Historical Ciphers and Unconditional Security

Gerardo Pelosi

Department of Electronics, Information and Bioengineering – (DEIB)
Politecnico di Milano

gerardo.pelosi - at - polimi.it

Overview

Lesson contents

- Historical symmetric-key ciphers (aka *Hand Ciphers*)
  - Substitution ciphers
    - Mono-alphabetic ciphers
    - Poly-alphabetic ciphers
  - Permutation ciphers and Affine ciphers
- Information Theoretic security
  - Shannon's theorem
  - Vernam cipher
  - Entropy, Spurious keys and Unicity distance

(Mandatory readings: Chapter 3 and Chapter 5 of Smart's book)

Monoalphabetic Cipher

Shift cipher

- Given an alphabet $\mathcal{A} = \{\text{ABCDEFGHIJKLMNOPQRSTUVWXYZ}\}$, each letter is identified with a number: A$=0$, B$=1,\ldots,$Z$=25$
- The message space $\mathcal{M}$ includes messages composed of a single letter.
- The key of the cipher is a number $0 \leq k \leq 25$
- The encryption replaces each plaintext letter by the letter which is k places forward in $\mathcal{A}$: $\mathbf{c} = \mathbf{p} + \mathbf{k} \mod \mathbf{26}$, where p and c are the numbers denoting the positions of the corresponding letters in $\mathcal{A}$.

- If $\mathbf{k} = \mathbf{3}$, this is known as **Caesar's cipher**
- If $\mathbf{k} = \mathbf{13}$, the encryption and decryption are the same process, and this is known as **rot13 cipher**.
- *Observation*.
  The keyspace cardinality $|\mathcal{K}|$ is 26: a bruteforce attack is immediate.

Generalization of the Shift Cipher: Substitution Ciphers

Monoalphabetic Substitution Cipher

- The message space is defined over an alphabet $\mathcal{A}_m$, e.g., $\mathcal{A}_m = \{\text{ABCDEFGHIJKLMNOPQRSTUVWXYZ}\}$
- The ciphertext space is defined over an alphabet $\mathcal{A}_c$, e.g., $\mathcal{A}_c = \{\text{abcdefghijklmnopqrstuvwxyz}\}$ (...Possibly, $\mathcal{A}_m \equiv \mathcal{A}_c$)
- The sizes of the two alphabets **must match**, i.e.: $|\mathcal{A}_m| = |\mathcal{A}_c|$
- The message spaces $\mathcal{M}, \mathcal{C}$ include messages composed of a single letter, respectively. That is: $\mathcal{M} \equiv \mathcal{A}_m$ and $\mathcal{C} \equiv \mathcal{A}_c$.
- **The encryption transformation can be defined as the application of any bijective map between the elements of $\mathcal{M}$ and the elements of $\mathcal{C}$**

Substitution Ciphers

Monoalphabetic Substitution Cipher

- $\mathcal{M}$: set of the capital letters of the English alphabet; $\mathcal{C}$: set of the small letters of the English alphabet; listed in lexicographical order
- The **cipher key** $\mathcal{K}$ identifies the **map** (between the two English alphabets listed in lexicographical order) to compute the ciphertext.

Map μ

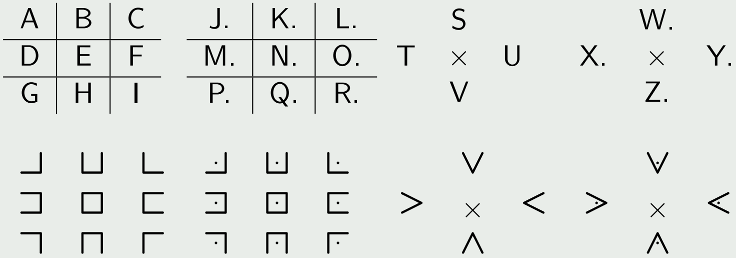
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
t	m	k	g	o	y	d	s	i	p	e	l	u	a	v	c	r	j	w	x	z	n	h	b	q	f

- The encryption of the word "CRYPTOGRAPHY" is obtained applying the map μ sequentially to each ptx letter: "kjqcxvdjtcsq"
- There are as many keys as the bijective maps, i.e., $(|\mathcal{A}_m|)!$
- For $|\mathcal{A}_m| = 26$, the keyspace is $26! \approx 2^{88}$ keys wide, which is too large to bruteforce

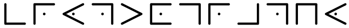
Substitution ciphers

Monoalphabetic Substitution Cipher: PigPen

A well known substitution cipher employs the following mapping between the english alphabet and graphical symbols



Example: The encryption of the word "CRYPTOGRAPHY" is



Substitution Ciphers

Cryptanalysis

There is a **1-to-1 correspondence** between plaintext (ptx) and ciphertext (ctx) letters (also digrams, trigrams, etc). Thus, assuming the Kerchoff's principle holds, a *Ciphertext-Only attack* (COA) easily reveals the key.

