CS458: Introduction to Information Security

Notes 3: Historical Crypto - Part II

Yousef M. Elmehdwi

Department of Computer Science

Illinois Institute of Technology

yelmehdwi@iit.edu

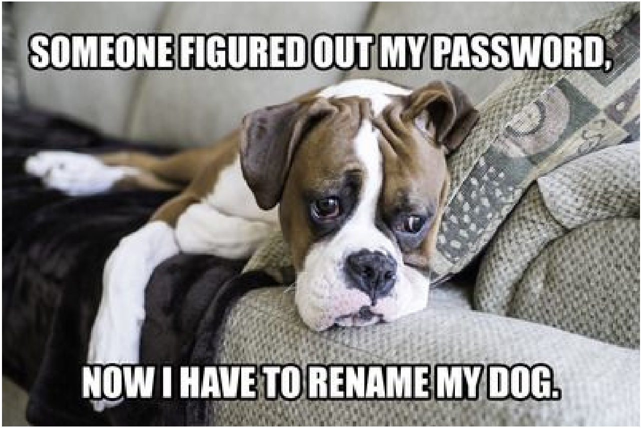
September 6, 2018

Slides: Modified from Michael J. Fischer, Ewa Syta

Outline

- Crypto Continued
- Modern Crypto – Next week

- Crypto Continued



Source: dogtime.com

Cryptanalysis: Terminology

- Cryptosystem is secure if best know attack is to try all keys.
  - Exhaustive key search, that is.
- Cryptosystem is insecure if any shortcut attack is known.
- Q: Are there any completely secure ciphers?

Terminology

- A shift cipher uses a letter substitution defined by a rotation of the alphabet.
- Any cipher that uses a substitution to replace a plaintext letter by a ciphertext letter is called a substitution cipher.
  - A shift cipher is a special case of a substitution cipher.
- Any cipher that encrypts a message by applying the same substitution to each letter of the message is called a monoalphabetic cipher.

Polyalphabetic ciphers

- Another way to strengthen substitution ciphers is to use different substitutions for different letter positions.
    - Choose $\mathtt{r}$ different alphabet permutations $\pi_1, \ldots, \pi_{\mathtt{r}}$ for some number $\mathtt{r}$.
    - Use π_1 for the first letter of $\mathtt{m}$, π_2 for the second letter, etc.
    - Repeat this sequence after every $\mathtt{r}$ letters.
- While this is much harder to break than monoalphabetic ciphers, letter frequency analysis can still be used.
- Every $\mathtt{r}^{\mathtt{th}}$ letter is encrypted using the same permutation, so the submessage consisting of just those letters still exhibits normal English language letter frequencies

Vigènere Cipher

- The Vigenère cipher is a polyalphabetic cipher in which the number of different substitutions $\mathtt{r}$ is also part of the key.
- Thus, the adversary must determine $\mathtt{r}$ as well as discover the different substitutions.
- All polyalphabetic ciphers can be broken using letter frequency analysis, but they are secure enough against manual attacks to have been used at various times in the past.
- The German Enigma encryption machine used in the second world war is also based on a polyalphabetic cipher.

Vigènere Cipher

- Like Caesar cipher, but use a phrase as key
- Example
  - `Message`: THE BOY HAS THE BALL
  - `Key`: VIG
  - `Encipher`: using Caesar cipher for each letter:

```
key    VIGVIGVIGVIGVIGV
plain  THEBOYHASTHEBALL
cipher OPKWWECIYOPKWIRG
```

The Vigènere Tableau

	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
A	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
B	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	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	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

Vigènere Cipher

	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
A	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
B	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	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	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

- Example
- key V, letter T: follow V column down to T row (giving “O”)
- key I, letter H: follow I column down to H row (giving “P”)

```
key    VIGVIGVIGVIGVIGV
plain  THEBOYHASTHEBALL
cipher OPKWWECIYOPKWIRG
```

Vigènere Example: Another way

```
Plaintext :  THEBOYHASTHEBALL
Key       :  VIGVIGVIGVIGVIGV
```

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

A	B	C	D	E	F	G	H	I	J	K	L	M
26	27	28	29	30	31	32	33	34	35	36	37	38
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
39	40	41	42	43	44	45	46	47	48	49	50	51

