Cryptosystems

Definition

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

1. $\mathcal{P}$ is a set of possible *plaintexts*
2. $\mathcal{C}$ is a set of possible *ciphertexts*
3. $\mathcal{K}$ is a set of possible *keys*
4. for each $\kappa \in \mathcal{K}$, there is an *encryption rule* $e_\kappa \in \mathcal{E}$ and a corresponding *decryption rule* $d_\kappa \in \mathcal{D}$. Each $e_\kappa : \mathcal{P} \to \mathcal{C}$ and $d_\kappa : \mathcal{C} \to \mathcal{D}$ are functions such that $d_\kappa(e_\kappa(m)) = m$ for every plaintext $m \in \mathcal{P}$.

Alphabet and encoding

letter	a	b	c	d	e	f	g	h	i
value	0	1	2	3	4	5	6	7	8
letter	j	k	l	m	n	o	p	q	r
value	9	10	11	12	13	14	15	16	17
letter	s	t	u	v	w	x	y	z	
value	18	19	20	21	22	23	24	25	26

Shift Cipher definition

$$\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}/26\mathbb{Z}$$

For $0 \leq \kappa \leq 25$

$$e_\kappa(m) = m + \kappa \bmod 26$$

and

$$d_\kappa(c) = c - \kappa \bmod 26$$

where $m, c \in \mathbb{Z}/26\mathbb{Z}$

Plain text

It was at least certain that Phileas Fogg had not absented himself from London for many years. Those who were honoured by a better acquaintance with him than the rest, declared that nobody could pretend to have ever seen him anywhere else.. . .
. . . , who proffered the viands in special porcelain, and on the finest linen; club decanters, of a lost mould, contained his sherry, his port, and his cinnamon-spiced claret; while his beverages were refreshingly cooled with ice, brought at great cost from the American lakes.

“Around the world in 80 days” by Jules Verne

Shift Cipher encryption example

It was at least certain that ...

itwasatleastcertainthat

t	i	t	w	a	s	a	t	l	e	a	s	t

Shift Cipher encryption example

It was at least certain that ...

itwasatleastcertainthat

t	i	t	w	a	s	a	t	l	e	a	s	t
v	8	19	22	0	18	0	19	11	4	0	18	19

Shift Cipher encryption example

It was at least certain that ...

itwasatleastcertainthat

t	i	t	w	a	s	a	t	l	e	a	s	t
v	8	19	22	0	18	0	19	11	4	0	18	19
k	c	c	c	c	c	c	c	c	c	c	c	c

Shift Cipher encryption example

It was at least certain that ...

itwasatleastcertainthat

t	i	t	w	a	s	a	t	l	e	a	s	t
v	8	19	22	0	18	0	19	11	4	0	18	19
k	c	c	c	c	c	c	c	c	c	c	c	c
$+$	2	2	2	2	2	2	2	2	2	2	2	2

Shift Cipher encryption example

It was at least certain that ...

itwasatleastcertainthat

t	i	t	w	a	s	a	t	l	e	a	s	t
v	8	19	22	0	18	0	19	11	4	0	18	19
k	c	c	c	c	c	c	c	c	c	c	c	c
$+$	2	2	2	2	2	2	2	2	2	2	2	2
v	10	21	24	2	20	2	21	13	8	2	20	21

Shift Cipher encryption example

It was at least certain that ...

itwasatleastcertainthat

t	i	t	w	a	s	a	t	l	e	a	s	t
v	8	19	22	0	18	0	19	11	4	0	18	19
k	c	c	c	c	c	c	c	c	c	c	c	c
$+$	2	2	2	2	2	2	2	2	2	2	2	2
v	10	21	24	2	20	2	21	13	8	2	20	21
t	k	v	y	c	u	c	v	n	g	c	u	v

kvycucvngcuvegtvckpvjcv

Shift Cipher encryption more examples

It was at least certain that ...

