E_i(x) = (x + k_i) \% 26\), for \(i = 1, \ldots, n\)

When \(n = 1\) this is a shift cipher

When \(n = 1, k_1 = 3\) this is the Caeser cipher

Sometimes work mod 27 and include blank as a character
Key could be any string, but is generally a word or phrase

Example:

key = cyclone = 2, 24, 2, 11, 14, 13, 4

\( E(\text{pedestrian}) = \text{RCFPGGVKYP} \)
Breaking a polyadic cipher

Consider the message to be \( n \) separate “messages”

\[
\begin{align*}
    c_1 & \quad c_{n+1} & \quad c_{2n+1} & \quad \ldots \\
    c_2 & \quad c_{n+2} & \quad c_{2n+2} & \quad \ldots \\
    \vdots & & & \\
    c_n & \quad c_{2n} & \quad c_{3n} & \quad \ldots 
\end{align*}
\]

Cryptanalyze each line as a separate substitution cypher.
Effective for messages longer than about $25n$ characters (English).

For this method to be effective, we must know the block size (i.e., the value of $n$).

- Kasiski’s method
- The index of coincidence
Kasiski’s Method

123456712345671234567123456712345671
theharderIworkthemoreluckiseemtohave
BYSJOPXYLQVOPBBYSZOPSRNUKWT SYSKVILLY
Assume a polyalphabetic cypher, with block size $n$

A string appears twice in plaintext and distance between appearances is a multiple of $n$

cyphertext will have a repeated string at same distance

Converse?
Suppose cyphertext has a repeated string. Can we assume that the distance between appearances is a multiple of block size?

Probability that a ciphertext trigram appears twice (spontaneously) is \( \left( \frac{1}{26} \right)^3 \approx .00006 \).
Kasiski’s method:

1. Find some repeated strings in cyphertext of length at least 3
2. Compute distances between them
3. Block size will be a common divisor of those numbers
WJE appears in positions 292, 346, 382
$346 - 292 = 54$, $383 - 346 = 36$

IWO appears in positions 69, 317, 323
$317 - 69 = 248$, $323 - 317 = 6$

All of the differences are divisible by 6 except 248. So that is probably the block size. (Or maybe the block size is 2 or 3.)
Index of coincidence

Let $a_1, a_2, \ldots, a_t$ be a block of text from some source on the alphabet $\{a, \ldots, z\}$. Define $\text{IC}(a_1, \ldots, a_t)$ to be the probability that for some (fixed) $i \neq j$, $a_i = a_j$.

Suppose first that the source consists of random letters. Then $\text{IC}(a_1, \ldots, a_t) = \frac{1}{26} \approx 0.038$. 
Probabilities for English

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>.082</td>
<td>b</td>
<td>.015</td>
<td>c</td>
<td>.028</td>
</tr>
<tr>
<td>f</td>
<td>.022</td>
<td>g</td>
<td>.020</td>
<td>h</td>
<td>.061</td>
</tr>
<tr>
<td>k</td>
<td>.008</td>
<td>l</td>
<td>.040</td>
<td>m</td>
<td>.024</td>
</tr>
<tr>
<td>p</td>
<td>.019</td>
<td>q</td>
<td>.001</td>
<td>r</td>
<td>.060</td>
</tr>
<tr>
<td>u</td>
<td>.028</td>
<td>v</td>
<td>.010</td>
<td>w</td>
<td>.023</td>
</tr>
<tr>
<td>z</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Write $p_a = .082$, $p_b = .015$, $p_c = .028$, $p_d = .043$, $p_e = .127$, $p_f = .022$, $p_g = .020$, $p_h = .061$, $p_i = .070$, $p_j = .002$, $p_k = .008$, $p_l = .040$, $p_m = .024$, $p_n = .067$, $p_o = .075$, $p_p = .019$, $p_q = .001$, $p_r = .060$, $p_s = .063$, $p_t = .091$, $p_u = .028$, $p_v = .010$, $p_w = .023$, $p_x = .001$, $p_y = .020$, $p_z = .001$. 
If \( a_1, \ldots, a_t \) are characters drawn from English-language text, then \( \text{IC}(a_1, a_2, \ldots, a_t) = p_a^2 + p_b^2 + \cdots + p_z^2 \approx 0.065 \).