- key V, letter T: T is 19, and V is 21, we add the numbers, 40 which means O is the ciphertext.
- key I, letter H: H is 7, and I is 8, we add the numbers, 15 which means P is the ciphertext.
- etc

Useful Terms

- period: length of key
  - In earlier example, period is 3
- tableau: table used to encipher and decipher
  - Vigènere cipher has key letters on top, plaintext letters on the left
- polyalphabetic: the key has several different letters
  - Caesar cipher is mono-alphabetic

Attacking the Cipher

- Approach
  - Establish period (find the keylength); call it r
  - Break message into r parts, each part being enciphered using the same key letter
  - Solve each part
- We will show each step

Index of Coincidence

- Suppose you have an encrypted message:

FMMXF	NMDHO	DQDOR	ODSKV	KAERR	YFEKH
VSRVU	TADLA	KARIR	MXFRD	SAFID	KKARF
BNDKF	SDFSV	KAREK	AVHRT	AVDSK	ARDEV
TSMFS	XNFXR	FERLF	MMROL	RMKHD	SVNEX
FNMHK	ARKAD	EOFMM	KARHR	ODUUR	EUEVP
RFLAV	KARED	SMFSX	NFXRL	NHKVP	HFSOM
FTHKA	REDQR	EXFEV	SSRHR	YFEFK	RHKAR
XFNMH	UEVPK	ARFBN	DKFSD	KARPF	ESRFS
OKARH	RDSRH	RYFEF	KRKAR	PUEVP	KARIR

- We can use the index of coincidence (IC) to determine what type of cipher was used.

Coincidence

- Suppose we have a text,

  `"Four score and seven years ago our fathers brought forth, on this continent, a new nation, ..."`

- If we pick two letters from the text at random, most of the time the letters will be different, but sometimes they will be the same
- In a typical English text, about 6.8% of the randomly chosen pairs will consist of identical letters, while a "text" of randomly chosen letters will have IC as low as 3.8%.
- This feature is preserved by a substitution cipher, so if a ciphertext has a high IC, we might conclude it was encrypted using a substitution cipher.

Index of Coincidence

- Suppose a particular letter appears k times among N letters.
- there are $\binom{N}{2} = \frac{N(N-1)}{2}$ ways we can pick two letters at random, and $\binom{k}{2} = \frac{k(k-1)}{2}$ ways we can pick the designated letter,
- So, the probability that both letters we pick are the designated letter will be

$$\frac{\frac{k(k-1)}{2}}{\frac{N(N-1)}{2}} = \frac{k(k-1)}{N(N-1)}$$

- It follows that the IC will be found by

$$IC = \frac{\sum k_i(k_i-1)}{N(N-1)}$$ where k_i is the number of times the i^{th} symbol appears.

index of Coincidence: Example

- Suppose you have ciphertext:

FMMXF	NMDHO	DQDOR	ODSKV	KAERR	YFEKH
VSRVU	TADLA	KARIR	MXFRD	SAFID	KKARF
BNDKF	SDFSV	KAREK	AVHRT	AVDSK	ARDEV
TSMFS	XNFXR	FERLF	MMROL	RMKHD	SVNEX
FNMHK	ARKAD	EOFMM	KARHR	ODUUR	EUEVP
RFLAV	KARED	SMFSX	NFXRL	NHKVP	HFSOM
FTHKA	REDQR	EXFEV	SSRHR	YFEFK	RHKAR
XFNMH	UEVPK	ARFBN	DKFSD	KARPF	ESRFS
OKARH	RDSRH	RYFEF	KRKAR	PUEVP	KARIR

- We count the occurrences of letters:

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

index of Coincidence: Example

- We can use the frequency to compute the IC

A	B	C	D	E	F	G	H	I	J	K	L	M
22	2	0	19	18	29	0	14	3	0	26	5	14

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
9	8	6	2	39	18	4	6	14	0	9	3	0

$$IC = \frac{22(21) + 2(1) + 19)(18) + ...}{270(269)} \approx 0.0718$$

- This is a relatively high IC, so we might conclude this ciphertext was produced using a substitution cipher