itwasatleastcertainthat

t	i	t	w	a	s	a	t	l	e	a	s	t
v	8	19	22	0	18	0	19	11	4	0	18	19
k	s	s	s	s	s	s	s	s	s	s	s	s
+	18	18	18	18	18	18	18	18	18	18	18	18
v	0	11	4	18	10	18	11	3	22	18	10	11
t	a	l	o	s	k	s	l	d	w	s	k	l

alosksldwskluwjlsaflzsl

Shift Cipher decryption

alosksldwskluwjlsaflzsl

t	a	l	o	s	k	s	l	d	w	s	k	l
v	0	11	4	18	10	18	11	3	22	18	10	11
k	s	s	s	s	s	s	s	s	s	s	s	s
–	18	18	18	18	18	18	18	18	18	18	18	18
v	8	19	22	0	18	0	19	11	4	0	18	19
t	i	t	w	a	s	a	t	l	e	a	s	t

itwasatleastcertainthat

Shift cipher cryptanalysis

brute force only 26 keys

statistics letter frequencies (total of 1699 letters)

Letter probabilities

letter	probability	letter	probability
A	0.082	N	0.067
B	0.015	O	0.075
C	0.028	P	0.019
D	0.043	Q	0.001
E	0.127	R	0.060
F	0.022	S	0.063
G	0.020	T	0.091
H	0.061	U	0.028
I	0.070	V	0.010
J	0.002	W	0.023
K	0.008	X	0.001
L	0.040	Y	0.020
M	0.024	Z	0.001

Letter statistics

1. e is most common with probability 0.12
2. t, a, o, i, n, s, h, r with probability 0.06 to 0.09
3. d, l with probability in 0.04
4. c, u, m, w, f, g, y, p, b, with probability 0.015 to 0.028
5. v, k, j, x, q, z, with probability less than 0.01

Plaintext letter occurrences

letter	occurrences	letter	occurrences
a	126	n	118
b	26	o	106
c	50	p	30
d	62	q	1
e	223	r	106
f	33	s	127
g	35	t	149
h	112	u	40
i	134	v	18
j	0	w	49
k	11	x	2
l	69	y	30
m	42	z	0

Plaintext letter probabilities

letter	probability	letter	probability
a	0.074	n	0.069
b	0.015	o	0.062
c	0.029	p	0.018
d	0.036	q	0.001
e	0.131	r	0.062
f	0.019	s	0.075
g	0.021	t	0.088
h	0.066	u	0.024
i	0.079	v	0.011
j	0.000	w	0.029
k	0.006	x	0.001
l	0.041	y	0.018
m	0.025	z	0.000

Ceaser letter occurrences

letter	occurrences	letter	occurrences
a	30	n	69
b	0	o	42
c	126	p	118
d	26	q	106
e	50	r	30
f	62	s	1
g	223	t	106
h	33	u	127
i	35	v	149
j	112	w	40
k	134	x	18
l	0	y	49
m	11	z	2

Vegenere Cipher

i	t	w	a	s	a	t	l	e	a	s	t	c	e	r
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
s	x	u	k	w	y	d	p	c	k	w	r	m	i	p
t	a	i	n	t	h	a	t	p	h	i	l	e	a	s
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
d	e	g	x	x	f	k	x	n	r	m	j	o	e	q
f	o	g	g	h	a	d	n	o	t	a	b	s	e	n
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
p	s	e	q	l	y	n	r	m	d	e	z	c	i	l
t	e	d	h	i	m	s	e	l	f	f	r	o	m	l
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
d	i	b	r	m	k	c	i	j	p	j	p	y	q	j

Vegenere letter occurrences

letter	occurrences	letter	occurrences
a	32	n	38
b	53	o	68
c	135	p	66
d	66	q	64
e	79	r	148
f	54	s	96
g	77	t	14
h	19	u	17
i	94	v	59
j	34	w	62
k	64	x	72
l	83	y	92
m	94	z	19

