SPCS Cryptography Class Lecture 1
Introduction

Ho Chung Siu

Stanford University

June 22, 2015

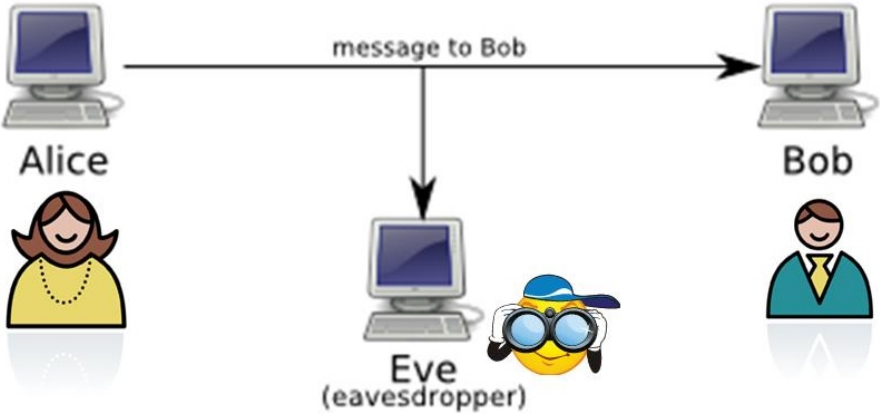
Course Logistics

- Textbook: *Cryptography*, by Simon-Rubinstein Salzedo
- Course webpage:
  `http://web.stanford.edu/~soarer/cryptography.html`
- Problem sets
- Presentations on last two days: Jul 8,9.

Setup of Cryptography

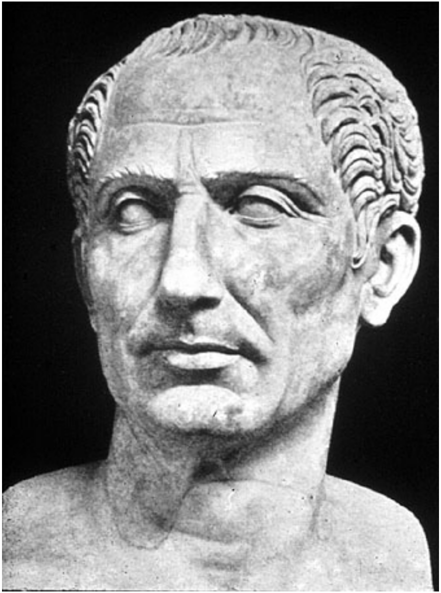
Cryptography

Setup of Cryptography



Cryptanalysis

Caesar cipher



- Left shift by 3 alphabets.
- For example, $d \to A$, $e \to B$.
- How would

  `cryptography`

  be encrypted?

  `ZOVMQLDOXMEV`
- How to decrypt a message?
- For example,

  `OFDEQ`

Shift cipher

- We can also do shifts other than -3 alphabets.
- For example, with a right shift of 10,

  `sleep`

  becomes

  `BUNNY`

Shift cipher

- Let's say Alice and Bob agrees on a key (the shift). If Eve does not know the key, what can she do to crack a Caesar cipher?
- For example,

  `JRNGURE`

  `http://www.dcode.fr/caesar-cipher`
- How can one make this harder to crack?

Substitution ciphers

- Suppose Alice and Bob agrees on a substitution scheme instead - where every letter would stand for another (perhaps random) letter.
- For example, if we have an encryption scheme

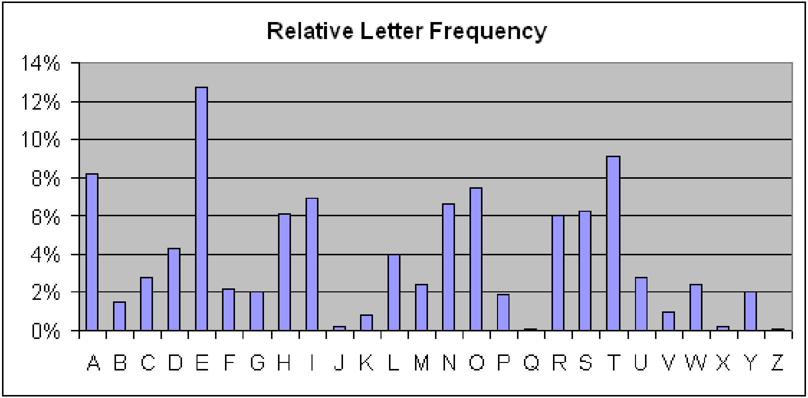
a	b	c	d	e	f	g	h	i	j	k	l	m
C	I	S	Q	V	N	F	O	W	A	X	M	T
n	o	p	q	r	s	t	u	v	w	x	y	z
G	U	H	P	B	K	L	R	E	Y	D	Z	J

- How would Alice send Bob the message

`weather`

- How does Bob decipher?
- What can Eve do? Frequency Analysis.