Mathematical cryptography

- Mathematical cryptography began when Friedrich Kasiski published a method of breaking Vigènere ciphers in 1863
- The fundamental weakness of Vigènere ciphers that if that keylength is known, the ciphertext can be split apart into individual shift ciphers
- So security relies on having the keylength unknown
- Kasiski: **repetitions** in the ciphertext occur when characters of the key appear over the same characters in the plaintext

- Kaskski: **repetitions** in the ciphertext occur when characters of the key appear over the same characters in the plaintext
  - The number of characters between the repetitions is a multiple of the period.
- Example

```
key    VIGVIGVIGVIGVIGV
plain  THEBOYHASTHEBALL
cipher OPKWWECIYOPKWIRG
```

- Note the key and plaintext line up over the repetitions (underlined). As distance between repetitions is 9, the period is a factor of 9 (which is 3 here)

Finding the Keylength

- Kasiski's insight was the following:
  - There are common bigrams and trigrams in the plaintext: `TH, MM, RE`
  - From times to times, two occurrences of a bigram/trigram will be separated by an exact multiple of the keylength.
  - This means that the two occurrences will be encrypted in the same way.
- This suggests:
  - Find the common bigrams and trigrams in the ciphertext '
  - Find the distance between them
  - This distance may be a multiple of the keylength

Example

- Consider the ciphertext:

CDGAV	NNANX	DOKVZ	XDGVG	OBMXG	HVLOL
QFZIA	PJAXB	OAGTZ	FBTGA	IBVUK	LOBZT
SMDNV	GKSII	OAGJO	BJLGO	CIDGV	ZZCMH
YFGUI	ZWSBY	VKGUV	FGXFU	ZFOLK	FJOMZ
CMGMO	AGLCC	MWCDG	NCXUW	YCAYG	JALJF
GSXBJ	MJWMC	IECFB	OVZGU	PMOHO	KVHYE
CONNC	XTAQY	MZCJJ	H1XUW	KUMTV	WNNCX
ISPFN	YTGHN	CXCIP	COTPA	OBZFC	JIYVG
FLCYN	XKFZM	ZICJV	NZMJW	HZMHO	LCYWX

- Search the ciphertext for repeated bigrams and trigrams.

Example

- Consider the ciphertext:

CDGAV	NNANX	DOKVZ	XDGVG	OBMXG	HVLOL
QFZIA	PJAXB	OAGTZ	FBTGA	IBVUK	LOBZT
SMDNV	GKSII	OAGJO	BJLGO	CIDGV	ZZCMH
YFGUI	ZWSBY	VKGUV	FGXFU	ZFOLK	FJOMZ
CMGMO	AGLCC	MWCDG	NCXUW	YCAYG	JALJF
GSXBJ	MJWMC	IECFB	OVZGU	PMOHO	KVHYE
CONNC	XTAQY	MZCJJ	H1XUW	KUMTV	WNNCX
ISPFN	YTGHN	CXCIP	COTPA	OBZFC	JIYVG
FLCYN	XKFZM	ZICJV	NZMJW	HZMHO	LCYWX

  - The trigram `CDG` occurs in positions 1 and 133

Example

- Consider the ciphertext:

CDGAV	NNANX	DOKVZ	XDGVG	OBMXG	HVLOL
QFZIA	PJAXB	OAGTZ	FBTGA	IBVUK	LOBZT
SMDNV	GKSII	OAGJO	BJLGO	CIDGV	ZZCMH
YFGUI	ZWSBY	VKGUV	FGXFU	ZFOLK	FJOMZ
CMGMO	AGLCC	MWCDG	NCXUW	YCAYG	JALJF
GSXBJ	MJWMC	IECFB	OVZGU	PMOHO	KVHYE
CONNC	XTAQY	MZCJJ	H1XUW	KUMTV	WNNCX
ISPFN	YTGHN	CXCIP	COTPA	OBZFC	JIYVG
FLCYN	XKFZM	ZICJV	NZMJW	HZMHO	LCYWX

  - The trigram `CDG` occurs in positions 1 and 133.
  - The bigram `DG` occurs in positions 2,17,83, and 134.