Vegenere Cipher

i	t	w	a	s	a	t	l	e	a	s	t	c	e	r
↓		↓		↓		↓		↓		↓		↓		↓
s	x	u	k	w	y	d	p	c	k	w	r	m	i	p
t	a	i	n	t	h	a	t	p	h	i	l	e	a	s
	↓		↓		↓		↓		↓		↓		↓	
d	e	g	x	x	f	k	x	n	r	m	j	o	e	q
f	o	g	g	h	a	d	n	o	t	a	b	s	e	n
↓		↓		↓		↓		↓		↓		↓		↓
p	s	e	q	l	y	n	r	m	d	e	z	c	i	l
t	e	d	h	i	m	s	e	l	f	f	r	o	m	l
	↓		↓		↓		↓		↓		↓		↓	
d	i	b	r	m	k	c	i	j	p	j	p	y	q	j

Vegenere letter occurrences skip 1

letter	occurrences	letter	occurrences
a	20	n	22
b	24	o	28
c	67	p	43
d	32	q	35
e	46	r	72
f	31	s	45
g	44	t	8
h	11	u	8
i	43	v	23
j	19	w	28
k	28	x	32
l	44	y	45
m	46	z	6

Vegenere Cipher

i	t	w	a	s	a	t	l	e	a	s	t	c	e	r
↓			↓			↓			↓			↓		
s	x	u	k	w	y	d	p	c	k	w	r	m	i	p
t	a	i	n	t	h	a	t	p	h	i	l	e	a	s
↓			↓			↓			↓			↓		
d	e	g	x	x	f	k	x	n	r	m	j	o	e	q
f	o	g	g	h	a	d	n	o	t	a	b	s	e	n
↓			↓			↓			↓			↓		
p	s	e	q	l	y	n	r	m	d	e	z	c	i	l
t	e	d	h	i	m	s	e	l	f	f	r	o	m	l
↓			↓			↓			↓			↓		
d	i	b	r	m	k	c	i	j	p	j	p	y	q	j

Vegenere letter occurrences skip 2

letter	occurrences	letter	occurrences
a	1	n	24
b	32	o	65
c	47	p	7
d	49	q	12
e	16	r	41
f	9	s	48
g	21	t	0
h	0	u	5
i	11	v	21
j	0	w	11
k	41	x	31
l	9	y	42
m	15	z	9

Vegenere Cipher

i	t	w	a	s	a	t	l	e	a	s	t	c	e	r
	$\downarrow$			$\downarrow$			$\downarrow$			$\downarrow$			$\downarrow$	
s	x	u	k	w	y	d	p	c	k	w	r	m	i	p
t	a	i	n	t	h	a	t	p	h	i	l	e	a	s
	$\downarrow$			$\downarrow$			$\downarrow$			$\downarrow$			$\downarrow$	
d	e	g	x	x	f	k	x	n	r	m	j	o	e	q
f	o	g	g	h	a	d	n	o	t	a	b	s	e	n
	$\downarrow$			$\downarrow$			$\downarrow$			$\downarrow$			$\downarrow$	
p	s	e	q	l	y	n	r	m	d	e	z	c	i	l
t	e	d	h	i	m	s	e	l	f	f	r	o	m	l
	$\downarrow$			$\downarrow$			$\downarrow$			$\downarrow$			$\downarrow$	
d	i	b	r	m	k	c	i	j	p	j	p	y	q	j

Vegenere letter occurrences skip 2

letter	occurrences	letter	occurrences
a	16	n	0
b	2	o	3
c	10	p	23
d	0	q	14
e	46	r	48
f	9	s	35
g	20	t	7
h	19	u	0
i	80	v	38
j	9	w	42
k	6	x	41
l	35	y	11
m	50	z	2

Digram and trigram statistics

digrams most common are
th, he, in, er, an, re, ed, on, es, st, en. . .

trigrams most common are
the, ing, and, her, ere, ent, tha, nth. . .