(COA: ctxs corresponding to several messages encrypted with the same key should be known)

- The statistics of the plaintext language (frequency distribution of the symbols of $\mathcal{A}_m$ in $\mathcal{M}$) is known
- The statistics of the ciphertext space are derived for the symbols of $\mathcal{A}_c$ in $\mathcal{C}$, over the available ciphertexts
- The substitution map (i.e., the key) **can be inferred** easily by **matching the symbols occurring with similar frequencies**

A *Known Plaintext Analysis* or a *Chosen Plaintext Analysis* can easily reveal the key, as well.

Cryptanalysis COA of a Shift Cipher (1/3)

Given the following ciphertext:

```
BPMZM WVKM EIA IV COTG LCKSTQVO EQBP NMIBPMZA ITT ABCJJG IVL JZWEV IVL BPM WBPMZ
JQZLA AIQL QV AW UIVG EWZLA OMB WCB WN BWEV OMB WCB, OMB WCB, OMB WCB WN BWEV
IVL PM EMVB EQBP I YCIKS IVL I EILLTM IVL I YCIKS QV I NTCZZG WN MQLMZLWEV BPIB
XWWZ TQBBTM COTG LCKSTQVO EMVB EIVLMZQVO NIZ IVL VMIZ JCB IB MDMZG XTIKM BPMG
AIQL BW PQA NIKM VWE OMB WCB, OMB WCB, OMB WCB WN PMZM IVL PM EMVB EQBP I YCIKS
IVL I EILLTM IVL I YCIKS IVL I DMZG CVPIXXG BMIZ
```

We need to compare the frequency distribution of the letters in this text with the one of the standard English language

Cryptanalysis COA of a Shift Cipher (2/3)

English letter frequencies

A	8.05	J	0.10	S	6.59
B	1.62	K	0.52	T	9.59
C	3.20	L	4.03	U	3.10
D	3.65	M	2.25	V	0.93
E	12.31	N	7.19	W	2.03
F	2.28	O	7.94	X	0.20
G	1.61	P	2.29	Y	1.88
H	5.14	Q	0.20	Z	0.09
I	7.18	R	6.03		

Most common bigrams: `TH,HE,IN,ER,`
`AN,RE,ED,ON,`
`ES,ST,EN,AT,`
`TO,NT,HA`

Most common trigrams: `THE,ING,AND,`
`HER,ERE,ENT,`
`THA,NTH,WAS,`
`ETH,FOR`

Letter frequencies in the ciphertext

A	2.59	J	1.44	S	1.73
B	10.37	K	2.59	T	3.46
C	5.48	L	6.63	U	0.29
D	0.58	M	10.09	V	8.36
E	4.61	N	2.31	W	6.63
F	0.00	O	3.46	X	1.15
G	2.59	P	4.03	Y	1.15
H	0.00	Q	4.03	Z	4.90
I	11.53	R	0.00		

Cracking the cipher
The shift of `E` seems to be 4, 8, 17, 18, or 23
The shift of `A` seems to be 1, 8, 12, 21, or 22
Hence, a first guess about the key value is $k=8$
we can now decrypt the ciphertext to reveal:

Cryptanalysis COA of a Shift Cipher (3/3)

Inferred plaintext:

```
THERE ONCE WAS AN UGLY DUCKLING WITH FEATHERS ALL STUBBY AND BROWN AND THE OTHER
BIRDS SAID IN SO MANY WORDS GET OUT OF TOWN GET OUT, GET OUT, GET OUT OF TOWN
AND HE WENT WITH A QUACK AND A WADDLE AND A QUACK IN A FLURRY OF EIDERDOWN THAT
POOR LITTLE UGLY DUCKLING WENT WANDERING FAR AND NEAR BUT AT EVERY PLACE THEY
SAID TO HIS FACE NOW GET OUT, GET OUT, GET OUT OF HERE AND HE WENT WITH A QUACK
AND A WADDLE AND A QUACK AND A VERY UNHAPPY TEAR
```