Example

- Consider the ciphertext:

CDGAV	NNANX	DOKVZ	XDGVG	OBMXG	HVLOL
QFZIA	PJAXB	OAGTZ	FBTGA	IBVUK	LOBZT
SMDNV	GKSII	OAGJO	BJLGO	CIDGV	ZZCMH
YFGUI	ZWSBY	VKGUV	FGXFU	ZFOLK	FJOMZ
CMGMO	AGLCC	MWCDG	NCXUW	YCAYG	JALJF
GSXBJ	MJWMC	IECFB	OVZGU	PMOHO	KVHYE
CONNC	XTAQY	MZCJJ	H1XUW	KUMTV	WNNCX
ISPFN	YTGHN	CXCIP	COTPA	OBZFC	JIYVG
FLCYN	XKFZM	ZICJV	NZMJW	HZMHO	LCYWX

  - The trigram CDG occurs in positions 1 and 133.
  - The bigram DG occurs in positions 2, 17, 83, and 134.
  - The bigram NN occurs in positions 6, 183, and 207.

Example

- Consider the ciphertext:

CDGAV	NNANX	DOKVZ	XDGVG	OBMXG	HVLOL
QFZIA	PJAXB	OAGTZ	FBTGA	IBVUK	LOBZT
SMDNV	GKSII	OAGJO	BJLGO	CIDGV	ZZCMH
YFGUI	ZWSBY	VKGUV	FGXFU	ZFOLK	FJOMZ
CMGMO	AGLCC	MWCDG	NCXUW	YCAYG	JALJF
GSXBJ	MJWMC	IECFB	OVZGU	PMOHO	KVHYE
CONNC	XTAQY	MZCJJ	H1XUW	KUMTV	WNNCX
ISPFN	YTGHN	CXCIP	COTPA	OBZFC	JIYVG
FLCYN	XKFZM	ZICJV	NZMJW	HZMHO	LCYWX

  - The trigram `CDG` occurs in positions 1 and 133.
  - The bigram `DG` occurs in positions 2, 17, 83, and 134.
  - The bigram `NN` occurs in positions 6, 183, and 207.
  - The trigram `OAG` occurs in positions 41, 71, and 125.
- There are others, but we'll start with these.

The Art of the Key

- By assumption, some (but not necessarily all) of these repeated bigrams and trigrams are separated by multiples of k, the keylength:
  - The trigram `CDG` occurs in positions 1 and 133: These are 133-1=132 spaces apart, so 132 might be a multiple of k
  - The bigram `DG` occurs in positions 2, 17, 83, and 134: These are 17-2=15, 83-2=81, 134-2=132, 83-17=66, 134-17=117, and 134-83=51 spaces apart, so some of these might be multiples of k .
  - The bigram `NN` occurs in positions 6, 183, and 207: these are 183-6=177, 207-6=201, and 207-183=24 spaces apart, so some of these might be multiples of k .
  - The trigram `OAG` occurs in positions 41, 71, and 125: These are 71-41=30, 125-41=84, and 125-71=54 spaces apart, so some of these might be multiples of k.

The Art of the Key

- If you find every occurrence of every bigram and trigram, you'll generally find ... nothing useful.
- This is because you'll have a large set of numbers, and the only thing that *all* the numbers will be multiples of is 1, which would produce a shift cipher (and if it's a shift cipher, you wouldn't bother with this approach).

The Art of the Key

- By assumption, some (but not necessarily all) of these repeated bigrams and trigrams are separated by multiples of k, the keylength:
  - The trigram `CDG` occurs in positions 1 and 133: These are 133-1=132 spaces apart.
  - The bigram `DG` occurs in positions 2, 17, 83, and 134: These are 17-2=15, 83-2=81, 134-2=132, 83-17=66, 134-17=117, and 134-83=51 spaces apart.
  - The bigram `NN` occurs in positions 6, 183, and 207: These are 183-6=177, 207-6=201, and 207-183=24 spaces apart.
  - The trigram `OAG` occurs in positions 41, 71, and 125: These are 71-41=30, 125-41=84, and 125-71=54 spaces apart.
- The art of the Kasiski attack is finding a number that divides most but not all of the separations.
- Here, most of the numbers are divisible by 6, but not all.

