CS641

Modern Cryptology

Lecture 2

Frequency Analysis

- Used to break a substitution cipher.
- Key observation is that letters in English alphabet have uneven distribution in normal English text.
- Use this to recover unknown substitution.
- Frequency distribution table for English text:

Letter	E	T	A	O	I	N	S	H	R
Frequency	12.7	9.1	8.2	7.5	7.0	6.7	6.3	6.1	6.0
Letter	D	L	U	C	M	W	F	Y	G
Frequency	4.3	4.0	2.8	2.8	2.4	2.4	2.2	2.0	2.0
Letter	P	B	V	K	X	J	Q	Z	
Frequency	1.9	1.5	1.0	0.8	0.2	0.2	0.1	0.1	

Frequency Analysis

Encrypted Text

pzipt vwdrviwi jwff xytz apaxwwz xyroftzi dtfwf ra drvrztspvof ipfwtfw arv xgr itbf pz t vrg tf arovxwwz xyroftzi xgr tzi tzi apaxb fpc zwg pzawdxprzf gwvw vwervxwi pz xyw jtfx xgwzxb arov yrovf, tddrvipzu xr xyw ozprz ywtjxy qpzpfxvb rz ftxovitb qrvzpzu. xywvw gwvw rzw yozivwi tzi apaxb xgr iwtxyf pz xyw ftqw ewvpri tdvrff xyw drozxvb tzi xyw atxtjpxpwf ytsw vwqtpzwi hwjrg xgr yozivwi arv xyw jtfx xwz itbf, ywtjxy qpzpfxvb'f itxt fyrgwi. xyw drozxvb, gypdy pf xyw fwdrzi grvfx-taawdxwi pz xyw grvji, ytf fwwz qrvw xytz rzw dvrvw dtfwf fr atv.

Frequency Analysis

- Total alphabet: 445
- Frequency distribution table:

Letter	w	x	z	r	t	f	v	i	y
Frequency	12.1	9.7	8.5	8.3	8.1	7.9	7.4	6.1	6.1

Letter	p	a	d	g	o	b	j	q	u
Frequency	6.1	3.6	3.2	3.2	2.7	2.3	1.8	1.4	0.5

Letter	e	s	h	c	l	n	k	m	
Frequency	0.5	0.5	0.2	0.2	0.0	0.0	0.0	0.0	

Frequency Analysis

Decrypted Text

pnipt vedrviei jeff tytn apateen tyroftni dtfef ra drvrntspvof ipfetfe arv tgr itbf pn t vrg tf arovteen tyroftni tgr tni tni apatb fpc neg pnaedtprnf geve veervtei pn tye jtft tgentb arov yrovf, tddrvipnu tr tye onprn yetjty qpnpftvb rn fttovitb qrvnpnu. Tyeve geve rne yonivei tni apatb tgr iettyf pn tye ftqe eevpri tdvrff tye drontvb tni tye atttjptpef ytse veqtpnei hejrg tgr yonivei arv tye jtft ten itbf, yetjty qpnpftvb'f ittt fyrgei. Tye drontvb, gypdy pf tye fedrni grvft-taaedtei pn tye grvji, ytf feen qrve tytn rne dvrve dtfef fr atv.

Letter	e	t	n	r	t	f	v	i	y
Frequency	12.1	9.7	8.5	8.3	8.1	7.9	7.4	6.1	6.1
Letter	p	a	d	g	o	b	j	q	u
Frequency	6.1	3.6	3.2	3.2	2.7	2.3	1.8	1.4	0.5
Letter	e	s	h	c	l	n	k	m	
Frequency	0.5	0.5	0.2	0.2	0.0	0.0	0.0	0.0	