digram encryptions appear random

i	t	w	a	s	a	t	l	**e**	**a**	s	t	c	e	r
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
s	x	u	k	w	y	d	p	c	k	w	r	m	i	p
t	a	i	**n**	**t**	h	a	t	p	h	i	l	**e**	**a**	s
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
d	e	g	x	x	f	k	x	n	r	m	j	o	e	q
f	o	g	g	h	a	d	n	o	t	a	b	s	e	**n**
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
p	s	e	q	l	y	n	r	m	d	e	z	c	i	l
t	e	d	h	i	m	s	e	l	f	f	r	o	m	l
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
d	i	b	r	m	k	c	i	j	p	j	p	y	q	j

common can collide with rare

o	n	d	o	n	f	o	r	m	a	n	y	y	e	a
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
y	r	b	y	r	d	y	v	k	k	r	w	i	i	y
r	s	**t**	**h**	o	s	e	w	h	o	w	e	r	e	h
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
b	w	r	r	s	q	o	a	f	y	a	c	b	i	f
o	n	o	u	r	e	d	b	y	a	b	e	t	t	e
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
y	r	m	e	v	c	n	f	w	k	f	c	d	x	c
r	a	c	q	u	a	i	**n**	**t**	a	n	c	e	w	i
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
b	e	a	a	y	y	s	r	r	k	r	a	o	a	g

but mostly useful

t	**h**	h	i	m	**t**	**h**	a	n	**t**	**h**	e	r	e	s
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
d	l	f	s	q	r	r	e	l	d	l	c	b	i	q
t	d	e	c	l	a	r	e	d	**t**	**h**	a	t	n	o
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
d	h	c	m	p	y	b	i	b	d	l	y	d	r	m
b	o	d	y	c	o	u	l	d	p	r	e	t	e	n
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
l	s	b	i	g	m	e	p	b	z	v	c	d	i	l
d	t	o	h	a	v	e	e	v	e	r	s	e	e	n
K	E	Y	K	E	Y	K	E	Y	K	E	Y	K	E	Y
n	x	m	r	e	t	o	i	t	o	v	q	o	i	l

Finding a digram or a trigram

vrriqkocyjnvewsredlcqekoaycmlr
mqocccceayrrowrkwrbyeqpcgmrrebs
jdsgsvxwiirkqmdmmxpccwsxackvws
recxpekeviayreorgkprylgcxycxcc
tfspckwdykegeqxsrurmgrrylyfics

Index of coincidence

Definition (Index of coincidence)

Suppose $\vec{x} = x_1 x_2 x_3 \ldots x_n$ is a string of length n over an alphabet of size a. The *index of coincidence* is defined the probability that two random elements of $\vec{x}$ are equal.

English $a = 26$. Say frequencies are $f_0, \ldots, f_{25}$ then

$$IC(\vec{x}) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)}$$

Letter probabilities

letter	probability	letter	probability
A	0.082	N	0.067
B	0.015	O	0.075
C	0.028	P	0.019
D	0.043	Q	0.001
E	0.127	R	0.060
F	0.022	S	0.063
G	0.020	T	0.091
H	0.061	U	0.028
I	0.070	V	0.010
J	0.002	W	0.023
K	0.008	X	0.001
L	0.040	Y	0.020
M	0.024	Z	0.001

Letter probabilities

For random English text

$$IC(eng) = \sum_{i=0}^{25} p_i^2 \approx 0.065$$

$$IC(rand) = \sum_{i=0}^{25} \left(\frac{1}{26}\right)^2 \approx 0.038$$

recall Vegenere letter occurrences

letter	occurrences	letter	occurrences
a	32	n	38
b	53	o	68
c	135	p	66
d	66	q	64
e	79	r	148
f	54	s	96
g	77	t	14
h	19	u	17
i	94	v	59
j	34	w	62
k	64	x	72
l	83	y	92
m	94	z	19

index of coincidence

$$IC(\vec{x}) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)} \approx 0.048$$

frequencies $mod2$

l	#	l	#
a	20	n	22
b	24	o	28
c	67	p	43
d	32	q	35
e	46	r	72
f	31	s	45
g	44	t	8
h	11	u	8
i	43	v	23
j	19	w	28
k	28	x	32
l	44	y	45
m	46	z	6