Example: Decryption

- If we assume a keylength of 6, then every 6th letter comes from same shift:

CDGAV	NNANX	DOKVZ	XDGVG	OBMXG	HVLOL
QFZIA	PJAXB	OAGTZ	FBTGA	IBVUK	LOBZT
SMDNV	GKSII	OAGJO	BJLGO	CIDGV	ZZCMH
YFGUI	ZWSBY	VKGUV	FGXFU	ZFOLK	FJOMZ
CMGMO	AGLCC	MWCDG	NCXUW	YCAYG	JALJF
GSXBJ	MJWMC	IECFB	OVZGU	PMOHO	KVHYE
CONNC	XTAQY	MZCJJ	H1XUW	KUMTV	WNNCX
ISPFN	YTGHN	CXCIP	COTPA	OBZFC	JIYVG
FLCYN	XKFZM	ZICJV	NZMJW	HZMHO	LCYWX

Example: Decryption

- If we assume a keylength of 6, then every 6th letter comes from same shift:

C	N	K	V	G
Q	J	G	G	K
S	K	G	G	V
Y	W	G	F	K
C	G	C	U	G
G	J	C	G	O
C	T	C	U	V
I	T	C	P	C
F	K	C	J	O

- So the 1st, 7th, 13th, etc., letters are all from the same shift
- Given the high frequencey of the ciphertext `G`, it's resonable to assume `E` $\to$ `G`, suggesting a shift 2, and giving the first letter of the keyword: `C`

Example: Decryption

- 1st, 7th, 13th, ... letters are decrypted:

ADGAV	N**L**ANX	DO**I**VZ	XDG**T**G	OBMX**E**	HVLOL
OFZIA	P**H**AXB	OA**E**TZ	FBT**E**A	IBVU**I**	LOBZT
QMDNV	G**I**SII	OA**E**JO	BJL**E**O	CIDG**T**	ZZCMH
WFGUI	Z**U**SBY	VK**E**UV	FGX**D**U	ZFOL**I**	FJOMZ
AMGMO	A**E**LCC	MW**A**DG	NCX**S**W	YCAY**E**	JALJF
ESXBJ	M**H**WMC	IE**A**FB	OVZ**E**U	PMOH**M**	KVHYE
AONNC	X**R**AQY	MZ**A**JJ	H1X**S**W	KUMT**T**	WNNCX
GSPFN	Y**R**GHN	CX**A**IP	COT**N**A	OBZF**A**	JIYVG
DLCYN	X**I**FZM	ZI**A**JV	NZM**H**W	HZMH**M**	LCYWX

Example: Decryption

- If we assume a keylength of 6, then every 6th letter comes from same shift:

D	A	V	G	H
F	A	T	A	L
M	S	J	O	Z
F	S	U	U	F
M	L	D	W	J
S	W	F	U	K
O	A	J	W	W
S	G	I	A	J
L	F	J	W	L

- Next, take the 2nd, 8th, etc. letters,
- There are 5 `A`s, 5`F`s, 5 `J`s and 5 `W`s, so it's harder to tell which might be `E`. So, we might try to look at the frequency histograms.
  - If `E` $\rightarrow$ `A`, `J` $\rightarrow$ `F`, suggesting a plaintext with many `J`s
  - If `E` $\rightarrow$ `F`, then `Z` $\rightarrow$ `A`, suggesting a plaintext with many `Z`s
  - If `E` $\rightarrow$ `J`, then `V` $\rightarrow$ `A`, suggesting a plaintext with many `V`s
  - If `E` $\rightarrow$ `W`, then `I` $\rightarrow$ `A`, `N` $\rightarrow$ `F`, `R` $\rightarrow$ `J` suggesting a plaintext with many `I`s, `N`s, and `R`s.It would be reasnoble to conclude `E` $\rightarrow$ `W`, giving `S` as the second letter of the keyword.

Example: Decryption

- 1st, 2nd, 7th, 8th, ... letters are decrypted:

ALGAV	N**LI**NX	DOI**DZ**	XDG**TO**	OBMX**E**	**P**VLOL
ONZIA	P**HI**XB	OA**EB**Z	FBT**EI**	IBVU**I**	**T**OBZT
QUDNV	G**IA**II	OA**ER**O	BJL**EW**	CIDG**T**	**H**ZCMH
WNGUI	Z**UA**BY	VK**EC**V	FGX**DC**	ZFOL**I**	**N**JOMZ
AUGMO	A**ET**CC	MW**AL**G	NCX**SE**	YCAY**E**	**R**ALJF
EAXBJ	M**HE**MC	IE**AN**B	OVZ**EC**	PMOH**M**	**S**VHYE
AWNNC	X**RI**QY	MZ**AR**J	H1X**SE**	KUMT**T**	**E**NNCX
GAPFN	Y**RO**HN	CX**AQ**P	COT**NI**	OBZF**A**	**R**IYVG
DTCYN	X**IN**ZM	ZI**AR**V	NZM**HE**	HZMH**M**	**T**CYWX