Cryptanalysis of a Monoalphabetic Substitution Cipher (1/5)

Given the following ciphertext:

```
XSO MJIWXVL JODIVA STW VAO VY OZJVCO’W LTJDOWX KVAKOAXJTXI- VAW VY SIDS
XOKSAVLVDQ IAGZWXJQ. KVUCZXOJW, KVUUZAIKTXIVAW TAG UIKJVOLOKXJVAIKW TJO HOLL
JOCJOWOAXOG, TLVADWIGO GIDIXTL UOGIT, KVUCZXOJ DTUOW TAG OLOKXJVAIK KVUUOJKO. TW
HOLL TW SVWXIAD UTAQ JOWOTJKS TAG CJVGZKX GONOLVCUOAX KOAXJOW VY UTPVJ DLVMTL
KVUCTAIOW, XSO JODIVA STW T JTCIGLQ DJVHIAD AZU- MOJ VY IAAVNTXINO AOH
KVUCTAIOW. XSO KVUCZXOJ WKIOAKO GOC- TJXUOAX STW KLVWO JOLTXIVAWSICW HIXS UTAQ
VY XSOWO VJDTAI- WTXIVAW NIT KVLLTMVJTXINO CJVPOKXW, WXTYY WOKVAGUOAXW TAG
NIWIXIAD IAGZWXJITL WXTYY. IX STW JOKOAXLQ IAXJVGZKOG WONO- JTL UOKSTAIWUW YVJ
GONOLVCIAD TAG WZCCVJXIAD OAXJOCJOAOZJITL WXZGOAXW TAG WXTYY, TAG TIUW XV CLTQ T
WIDAIYIKTAX JVLO IA XSO GONOLVCUOAX VY SIDS-XOKSAVLVDQ IAGZWXJQ IA XSO JODIVA.
```

Again, we need to compare the frequency distribution of the letters in this text with the one of the standard English language

Cryptanalysis of a Monoalphabetic Substitution Cipher (2/5)

English letter frequencies

A	8.05	J	0.10	S	6.59
B	1.62	K	0.52	T	9.59
C	3.20	L	4.03	U	3.10
D	3.65	M	2.25	V	0.93
E	12.31	N	7.19	W	2.03
F	2.28	O	7.94	X	0.20
G	1.61	P	2.29	Y	1.88
H	5.14	Q	0.20	Z	0.09
I	7.18	R	6.03		

Most common bigrams: TH,HE,IN,ER, AN,RE,ED,ON, ES,ST,EN,AT, TO,NT,HA

Most common trigrams: THE,ING,AND, HER,ERE,ENT, THA,NTH,WAS, ETH,FOR

Letter frequencies in the ciphertext

A	8.99	J	6.51	S	3.26
B	0.00	K	4.81	T	8.06
C	2.95	L	4.34	U	3.57
D	3.10	M	0.62	V	7.60
E	0.00	N	1.40	W	7.13
F	0.00	O	11.63	X	7.75
G	3.72	P	0.31	Y	1.61
H	0.78	Q	1.40	Z	2.17
I	7.75	R	0.00		

Most common bigrams: TA,AX,IA,VA WX,XS,AG,OA JO,JV

Most common trigrams: OAX,TAG,IVA XSO,KVU,TXI UOA,AXS

Cryptanalysis of a Monoalphabetic Substitution Cipher (3/5)

Analysis

Since O in the ciphertext occurs with frequency 11.63 we can guess E=O
Common trigrams in the ciphertext are: OAX=E**, and XSO=**E
Common similar trigrams in English are: ENT, ETH, and THE

Hence a first guess to partially decrypt the ciphertext may be: E=O, T=X, H=S, N=A

For the sake of conciseness, from now on we only look at the first two sentences of the ciphertext:

THE MJIWTVL JEDIVN HTW VNE VY EZJVCE’W LTJDEWT KVNKENTJTTIV NW VY HIDH

TEKHNVLVDQ INGZWTJQ. KVUCZTEJW, KVUU ZNIKATIVNW AND UIKJVELEKTJVNIKW AJE HELL

JECJEWENTED, ALVNDWIDE DIDITAL UEDIA, KVUCZTEJ DAUEW AND ELEKTJVNIK KVUUEJKE.