l	#	l	#
a	12	n	16
b	29	o	40
c	68	p	23
d	34	q	29
e	33	r	76
f	23	s	51
g	33	t	6
h	8	u	9
i	51	v	36
j	15	w	34
k	36	x	40
l	39	y	47
m	48	z	13

index of coincidence

$$IC(x_0 x_2 x_4 \dots) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} \approx 0.047$$

$$IC(x_1 x_3 x_5 \dots) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} \approx 0.048$$

frequencies $mod3$

l	#	l	#
a	1	n	24
b	32	o	65
c	47	p	7
d	49	q	12
e	16	r	41
f	9	s	48
g	21	t	0
h	0	u	5
i	11	v	21
j	0	w	11
k	41	x	31
l	9	y	42
m	15	z	9

l	#	l	#
a	16	n	0
b	2	o	3
c	10	p	23
d	0	q	14
e	46	r	48
f	9	s	35
g	20	t	7
h	19	u	0
i	80	v	38
j	9	w	42
k	6	x	41
l	35	y	11
m	50	z	2

l	#	l	#
a	15	n	14
b	19	o	0
c	78	p	36
d	17	q	38
e	17	r	59
f	36	s	13
g	36	t	7
h	0	u	12
i	3	v	0
j	25	w	9
k	17	x	0
l	39	y	39
m	29	z	8

index of coincidence

$$
\begin{aligned}
IC(x_0x_3x_6\ldots) &= \frac{\sum_{i=0}^{25}\binom{f_i}{2}}{\binom{n}{2}} \approx 0.064 \\
IC(x_1x_4x_7\ldots) &= \frac{\sum_{i=0}^{25}\binom{f_i}{2}}{\binom{n}{2}} \approx 0.070 \\
IC(x_2x_5x_8\ldots) &= \frac{\sum_{i=0}^{25}\binom{f_i}{2}}{\binom{n}{2}} \approx 0.066
\end{aligned}
$$

index of coincidence

$$IC(\vec{x}) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)}$$

to

$$IC(\vec{x}) = \sum_{i=0}^{25} p_i \frac{f_{i+j}}{n}$$

1st keyword letter

0.039	0.034	0.035	0.038	0.036
0.033	0.043	0.033	0.03	0.04
0.065	0.039	0.031	0.037	0.047
0.033	0.035	0.039	0.034	0.032
0.039	0.044	0.037	0.04	0.043
0.047				

2nd keyword letter

0.045	0.034	0.032	0.04	0.068
0.038	0.029	0.033	0.048	0.034
0.034	0.037	0.034	0.037	0.038
0.044	0.037	0.043	0.039	0.044
0.038	0.035	0.033	0.038	0.038
0.032				

3rd keyword letter

0.033	0.034	0.046	0.034	0.035
0.037	0.034	0.033	0.035	0.046
0.039	0.044	0.038	0.046	0.037
0.035	0.034	0.04	0.035	0.031
0.041	0.034	0.033	0.04	0.066
0.04				

index of coincidence

f_i indices $0 \bmod 3$ g_i indices $1 \bmod 3$ h_i indices $2 \bmod 3$

$$IC(\vec{x}) = \sum_{i=0}^{25} \frac{f_i}{n} \frac{g_{i+j}}{n}$$

$$IC(\vec{x}) = \sum_{i=0}^{25} \frac{f_i}{n} \frac{h_{i+j}}{n}$$

1-2rd keyword letter difference

0.035	0.036	0.036	0.036	0.039
0.046	0.041	0.04	0.036	0.043
0.039	0.033	0.033	0.038	0.037
0.033	0.048	0.037	0.031	0.039
0.067	0.038	0.027	0.032	0.047
0.034				

1-3rd keyword letter difference

0.044	0.042	0.036	0.045	0.038
0.033	0.034	0.04	0.034	0.031
0.044	0.037	0.033	0.04	0.065
0.041	0.031	0.033	0.045	0.034
0.034	0.036	0.035	0.033	0.036
0.048				