Example: Decryption

- *All Gaul is divided into three parts, one of which the Belgae inhabit, the Aquitani another, those who in their own language are called Celts, in our Gauls, the third. All these differ from each other in language, customs and laws. The river Garonne separates the Gauls from the Aquitani; the Marne and the Seine separate them from the Belgae.* **The Gallic Wars, By Julius Caesar**

Vernam cipher (One-time pad)

- The Vernam cipher (one-time pad) is an information-theoretically secure cryptosystem.
- This means that Eve, knowing only the ciphertext, can extract absolutely no information about the plaintext other than its length.
- Perfect secrecy: observation of the ciphertext provides no information to an adversary
    - Informally, perfect secrecy means that an attacker can not obtain any information about the plaintext, by observing the ciphertext.

Vernam cipher (One-time pad)

- One-time pad is a cipher that cannot be broken if it is used correctly.
- Rules:
  - The key is as long as the message.
  - The key is random.
  - The key is never reused.

Exclusive-or on bits

- Vernam cipher is based on exclusive-or (XOR), which we write as $\oplus$
  - $x \oplus y$ is true when exactly one of x and y is true.
  - $x \oplus y$ is false when x and y are both true or both false.
- Exclusive-or is just sum modulo two if 1 represents true and 0 represents false.

$$x \oplus y = (x + y) \bmod 2$$

- XOR is associative and commutative. 0 is the identity element.

$$k \oplus 0 = 0 \oplus k = k$$

- XOR is its own inverse.

$$k \oplus k = 0$$

One-Time Pad: Informal description

- The one-time pad encrypts a message m by XORing it with the key k, which must be as long as m.
- Assume both m and k are represented by strings of bits. Then ciphertext bit $c_i = m_i \oplus k_i$.
- Note that $c_i = m_i$ if $k_i = 0$, and $c_i = \neg m_i$ if $k_i = 1$.
- Decryption is the same, i.e., $m_i = c_i \oplus k_i$

One-Time Pad: Encryption

- Let `a=000, h=001, i=011, k=100, p=101, y=111`
- **Encryption**: `Plaintext` $\oplus$ `Key = Ciphertext`

	h	a	p	p	y
Plaintext	001	000	101	101	111
Key	101	111	110	101	011
Ciphertext	100	111	011	000	100
	k	y	i	a	k

One-Time Pad: Decryption

- Let a=000, h=001, i=011, k=100, p=101, y=111
- **Decryption**: Ciphertext $\oplus$ Key = Plaintext

	k	y	i	a	k
Ciphertext	100	111	011	000	100
Key	101	111	110	101	011
Plaintext	001	000	101	101	111
	h	a	p	p	y

The one-time pad cryptosystem formally defined

- $\mathcal{M} = \mathcal{C} = \mathcal{K} = \{0,1\}^r$ for some length r.
- $E_k(m) = k \oplus m$, where $\oplus$ is applied to corresponding bits of k & m.
- $D_k(c) = k \oplus c$, where $\oplus$ is applied to corresponding bits of k & c.
- It works because

$$D_k(E_k(m)) = k \oplus (k \oplus m) = (k \oplus k) \oplus m = 0 \oplus m = m$$

One-time pad: Security

- Like the 1-letter Caesar cipher, for given m and c, there is exactly one key k such that $E_k(m) = c$ (namely, $k = m \oplus c$)
- For fixed c, m varies over all possible messages as k ranges over all possible keys, so c gives no information about m.
- It will follow that the one-time pad is information-theoretically secure

Importance of the Vernam cipher

- It is important because
  - it is sometimes used in practice;
  - it is the basis for many stream ciphers, where the truly random key is replaced by a pseudo-random bit string.