- Normal English has certain letter frequency. For example, *e* appears way more often than *x*.



- Same for bigrams (pair of letters). For example,

Bigrams Frequency			
th	1.52 %	en	0.55 %
he	1.28 %	ed	0.53 %
in	0.94 %	to	0.52 %
er	0.94 %	it	0.50 %
an	0.82 %	ou	0.50 %
re	0.68 %	ea	0.47 %
nd	0.63 %	hi	0.46 %
at	0.59 %	is	0.46 %
on	0.57 %	or	0.43 %
nd	0.56 %	ti	0.34 %
ha	0.56 %	as	0.33 %
es	0.56 %	te	0.27 %
st	0.55 %	et	0.19 %

- Substitution ciphers do NOT change this profile!

Substitution ciphers

Example

We wish to decrypt

```
JNRZR BNIGI BJRGZ IZLQR OTDNJ GRIHT USDKR
ZZWLG OIBTM NRGJN IJTZJ LZISJ NRSBL QVRSI
ORIQT QDEKJ JNRQW GLOFN IJTZX QLFQL WBIMJ
ITQXT HHTBL KUHQL JZKMM LZRNT OBIMI EURLW
BLQZJ GKBJT QDIQS LWJNR OLGRI EZJGK ZRBGS
MJLDG IMNZT OIHRK MOSOT QHIJL QBRJN IJJNT
ZFIZL WIZTO MURZM RBTRZ ZKBNN LFRVR GIZFL
KUHIM MRIGJ LJNRB GKHRT QJRUU RBJLW JNRZI
TULGI EZLUK JRUST QZLUK EURFT JNLKJ JNRXR S
```

http://scottbryce.com/cryptograms/index.htm

Substitution ciphers

Solution

- *The most common English letter is E, T and A.*
- *The most common English bigram is TH, HE, IN.*
- *H is in the two most frequent English bigrams! This would give us a hint about what letters are T,H,E.*

Substitution ciphers

Here is a frequency table of the text

	R	J	I	L	Z	T	N	Q	B	G	K
Freq	33	30	27	25	24	20	19	16	15	15	13
	U	M	O	S	H	W	F	E	D	X	V
Freq	12	12	10	9	8	7	6	5	5	3	2

The most frequent bigrams are: JN (11 times), NR (8 times), TQ (6 times), and LW, RB, RZ, and JL (5 times each).

What should JN,NR be?

Substitution ciphers

Half way through..

```
theZe BhIGI BteGZ IZLQe OTDht GeIHT USDKe
ZZWLG OIBTM heGth ItTZt LZISt heSBL QVeSI
OeIQT QDEKt theQW GLOFh ItTZX QLFQL WBIMt
ITQXT HHTBL KUHQL tZKMM LZehT OBIMI EUeLW
BLQZt GKBtT QDIQS LWthe OLGeI EZtGK ZeBGS
MtLDG IMhZT OIHeK MOSOT QHItL QBeth ItthT
ZFIZL WIZTO MUeZM eBTeZ ZKBhh LFeVe GIZFL
KUHIM MeIGt LtheB GKHeT QteUU eBtLW theZI
TULGI EZLUK teUST QZLUK EUeFT thLKt theXe S
```

Substitution ciphers

	R	J	I	L	Z	T	N	Q	B	G	K
Freq	33	30	27	25	24	20	19	16	15	15	13
	U	M	O	S	H	W	F	E	D	X	V
Freq	12	12	10	9	8	7	6	5	5	3	2

Frequent bigrams are: JN (11 times), NR (8 times), TQ (6 times), and LW, RB, RZ, and JL (5 times each).

- Next frequent characters we don't know are I,L,Z,T.
- *RZ* appears often, and there is theZe in the beginning. This suggests that *Z* should be either *r* or *s*.
- *TQ* appears 6 times, and *LQ* appears 4 times, and they are among the characters we don't know yet. Looking at the bigram table,

Substitution ciphers

Bigrams Frequency			
th	1.52 %	en	0.55 %
he	1.28 %	ed	0.53 %
in	0.94 %	to	0.52 %
er	0.94 %	it	0.50 %
an	0.82 %	ou	0.50 %
re	0.68 %	ea	0.47 %
nd	0.63 %	hi	0.46 %
at	0.59 %	is	0.46 %
on	0.57 %	or	0.43 %
nd	0.56 %	ti	0.34 %
ha	0.56 %	as	0.33 %
es	0.56 %	te	0.27 %
st	0.55 %	et	0.19 %

- This suggests that they are an, in or on. In particular, Q should be n, and T, L would be among $\{a, i, o\}$.
- $IJ = It$ appears 4 times as well - this would suggest that I is likely a, although it may also be s.

