Attacks Against Classical Ciphers

Ahto Buldas

September 23, 2019

Cryptosystem

$\mathbf{X}$ – set of all possible plaintexts
$\mathbf{Y}$ – set of all possible ciphertexts
$\mathbf{Z}$ – set of all possible keys

Encryption and Decryption: For every $z \in \mathbf{Z}$, there are functions

$$E_z \colon \mathbf{X} \to \mathbf{Y} \qquad \text{and} \qquad D_z \colon \mathbf{Y} \to \mathbf{X} \ ,$$

such that $D_z(E_z(x)) = x$ for every $x \in \mathbf{X}$

Substitution Cipher

Every letter is substituted with another letter, by using a 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
Q F Y B R I W Z D J G X O P K N V S A H C L T E M U
```

For example a plaintext `MESSAGE` is encrypted to `ORAAQWR`:

```
M E S S A G E
O R A A Q W R
```

$\mathbf{X}$ – all possible texts
$\mathbf{Z}$ – all possible permutations of the 26-letter alphabet

$|\mathbf{Z}| = 26! = 2 \cdot 3 \cdot \ldots \cdot 25 \cdot 26 \approx 2^{88}$

Shift Cipher

Convert letters to numbers:

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Shift cipher $y = E_z(x)$, where $x, y, z \in \{0, 1, 2, \ldots, 25\}$:

$$y = E_z(x) = x + z \bmod 26 = \begin{cases} x + z & \text{if } x + z < 26 \\ x + z - 26 & \text{if } x + z \geq 26 \end{cases}$$

Breaking a Shift Cipher

Assume we have a ciphertext:

LSAQERCQMGWHSAIVMTSRXLIHEMPC

and we suspect the use of the shift cipher.
Try to decrypt with all keys, starting from $z = 1$:

z	Decrypted text:
1	KRZPDQBPLFVGRZHULSRQWKHGDLOB

Breaking a Shift Cipher

Assume we have a ciphertext:

LSAQERCQMGWHSAIVMTSRXLIHEMPC

and we suspect the use of the shift cipher.
Try to decrypt with all keys, starting from $z = 1$:

z	Decrypted text:
1	KRZPDQBPLFVGRZHULSRQWKHGDLOB
2	JQYOCPAOKEUFQYGTKRQPVJGFCKNA

Breaking a Shift Cipher

Assume we have a ciphertext:

`LSAQERCQMGWHSAIVMTSRXLIHEMPC`

and we suspect the use of the shift cipher.
Try to decrypt with all keys, starting from $z = 1$:

z	Decrypted text:
1	KRZPDQBPLFVGRZHULSRQWKHGDLOB
2	JQYOCPAOKEUFQYGTKRQPVJGFCKNA
3	IPXNBOZNJDTEPXFSJQPOUIFEBJMZ

Breaking a Shift Cipher

Assume we have a ciphertext:

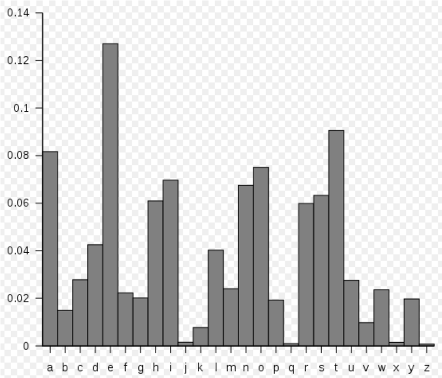
LSAQERCQMGWHSAIVMTSRXLIHEMPC

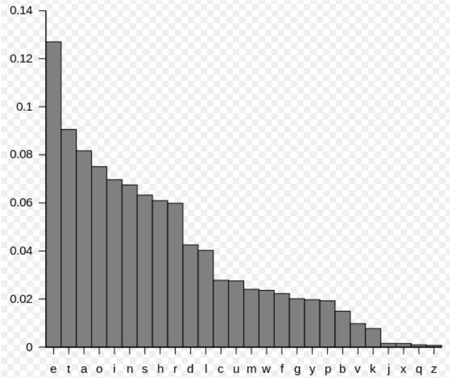
and we suspect the use of the shift cipher.
Try to decrypt with all keys, starting from $z = 1$:

z	Decrypted text:
1	KRZPDQBPLFVGRZHULSRQWKHGDLOB
2	JQYOCPAOKEUFQYGTKRQPVJGFCKNA
3	IPXNBOZNJDTEPXFSJQPOUIFEBJMZ
4	HOWMANYMICSDOWERIPONTHEDAILY

Frequency Analysis

Frequencies of English letters:





Breaking a Substitution Cipher

The next example is from the wikipedia page "Frequency analysis"

Suppose we have a ciphertext:

```
LIVITCSWPIYVEWHEVSRIQMXLEYVEOIEWHRXEXIPFEMVEWHKVSTYLXZIXLIKIIXPIJVSZEYPERRGERIM
WQLMGLMXQERIWGPSRIHMXQEREKIETXMJTPRGEVEKEITREWHEXXLEXXMZITWAWSQWXSWEXTVEPMRXRSJ
GSTVRIEYVIEXCVMUIMWERGMIWXMJMGCSMWXSJOMIQXLIVIQIVIXQSVSTWHKPEGARCSXRWIEVSWIIBXV
IZMXFSJXLIKEGAEWHEPSWYSWIWIEVXLISXLIVXLIRGEPIRQIVIIBGIIHMWYPFLEVHEWHYPSRRFQMXLE
PPXLIECCIEVEWGISJKTVWMRLIHYSPHXLIQIMYLXSJXLIMWRIGXQEROIVFVIZEVAEKPIEWHXEAMWYEPP
XLMWYRMWXSGSWRMHIVEXMSWMGSTPHLEVHPFKPEZINTCMXIVJSVLMRSCMWMSWVIRCIGXMWYMX
```

X~t means a guess that ciphertext letter X represents the plaintext letter t.

Breaking a Substitution Cipher

The next example is from the wikipedia page "Frequency analysis"

Suppose we have a ciphertext:

```
LIVITCSWPIYVEWHEVSRIQMXLEYVEOIEWHRXEXIPFEMVEWHKVSTYLXZIXLIKIIXPIJVSZEYPERRGERIM
WQLMGLMXQERIWGPSRIHMXQEREKIETXMJTPRGEVEKEITREWHEXXLEXXMZITWAWSQWXSWEXTVEPMRXRSJ
GSTVRIEYVIEXCVMUIMWERGMIWXMJMGCSMWXSJOMIQXLIVIQIVIXQSVSTWHKPEGARCSXRWIEVSWIIBXV
IZMXFSJXLIKEGAEWHEPSWYSWIWIEVXLISXLIVXLIRGEPIRQIVIIBGIIHMWYPFLEVHEWHYPSRRFQMXLE
PPXLIECCIEVEWGISJKTVWMRLIHYSPHXLIQIMYLXSJXLIMWRIGXQEROIVFVIZEVAEKPIEWHXEAMWYEPP
XLMWYRMWXSGSWRMHIVEXMSWMGSTPHLEVHPFKPEZINTCMXIVJSVLMRSCMWMSWVIRCIGXMWYMX
```

X˜t means a guess that ciphertext letter X represents the plaintext letter t.

Observations:

- I is the most common single letter (in English: e)
- XL most common bigram (in English: th)
- XLI is the most common trigram (in English: the)

This strongly suggests that X˜t, L˜h and I˜e.

Breaking a Substitution Cipher

The second most frequent ciphertext letter is E.

As the first and second most frequent letters in the English language: e and t already accounted) we guess that E˜a.

We obtain the next partial decrypted message:

```
heVeTCSWPeYVaWHaVSReQMthaYVaOeaWHRtatePFaMVaWHKVSTYhtZetheKeetPeJVSZaYPaRRGaReM
WQhMGhMtQaReWGPSReHMtQaRaKeaTtMJTPRGaVaKaeTRaWHatthattMZeTWAWSQWtSWatTVaPMRtRSJ
GSTVReaYVeatCVMUeMWaRGMeWtMJMGCSMWtSJOMeQtheVeQeVetQSVSTWHKPaGARCStRWeaVSWeeBtV
eZMtFSJtheKaGAaWHaPSWYSWeWeaVtheStheVtheRGaPeRQeVeeBGeeHMWYPFhaVHaWHYPSRRFQMtha
PPtheaCCeaVaWGeSJKTVWMRheHYSPHtheQeMYhtSJtheMWReGtQaROeVFVeZaVAaKPeaWHtaAMWYaPP
thMWYRMWtSGSWRMHeVatMSWMGSTPHhaVHPFKPaZeNTCMteVJSVhMRSCMWMSWVeRCeGtMWYMt
```

Now we can spot patterns, such as "that", and other patterns:

- "Rtate" might be "state", which suggests R˜s.
- "atthattMZe" could be "atthattime", which yields M˜i and Z˜m.
- "heVe" might be "here", suggesting V˜r.

Breaking a Substitution Cipher

We now have the following partially decrypted message:

```
hereTCSWPeYraWHarSseQithaYraOeaWHstatePFairaWHKrSTYhtmetheKeetPeJrSmaYPassGasei
WQhiGhitQaseWGPSseHitQasaKeaTtiJTPsGaraKaeTsaWHatthattimeTWAWSQWtSWatTraPistsSJ
GSTrseaYreatCriUeiWasGieWtiJiGCSiWtSJOieQthereQeretQSrSTWHKPaGAsCStsWearSWeeBtr
emitFSJtheKaGAaWHaPSWYSWeWeartheStherthesGaPesQereeBGeeHiWYPFharHaWHYPSssFQitha
PPtheaCCearaWGeSJKTrWisheHYSPHtheQeiYhtSJtheiWseGtQasOerFremarAaKPeaWHtaAiWYaPP
thiWYsiWtSGSWsiHeratiSWiGSTPHharHPFKPameNTCiterJSrhisSCiWiSWresCeGtiWYit
```

Some more guessing leads to:

```
hereuponlegrandarosewithagraveandstatelyairandbroughtmethebeetlefromaglasscasei
nwhichitwasenclosedItwasabeautifulscarabaeusandatthattimeunknowntonaturalistsof
courseagreatprizeinascientificpointofviewtherewereworoundblackspotsnearoneextr
emityofthebackandalongoneneartheotherthescaleswereexceedinglyhardandglossywitha
lltheappearanceofburnishedgoldtheweightoftheinsectwasveryremarkableandtakingall
thingsintoconsiderationicouldhardlyblamejupiterforhisopinionrespectingit
```

Breaking a Substitution Cipher

Now we add the spaces and punctuation and get the decrypted text:

Hereupon Legrand arose, with a grave and stately air, and brought me the beetle from a glass case in which it was enclosed. It was a beautiful scarabaeus, and, at that time, unknown to naturalists—of course a great prize in a scientific point of view. There were two round black spots near one extremity of the back, and a long one near the other. The scales were exceedingly hard and glossy, with all the appearance of burnished gold. The weight of the insect was very remarkable, and, taking all things into consideration, I could hardly blame Jupiter for his opinion respecting it.

The text is from "The Gold-Bug": a story by Edgar Allan Poe from 1843.

Vigenere Cipher

$\mathbf{Z}$ – all possible m-letter keys: $z_0 z_1 \ldots z_{m-1}$
$\mathbf{X}$ – all possible n-letter messages: $x_1 x_2 \ldots x_n$
$\mathbf{Y}$ – all possible n-letter ciphertexts: $y_1 y_2 \ldots y_n$

Encrypt every letter x_i with the key $z_{i \bmod m}$:

$$y_i = x_i + z_{i \bmod m} \mod 26$$

How to Attack Vigenere Ciphers

- Find m by using statistical methods
- Find the differences between the keys $z_0, z_1, \ldots, z_{m-1}$
- Express all keys as linear functions from one single key z_i
- Try all values of z_i

Finding m by Kasiski Examination

The Kasiski examination as a method was first published in 1863 by Friedrich Kasiski (1805–1881) who was a German infantry officer, cryptographer and archeologist.

If there are similar groups of (at least 3) letters in the ciphertext, like:

AFRTASKGHTUCXZAFRTDSFHHJJ

Then the most probable explanation is that they correspond to similar groups of letters in the plaintext

Hence, the difference in their positions in the text is divisible by m

Index of Coincidence

The index of coincidence was discovered by US Army cryptographer William Frederick Friedman (1891–1969). He ran the research division of the Army's Signal Intelligence Service (SIS) in the 1930s.

Let X be an N-letter text, and $n_a, n_b, \ldots$ denote the numbers of ocurrences of a, b, ... in X.

The *index of coincidence* $\mathbf{IC}(X)$ of X is the probability that two random letters of X are equal. It is easy to see that

$$\mathbf{IC}(X) = \frac{n_a}{N} \cdot \frac{n_a - 1}{N-1} + \frac{n_b}{N} \cdot \frac{n_b - 1}{N-1} + \ldots + \frac{n_z}{N} \cdot \frac{n_z - 1}{N-1}$$

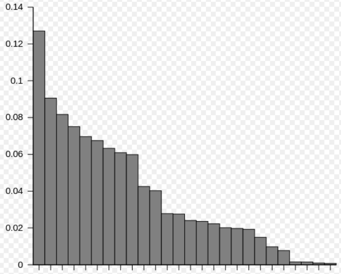
$\mathbf{IC}(X) \approx 0.038$ for a random text ≈ 0.065 for a meaningful text.

An Important Property of $\mathbf{IC}$

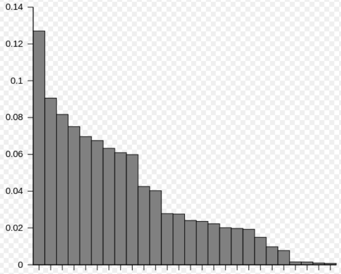
If Y is a ciphertext obtained from a plaintext X via enciphering it using a substitution cipher, then:

$$\mathbf{IC}(Y) = \mathbf{IC}(X)$$

Explanation: The sorted frequency distributions of X and Y are the same:



$X : e, t, a, o, \ldots$



$Y : E(e), E(t), E(a), E(o), \ldots$

Mutual Index of Coincidence

Let X be an N-letter text, where $n_a, n_b, ...$ denote the numbers of ocurrences of a, b, ... in X
Let Y be an N'-letter text, where $n'_a, n'_b, ...$ denote the number of occurrences of a, b, ... in Y

The mutual index of coincidence

$$\mathbf{IC}(X,Y) = \frac{n_a}{N}\frac{n'_a}{N'} + \frac{n_b}{N}\frac{n'_b}{N'} + \ldots + \frac{n_z}{N}\frac{n'_z}{N'}$$

of X and Y is the probability that $x = y$, where x and y are randomly chosen letters from X and Y, respectively.

An Important Property of $\mathbf{IC}(X, Y)$

Say $Y = y_1 y_2 \ldots y_n$ and $Y' = y'_1 y'_2 \ldots y'_m$ are two ciphertexts obtained from meaningful (English) plaintexts:

$$X = x_1 x_2 \ldots x_n \qquad \text{and} \qquad X' = x'_1 x'_2 \ldots x'_m$$

by using the *shift cipher* with the keys z and z', respectively:

$$y_i = x_i + z \mod 26 \qquad \text{and} \qquad y'_i = x'_i + z' \mod 26$$

Then:

$$\mathbf{IC}(Y, Y') \approx \begin{cases} 0.065 & \text{if } z = z' \\ 0.038 & \text{if } z \neq z' \end{cases}$$

Hence, we can see whether Y and Y' are encrypted with the same key or not.

Finding the difference $z - z'$ of two keys

Let $D_g(Y)$ denote the decryption functionality of the shift cipher, i.e. for any ciphertext letter y_i

$$D_g(y_i) = y_i - g \mod 26$$

Then for any $g = 0, 1, 2, \ldots, 25$:

$$\begin{aligned} \mathbf{IC}(Y, D_g(Y')) &= \mathbf{IC}(E_z(X), E_{z-g}(X')) \\ &\approx \begin{cases} 0.065 & \text{if } g = z' - z \mod 26 \\ 0.038 & \text{if } g \neq z' - z \mod 26 \end{cases} \end{aligned}$$

Say we have a ciphertext:

```
CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX
VRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
AMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBI
PEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHP
WQAIIWXNRMGWOIIFKEE
```

(From: Douglas R. Stinson. Cryptography: Theory and Practice. 1995.)

Kasiski examination

CHR repeats in positions: 1, 166, 236, 276 and 286

```
CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX
VRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
AMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBI
PEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHP
WQAIIWXNRMGWOIIFKEE
```

Differences of positions are: 165, 235, 275, and 285.
As $\gcd(165, 235, 275, 285) = 5$, we guess that $m = 5$.

Partial Texts: Encrypted with the same key

Y_1:CVABWEBQBUAWWQRWWXANTBDPXXRDWBFAXCWMNJJFAIACNRNCATBWKDMCDCQQXWK
Y_2:HOEITESEWOOEGMFTIFUDSTNSNVTNDPASNHESBGSEGEMRDRSHEAIEORTHNHOANOE
Y_3:RARANOBQRASAJNVRAPTCXUGRJRUQTHLVGRLJHNGYNQRRGINRYQPVEEBRVRHIRIE
Y_4:EHAXXPQEVKXHMKGZKSEMMIMEEVLWYXJBLZEIWMLPRJVELMRQEEEKWMHTRCPIMI
Y_5:EMTXBHMRXXXBMGXXLKMGXAGLLPHTGTHFLBKKRGXHBTLMXGWHVBEAAXHKZLWWGF

Check the indices of coincidence:

$$\mathbf{IC}(Y_1) = 0.063, \mathbf{IC}(Y_2) = 0.068, \mathbf{IC}(Y_3) = 0.061, \mathbf{IC}(Y_4) = 0.072\ .$$

This confirms that $m = 5$

Finding the Differences of Keys

Compute mutual indices:

$$\mathbf{IC}(X_i, D_g(X_j)) = \sum_{h=0}^{25} f_h \cdot f'_{h-g} \approx \sum_{h=0}^{25} p_h \cdot p_{h+(k_i-k_j)-g}$$

for all pairs $i \neq j$ and for all values of $g = 0, 1, \ldots, 25$
If $g = k_i - k_j$, then $(k_i - k_j) - g = 0$ and hence

$$\mathbf{IC}(X_i, D_g(X_j)) = \sum_{h=0}^{25} p_h \cdot p_h \approx 0.065 \ .$$

$i,\ j$	$\mathbf{IC}(X_i, D_g(X_j))$, where $g = 0, 1, \ldots 25$								
1,2	0.029	0.028	0.028	0.034	0.040	0.038	0.026	0.026	0.052
	0.069	0.045	0.026	0.038	0.043	0.038	0.044	0.038	0.029
$g = 9$	0.042	0.041	0.034	0.037	0.052	0.046	0.042	0.037	
1,3	0.040	0.034	0.040	0.034	0.028	0.054	0.049	0.034	0.030
	0.056	0.051	0.046	0.040	0.041	0.036	0.038	0.033	0.027
	0.038	0.037	0.032	0.037	0.055	0.030	0.025	0.037	