Attraction of one-time pad

- The one-time pad would seem to be the perfect cryptosystem.
  - It works for messages of any length (by choosing a key of the same length).
  - It is easy to encrypt and decrypt.
  - It is information-theoretically secure.
- In fact, it is sometimes used for highly sensitive data.

Drawbacks of one-time pad

- It has two major drawbacks:
  1. The key k must be as long as the message to be encrypted.
  2. The same key must never be used more than once. (Hence the term "one-time".)
- Together, these make the problem of key distribution and key management very difficult.

Drawbacks of one-time pad[1]

- Example taken from "Security Engineering", Ross Anderson, 2nd edition (Wiley)
- One-time pad was used in World War 2: one-time key material was printed on silk, which agents could conceal inside their clothing; whenever a key had been used, it was torn off and burnt
- Now suppose you intercepted a message from a wartime German agent which you know started with "Heil Hitler", and the first 10 letters of ciphertext were `DGTYI BWPJA`
- Means the first 10 letters of the one-time pad were `wclnb tdefj` since (A spy's message)

```
Plaintext :  heilhitler
Key       :  wclnbtdefj
Ciphertext:  DGTYIBWPJA
```

[1]Slides credit: Fabio Martignon

Example

```
Plaintext :  heilhitler
Key       :  wclnbtdefj
```

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

A	B	C	D	E	F	G	H	I	J	K	L	M
26	27	28	29	30	31	32	33	34	35	36	37	38
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
39	40	41	42	43	44	45	46	47	48	49	50	51

- key w, letter h: h is 7, and w is 22, we add the numbers, 29 which means D is the ciphertext.
- key c, letter e: e is 4, and c is 2, we add the numbers, 6 which means G is the ciphertext.
- etc

Drawbacks of one-time pad

- But once he has burnt the piece of silk with his key material, the spy can claim he's actually a member of the anti-Nazi underground resistance, and the message actually said "Hang Hitler".
- This is quite possible, as the key material could just as easily have been `wggsb tdefj`:
- What the spy claimed he said:

```
Ciphertext:  DGTYIBWPJA
Key       :  wggsbtdefj
Plaintext :  hanghitler
```

Drawbacks of one-time pad

- Now we rarely get anything for nothing in cryptology, and the price of the perfect secrecy of the one-time pad is that it fails completely to protect message integrity.
- Suppose for example that you wanted to get this spy into trouble, you could change the ciphertext to `DCYTI BWPJA`
- Manipulating the message to entrap the spy he said:

```
Ciphertext:  DCYTIBWPJA
Key       :  wclnbtdefj
Plaintext :  hanghitler
```

Why the key cannot be reused

- If Eve knows just one plaintext-ciphertext pair (m_1, c_1), then she can recover the key $k = m_1 \oplus c_1$.
- This allows her to decrypt all future messages sent with that key.
- Even in a ciphertext-only situation, if Eve has two ciphertexts c_1 and c_2 encrypted by the same key k, she can gain significant partial information about the corresponding messages m_1 and m_2.
- In particular, she can compute $m_1 \oplus m_2$ without knowing either m_1 or m_2 since

$$m_1 \oplus m_2 = (c_1 \oplus k) \oplus (c_2 \oplus k) = c_1 \oplus c_2$$

How knowing $m_1 \oplus m_2$ might help an attacker

- Fact (important property of $\oplus$)
  - For bits b_1 and b_2, $b_1 \oplus b_2 = 0$ if and only if $b_1 = b_2$
  - Hence, blocks of `0's` in $m_1 \oplus m_2$ indicate regions where the two messages m_1 and m_2 are identical.
  - That information, together with other information Eve might have about the likely content of the messages, may be enough for her to seriously compromise the secrecy of the data.

Key Randomness in One-Time Pad

- One-time pad uses a very long key, what if the key is not chosen randomly, instead, texts from, e.g., a book is used.
  - this is not one-time pad anymore
  - this does not have perfect secrecy
  - this can be broken
- The key in one-time pad should never be reused.
  - If it is reused, it is two-time pad, and is insecure!

One-Time Pad Summary

- Provably secure:
  - Ciphertext provides no info about plaintext.
  - All plaintexts are equally likely
- But, only when used correctly!
  - Pad must be random, used only once.
  - Pad is known only to sender and receiver.
- Note: pad (key) is same size as message.
  - So, why not distribute msg instead of pad?