Substitution ciphers

So far we know,

$$J = t,\ N = h,\ R = e,\ Q = n,\ I = a/s,\ T,L = a/i/o,\ Z = r/s.$$

Back to the ciphertext for a moment,

```
theZe BhIGI BteGZ IZLQe OTDht GeIHT USDKe
ZZWLG OIBTM heGth ItTZt LZISt heSBL QVeSI
OeIQT QDEKt theQW GLOFh ItTZX QLFQL WBIMt
ITQXT HHTBL KUHQL tZKMM LZehT OBIMI EUeLW
BLQZt GKBtT QDIQS LWthe OLGeI EZtGK ZeBGS
MtLDG IMhZT OIHeK MOSOT QHItL QBeth IttHT
ZFIZL WIZTO MUeZM eBTeZ ZKBhh LFeVe GIZFL
KUHIM MeIGt LtheB GKHeT QteUU eBtLW theZI
TULGI EZLUK teUST QZLUK EUeFT thLKt theXe S
```

Substitution ciphers

This suggests,

$$J = t,\ N = h,\ R = e,\ Q = n,\ I = a,\ T = i,\ L = o,\ Z = s.$$

Putting them in,

```
these BhaGa BteGs asone OiDht GeaHi USDKe
ssWoG OaBiM heGth atist osaSt heSBo nVeSa
Oeani nDEKt thenW GoOFh atisX noFno WBaMt
ainXi HHiBo KUHno tsKMM osehi OBaMa EUeoW
Bonst GKBti nDanS oWthe OoGea EstGK seBGS
MtoDG aMhsi OaHeK MOSOi nHato nBeth atthi
sFaso WasiO MUesM eBies sKBhh oFeVe GasFo
KUHaM MeaGt otheB GKHei nteUU eBtoW thesa
iUoGa EsoUK teUSi nsoUK EUeFi thoKt theXe S
```

First line almost decrypted!

Substitution ciphers

- The beginning of message should look like

  `these BhaGaBteGs as one OiDht...`

- *Bha* is probably start of a word, and there is also an *aBt* part. One is led to guess that $B = c$, which leads to $G = r$.
- Moreover, *OiDht* strongly hints that D is g.

```
these chara cters asone Oight reaHi USgKe
ssWor OaciM her that is to sa S theSco nVeSa
Oeani ngEKt thenW roOFh atisX noFno WcaMt
ainXi HHico KUHno tsKMM osehi OcaMa EUeoW
const rKcti nganS oWthe Oorea EstrK secrS
Mtogr aMhsi OaHeK MOSOi nHato nceth atthi
sFaso WasiO MUesM ecies sKchh oFeVe rasFo
KUHaM Meart othec rKHei nteUU ectoW thesa
iUora EsoUK teUSi nsoUK EUeFi thoKt theXe S
```

Substitution ciphers

- One can keep going - for example, Oight suggests that $O = m$.
- The deciphered text is,

```
These characters, as one may readily guess, form a cipher;
that is to say, they convey a meaning.  But then from what
is known of Captain Kidd, I could not suppose him capable
of constructing any of the more abtruse cryptographs.  I
made up my mind at once that this was of a simple species.
Such however as would appear to the crude intellect of the
sailor absolutely insoluble without the key.
```

What next?

- So substitution cipher is cracked. What next?
- The reason substitution cipher is prone to frequency analysis, is because it is *monoalphabetic*, i.e. same letter always gets encrypted to the same thing.
- We can make it *polyalphabetic*!
- For example, g may get encoded by x at some point, but by m later in the message.

Vigenere cipher

Vigenere cipher is an example of a polyalphabetic cipher.

Example

Encrypt `cryptography` using the key `lemon`.

Solution

Plaintext	*cryptography*
Key	*lemonlemonle*
Ciphertext	*NVKDGZKDOCSC*

French calls it *le chiffre indéchiffrable*.

What can we do to crack this cipher?

Terminologies

Some terminologies: Alice wants to send Bob a message.

- The unencrypted message is the *plaintext*. The encrypted message is the *ciphertext*. Alice and Bob are doing *cryptography* in this process.
- A cryptosystem consists of a pair of algorithms, encryption (plaintext to ciphertext), and decryption (ciphertext to plaintext).
- For the encryption part, there are two components - the *protocol* for encoding the message, and a specific *key*.

Kerckhoff's principle

A good cryptosystem should remain secure even if Eve knows the protocol, but not the key.

- What Eve does is *cryptanalysis*, the analysis of encrypted messages.