Frequency Analysis

Decrypted Text

pnipa vedrviei jeff tyan apateen tyrofani dafef ra drvrnaspvof ipfeafe arv tgr iabf pn a vrg af arovteen tyrofani tgr ani ani apatb fpc neg pnaedtprnf geve veervtei pn tye jaft tgentb arov yrovf, addrvipnu tr tye onprn yeajty qpnpftvb rn fatoviab qrvnpnu. Tyeve geve rne yonivei ani apatb tgr ieatyf pn tye faqe eevpri advrff tye drontvb ani tye aatajptpef yase veqapnei hejrg tgr yonivei arv tye jaft ten iabf, yeajty qpnpftvb'f iata fyrgei. Tye drontvb, gypdy pf tye fedrni grvft-aaaedtei pn tye grvji, yaf feen qrve tyan rne dvrve dafef fr aav.

Letter	e	t	n	r	a	f	v	i	y
Frequency	12.1	9.7	8.5	8.3	8.1	7.9	7.4	6.1	6.1

Letter	p	a	d	g	o	b	j	q	u
Frequency	6.1	3.6	3.2	3.2	2.7	2.3	1.8	1.4	0.5

Letter	e	s	h	c	l	n	k	m	
Frequency	0.5	0.5	0.2	0.2	0.0	0.0	0.0	0.0	

Frequency Analysis

Decrypted Text

Iniia vedrviei jeff tyan aiateen tyrofani dafef ra drvrnasivof iifeafe arv tgr iabf in a vrg af arovteen tyrofani tgr ani ani aiatb fic neg inaedtirnf geve veervtei in tye jaft tgentb arov yrovf, addrviinu tr tye onirn yeajty qiniftvb rn fatoviab qrvninu. Tyeve geve rne yonivei ani aiatb tgr ieatyf in tye faqe eeviri advrff tye drontvb ani tye aatajitief yase veqainei hejrg tgr yonivei arv tye jaft ten iabf, yeajty qiniftvb'f iata fyrgei. Tye drontvb, gyidy if tye fedrni grvft-aaaedtei in tye grvji, yaf feen qrve tyan rne dvrve dafef fr aav.

Letter	e	t	n	r	a	f	v	i	y
Frequency	12.1	9.7	8.5	8.3	8.1	7.9	7.4	6.1	6.1

Letter	i	a	d	g	o	b	j	q	u
Frequency	6.1	3.6	3.2	3.2	2.7	2.3	1.8	1.4	0.5

Letter	e	s	h	c	l	n	k	m	
Frequency	0.5	0.5	0.2	0.2	0.0	0.0	0.0	0.0	

Frequency Analysis

Decrypted Text

Iniia redoriei jeff tyan fifteen tyoufani dafef of doronasiruf iifeafe for tgo iabf in a rog af fourteen tyoufani tgo ani ani fiftb fic neg infedtionf gere reeortei in tye jaft tgentb four yourf, addoriinu to tye Union yeajty qiniftrb on faturiab qorninu. Tyere gere one yunirei ani fiftb tgo ieatyf in tye faqe eerioi adroff tye dountrb ani tye fatajitief yase reqainei hejog tgo yunirei for tye jaft ten iabf, yeajty qiniftrb'f iata fyogei. Tye dountrb, gyidy if tye fedoni gorft-affedtei in tye gorji, yaf feen qore tyan one drore dafef fo far.

Letter	e	t	n	o	a	f	r	i	y	i	f	d	g	u	b	j	q	u	e	s	h	c	l	n	k	m	
Frequency	12.1	9.7	8.5	8.3	8.1	7.9	7.4	6.1	6.1	6.1	3.6	3.2	3.2	2.7	2.3	1.8	1.4	0.5	0.5	0.5	0.2	0.2	0.0	0.0	0.0	0.0	

Frequency Analysis

Decrypted Text

India redorded jess than fifteen thousand dases of doronasirus disease for tgo dabs in a rog as fourteen thousand tgo and and fiftb sic neg infedtions gere reeorted in the jast tgentb four hours, addordinu to the Union heajth qinistrb on Saturdab qorninu. There gere one hundred and fiftb tgo deaths in the saqe eeriod adross the dountrb and the fatajities hase reqained hejog tgo hundred for the jast ten dabs, heajth qinistrb's data shoged. The dountrb, ghidh is the sedond gorst-affedted in the gorjd, has seen qore than one drore dases so far.