Eve's goals

- Eve wants learn something. Eve is not bound by any rules. She can do as she wishes with the information she has available.
- We don't want her to be able to:
  - Recover the key.
  - Find the plaintext to a ciphertext.
  - Determine any character to the plaintext.
  - Derive any meaningful information about the plaintext.

Eve's information

- Until now, we've implicitly assumed that Eve has no information about the cryptosystem except for the encryption and decryption methods and the ciphertext c.
- In practice, Eve might know much more.
  - She probably knows (or has a good idea) of the message distribution.
  - She might have obtained several other ciphertexts.
  - She might have learned the decryptions of earlier ciphertexts.
  - She might have even chosen the earlier messages or ciphertexts herself
- This leads us to consider several attack scenarios.

Attack scenarios

- Ciphertext-only attack
  - Eve knows only the ciphertext to be decoded c and tries to recover m.
- Known plaintext attack
  - Eve knows the ciphertext to be decoded c and a sequence of plaintext-ciphertext pairs (m_1, c_1),...,(m_r, c_r) where $c \notin \{c_1, \ldots, c_r\}$.
  - She tries to recover m.

Known plaintext attacks

- A known plaintext attack can occur when
  1. Alice uses the same key to encrypt several messages;
  2. Eve later learns or successfully guesses the corresponding plaintexts.
- Some ways that Eve learns plaintexts.
  - The plaintext might be publicly revealed at a later time, e.g., sealed bid auctions.
  - The plaintext might be guessable, e.g., an email header.
  - Eve might later discover the decrypted message on Bob's computer.

Chosen text attack scenarios

- Still stronger attack scenarios allow Eve to choose one element of a plaintext-ciphertext pair and obtain the other
- Chosen plaintext attack
  - Like a known plaintext attack, except that Eve chooses messages $m_1, \dots, m_r$ before getting c and Alice (or Bob) encrypts them for her.
- Chosen ciphertext attack
  - Like a known plaintext attack, except that Eve chooses ciphertexts $c_1, \dots, c_r$ before getting c and Alice (or Bob) decrypts them for her.
- Mixed chosen plaintext/chosen ciphertext attack
  - Eve chooses some plaintexts and some ciphertexts and gets the corresponding decryptions or encryptions.

Why would Alice cooperate in a chosen plaintext attack?

- Eve might be authorized to generate messages that are then encrypted and sent to Bob, but she isn't authorized to read other people's messages.[2]
- Alice might be an internet server, not a person, that encrypts messages received in the course of carrying out a more complicated cryptographic protocol.[3]
- Eve might gain access to Alice's computer, perhaps only for a short time, when Alice steps away from her desk.

[2] Nothing we have said implies that Eve is unknown to Alice and Bob or that she isn't also a legitimate participant in the protocol.

[3] We will see such protocols later in the course.

Adaptive chosen text attack scenarios

- Adaptive versions of chosen text protocols are when Eve chooses her texts one at a time after learning the response to her previous text
- Adaptive chosen plaintext attack
  - Eve chooses the messages $m_1, m_2, \ldots$ one at a time rather than all at once.
  - Thus,
    - m_2 depends on (m_1, c_1)
    - m_3 depends on both (m_1, c_1) and (m_2, c_2), etc.
- Adaptive chosen ciphertext and adaptive mixed attacks
  - are defined similarly

Exhaustive Key Search

- Exhaustive key search
  - Eve can simply try all possible keys and test each to see if it is correct.
  - Remember, she has some ciphertexts so she knows when she found the right key.
- To prevent an exhaustive key search, a cryptosystem must have a large keyspace.
  - The set of all possible keys that can be used to generate a key.
  - Must be too many keys for Eve to try them all in any reasonable amount of time.

Beyond Exhaustive Search

- A large keyspace is necessary for security.
- But a large keyspace is not sufficient.
    - Shortcut attacks might exist.
    - In cryptography we can (almost) never prove that no shortcut attack exists

Key Points

- Two basic types of ciphers
  - Transposition ciphers and substitution ciphers
- Caesar cipher uses one key
- Vigènere cipher uses a sequence of keys
- Cryptanalysis
  - Exhaustive search
  - Statistical analysis

- Next: Modern Cryptography