1,4	0.034	0.043	0.026	0.027	0.039	0.050	0.040	0.033	0.030
	0.034	0.039	0.045	0.044	0.034	0.039	0.046	0.045	0.038
	0.056	0.047	0.033	0.027	0.040	0.038	0.040	0.035	
1,5	0.043	0.033	0.028	0.046	0.043	0.045	0.039	0.032	0.027
	0.031	0.036	0.041	0.042	0.024	0.020	0.048	0.070	0.044
$g = 16$	0.029	0.039	0.044	0.043	0.047	0.034	0.026	0.046	
2,3	0.046	0.049	0.041	0.032	0.036	0.035	0.037	0.030	0.025
	0.040	0.035	0.030	0.041	0.068	0.041	0.033	0.038	0.045
$g = 13$	0.033	0.033	0.028	0.034	0.046	0.053	0.042	0.030	

$i,\ j$	$\mathbf{IC}(X_i, D_g(X_j))$, where $g = 0, 1, \ldots 25$								
2,4	0.046	0.035	0.044	0.045	0.034	0.031	0.041	0.046	0.040
	0.048	0.045	0.034	0.024	0.028	0.042	0.040	0.027	0.035
	0.050	0.035	0.033	0.040	0.057	0.043	0.029	0.028	
2,5	0.033	0.033	0.037	0.047	0.027	0.018	0.044	0.081	0.051
	0.030	0.031	0.045	0.039	0.037	0.028	0.027	0.031	0.040
$g = 7$	0.040	0.038	0.041	0.046	0.045	0.043	0.035	0.031	
3,4	0.039	0.036	0.041	0.034	0.037	0.061	0.035	0.041	0.030
	0.059	0.035	0.036	0.034	0.054	0.031	0.033	0.036	0.037
	0.036	0.029	0.046	0.033	0.052	0.033	0.035	0.031	
3,5	0.036	0.034	0.034	0.036	0.030	0.044	0.044	0.050	0.026
	0.041	0.052	0.051	0.036	0.032	0.033	0.034	0.052	0.032
$g = 20$	0.027	0.031	0.072	0.036	0.035	0.033	0.043	0.027	
4,5	0.052	0.039	0.033	0.039	0.042	0.043	0.037	0.049	0.029
	0.028	0.037	0.061	0.033	0.034	0.032	0.053	0.034	0.027
$g = 11$	0.039	0.043	0.034	0.027	0.030	0.039	0.048	0.036	

Solve the System

$$\begin{cases} z_1 - z_2 & \equiv & 9 & \pmod{26} \\ z_1 - z_5 & \equiv & 16 & \pmod{26} \\ z_2 - z_3 & \equiv & 13 & \pmod{26} \\ z_2 - z_5 & \equiv & 7 & \pmod{26} \\ z_3 - z_5 & \equiv & 20 & \pmod{26} \\ z_4 - z_5 & \equiv & 11 & \pmod{26} \end{cases}$$

We obtain that the key is:

$$z_1, z_1 + 17, z_1 + 4, z_1 + 21, z_1 + 10 \ ,$$

where the addition is modulo 26.

Solution

The key is JANET and the plaintext:

THEALMONDTREEWASINTENTATIVEBLOSSOMTHEDAYSW
ERELONGEROFTENENDINGWITHMAGNIFICENTEVENING
SOFCORRUGATEDPINKSKIESTHEHUNTINGSEASONWASO
VERWITHHOUNDSANDGUNSPUTAWAYFORSIXMONTHSTHE
VINEYARDSWEREBUSYAGAINASTHEWELLORGANIZEDFA
RMERSTREATEDTHEIRVINESANDTHEMORELACKADAISI
CALNEIGHBORSHURRIEDTODOTHEPRUNINGTHEYSHOUL
DHAVEDONEINNOVEMBER