Letter	e	t	n	o	a	s	r	d	h
Frequency	12.1	9.7	8.5	8.3	8.1	7.9	7.4	6.1	6.1
Letter	i	f	d	g	u	b	j	q	u
Frequency	6.1	3.6	3.2	3.2	2.7	2.3	1.8	1.4	0.5
Letter	e	s	h	c	l	n	k	m	
Frequency	0.5	0.5	0.2	0.2	0.0	0.0	0.0	0.0	

Frequency Analysis

Decrypted Text

India recorded less than fifteen thousand cases of coronasirus disease for two days in a row as fourteen thousand two and and fifty sic new infections were reeorted in the last twenty four hours, accordinu to the Union health qinistry on Saturday qorninu. There were one hundred and fifty two deaths in the saqe eeriod across the country and the fatalities hase reqained helow two hundred for the last ten days, health qinistry's data showed. The country, which is the second worst-affected in the world, has seen qore than one crore cases so far.

Letter	e	t	n	o	a	s	r	d	h
Frequency	12.1	9.7	8.5	8.3	8.1	7.9	7.4	6.1	6.1

Letter	i	f	c	w	u	y	l	q	u
Frequency	6.1	3.6	3.2	3.2	2.7	2.3	1.8	1.4	0.5

Letter	e	s	h	c	l	n	k	m	
Frequency	0.5	0.5	0.2	0.2	0.0	0.0	0.0	0.0	

Frequency Analysis

Decrypted Text

India recorded less than fifteen thousand cases of coronavirus disease for two days in a row as fourteen thousand two and and fifty six new infections were reported in the last twenty four hours, according to the Union health ministry on Saturday morning. There were one hundred and fifty two deaths in the same period across the country and the fatalities have remained below two hundred for the last ten days, health ministry's data showed. The country, which is the second worst-affected in the world, has seen more than one crore cases so far.

Letter	e	t	n	o	a	s	r	d	h
Frequency	12.1	9.7	8.5	8.3	8.1	7.9	7.4	6.1	6.1
Letter	i	f	c	w	u	y	l	m	g
Frequency	6.1	3.6	3.2	3.2	2.7	2.3	1.8	1.4	0.5
Letter	p	v	b	x	l	n	k	m	
Frequency	0.5	0.5	0.2	0.2	0.0	0.0	0.0	0.0	

Permutation Cipher

$k_E : [1, b] \mapsto [1, b]$, k_E a permutation

$$E(\ell_1 \ell_2 \cdots \ell_{bn}) =$$
$$\ell_{k_E(1)} \ell_{k_E(2)} \cdots \ell_{k_E(b)} \ell_{b+k_E(1)} \cdots \ell_{b+k_E(b)} \cdots \ell_{b(n-1)+k_E(1)} \cdots \ell_{b(n-1)+k_E(b)}$$

$k_D = k_E^{-1}$ and $D = E$.

- Operates on a block of text of b letters
- Number of possible keys equals $b!$
- Can be broken using knowledge about occurrence of letter pairs

Enigma Code

- Intensively used by Germans during WW II
- Uses a different substitution cipher for every letter according to a policy decided by encryption key
- Complex sequence of operations were hardwired in the Enigma machine
- The machine was given highest security priority, so that the algorithm used could not leak out
- Eventually it did, and this allowed a group in UK to break the code

Lessons

#1

Everything that is fixed will eventually be known to Ela. So assume a-priory that encryption and decryption algorithms are known.

#2

A simple combination of substitution and permutation ciphers does not provide sufficient security.

#3

A thorough analysis is required to understand what kind of encryption algorithms are strong.