Finally, suppose that \( c_1, c_2, \ldots, c_t \) is obtained from English text by a monoalphabetic substitution. Then we would expect \( \text{IC}(c_1, \ldots, c_t) = 0.065 \).
Let $c_1, \ldots, c_t$ be some block of ciphertext.

# of appearances of ‘A’ is denoted $u_A$

# of appearances of ‘B’ is denoted $u_B$ etc.

Let $1 \leq i < j \leq t$.

\[
P(c_i = c_j = A) = \frac{u_A}{t} \cdot \frac{u_A-1}{t-1},
\]

\[
P(c_i = c_j = B) = \frac{u_B}{t} \cdot \frac{u_B-1}{t-1} \quad \text{etc.}
\]

Therefore $\text{IC}(c_1, \ldots, c_t) \approx \frac{u_A(u_A-1) + u_B(u_B-1) + \cdots + u_Z(u_Z-1)}{t(t-1)}$.  

© 2011 Clifford Bergman
Example: Using the (monoalphabetic) cyphertext from the last lecture:

IJXMYIAXMYKDIKZJYBKLXYLXIDMHBJHHJXIXYJAEVDR
HMWXWJYIMHXZDAMJYJSOVDWCLJVXIIUBBXIAIYJAJY
VRALDABJHHJXIEVDRHMXOUAALDALXIJFXAMFXIWJY
SUIXIUIORALZJMKMYBALXFKLXZXLXRWDYAOXIXXY

we obtain $IC \approx 0.0656$. 
Suppose we have cyphertext $c_1, c_2, \ldots, c_t$ that we suspect is polyalphabetic with block size $n$. Take the $n$ substrings:

\[
\begin{align*}
    c_1 & \quad c_{n+1} & \quad c_{2n+1} & \quad \ldots \\
    c_2 & \quad c_{n+2} & \quad c_{2n+2} & \quad \ldots \\
    \vdots & & & \\
    c_n & \quad c_{2n} & \quad c_{3n} & \quad \ldots
\end{align*}
\]

Index of coincidence of each line should be approx. 0.065.
Returning to the example from earlier

The IC of the entire string is 0.037

If we guess $n = 2$ we compute:

\[
\begin{align*}
\text{IC(} & \text{LSRR} \ldots \text{)} = 0.051 \\
\text{IC(} & \text{PRNK} \ldots \text{)} = 0.044
\end{align*}
\]

$n = 3$: 0.059, 0.054, 0.050  \\
$n = 4$: 0.052, 0.041, 0.049, 0.044  \\
$n = 5$: 0.038, 0.042, 0.039, 0.041, 0.040.  \\
$n = 6$: 0.076, 0.069, 0.055, 0.062, 0.059, 0.083

Block size is surely 6 (as we expected from Kasiski).
As the blocksize increases, these statistical calculations become less and less effective.

If we make the blocksize as long as the message, then the ciphertext will be indistinguishable from random text.

☆ The one-time pad
In practice, a one-time pad is implemented as a Vigènere cipher with a block size as long as the message.

\[ \text{pad} = t_1, t_2, t_3, \ldots, \text{with } t_i \in \{0, 1, \ldots, 26\} \text{ random} \]
Example:

Ciphertext:
GMQGIHXABURHFMMXGJDAHGN ▼ ECD ▼ PBYJ

If key is:
▼ DVI VSAQLPCOTOXSSDUZLI ▼ SYDWKAEB

then plaintext is:
give me liberty or give me death
Example:

Ciphertext:
GMQGIHXABURHFMMXGJDAHGN \text{ECD} \text{PBYJ}

But key could just as well be:
FAEGKTCJBSQPAMLFBJBWWS \text{TEJP} \text{VJYJ}

in which case plaintext is:
all your base are belong to us
Under a one-time pad it is impossible to determine which string (of the right length) is the correct plaintext decryption of a given ciphertext.
The Enigma machine was employed by the German military during World War II.

Implemented a polyalphabetic cipher of block size $26^3 \approx 17,000$.

However the individual substitutions were not independent.