Cryptanalysis of a Monoalphabetic Substitution Cipher (4/5)

Analysis

Since T in the ciphertext occurs with frequency 8.06 it is likely that T=A, thus looking also at the bigrams and trigams we can infer that TA=AN, TAG=AN*, therefore a second guess can be: T=A and G=D

Looking at the bigrams: IX=*T, and XV=T*, due to the fact that the plaintext is in English it is highly likely that: I=I and V=O, obtaining:

```
THE MJIWTVL JEDIVN HAW VNE VY EZJVCE'W LAJDEWT KVNKENTJATIV NW VY HIDH
TEKHNVLVDQ INDZWTJQ. KVUCZTEJW, KVUU ZNIKTTIVNW TNG UIKJVELEKTJVNIKW TJE HELL
JECJEWENTEG, TLVNDWIGE GIDITTL UEGIT, KVUCZTEJ DTUEW TNG ELEKTJVNIK KVUUEJKE.
```

Cryptanalysis of a Monoalphabetic Substitution Cipher (5/5)

Analysis

Two more letters: VY=O* Hence Y must be one of F, N, R due to English. We already solved the map for the plain letter N, while Y has probability 1.61, F has probability 2.28, R has probability 6.03; Therefore Y=F.

We also have: IW=I* Therefore W must be one of F, N, S, T.
Already have the maps for F, N, T, thus W=S.

Summarizing: I=I, V=O, Y=F, W=S

```
THE MJISTOL JEDION HAS VONE OF EZJOCE'S LAJDEST KONKENTJATIONS OF HIDH
TEKHNOLODQ INDZSTJQ. KOUCZTEJS, KOUU ZNIKATIONS AND UIKJOELEKTJONIKS AJE HELL
JECJESENTED, ALONDSIDE DIDITAL UEDIA, KOUCZTEJ DAUES AND ELEKTJONIK KOUUEJKE.
```

Now, (thanks to the natural redundancy of English sentences) it is easy to infer what the underlying plaintext is

Exercise: Cryptanalysis of the Gold Bug Cipher (1/2)

"The Gold-Bug" is a short story by Edgar Allan Poe published in 1843. The plot follows William Legrand who was bitten by a gold-colored bug he found on the coastline together with a scarp piece of parchment. The parchment proved to contain a cryptogram written with invisible ink, and revealed by the heat of fire burning on the ground.

The Gold Bug Cryptogram

```
53‡‡†305))6*;4826)4‡.)4‡);806*;48†8
$60))85;1‡(;:‡*8†83(88)5*†;46(;88*96
*?;8)*‡(;485);5*†2:*‡(;4956*2(5*-4)8
$8*;4069285);)6†8)4‡‡;1(‡9;48081;8:8‡
1;48†85;4)485†528806*81(‡9;48;(88;4
(‡?34;48)4‡;161;:188;‡?;
```

Symbol	freq.	in the ctx	
8	33	1	8
;	26	0	6
4	19	9	5
‡	16	2	5
)	16	:	4
*	13	3	4
5	12	?	3
6	11	$	2
†	8	-	1
		.	1

Solution: Cryptanalysis of the Gold Bug Cipher (2/2)

The decrypted message with spaces, punctuation, and capitalization is:

A good glass in the bishop’s hostel in the devil’s seat twenty-one degrees and thirteen minutes northeast and by north main branch seventh limb east side shoot from the left eye of the death’s-head a bee line from the tree through the shot fifty feet out.

Substitution Ciphers

Statistics on the letters of the ptxs and ctxs could be used to break monoalphabetic substitution ciphers! From the early 1800s onwards cipher designers tried to break this link

Polyalphabetic Cipher

- The plaintext and ciphertext spaces include finite sequence of letters (*words*) from the alphabets $\mathcal{A}_m$ and $\mathcal{A}_c$, respectively.
- The encryption transformation is defined as the application of "**L** $>$ **1 bijective maps**" between the two alphabets: $\boldsymbol{\mu_0}, \boldsymbol{\mu_1}, \ldots,$
- The encryption transformation applies
  - $\boldsymbol{\mu_0}$ to the 1st letter of the ptx, $\boldsymbol{\mu_1}$ to the 2nd letter, and so on
  - ... periodically in **L**, i.e., $\boldsymbol{\mu_0}$ is applied again to the $(L+1)$-th letter of the ptx word, etc...
- The key $k = (\mu_0, \mu_1, \ldots, \mu_{L-1})$ is constituted by the L maps employed to encrypt the ciphertext
- For $|\mathcal{A}_m| = 26, L = 5$, the keyspace is $((|\mathcal{A}_m|)!)^L = (26!)^5 \approx 2^{441}$ keys wide, which is way too large to bruteforce

Historical Ciphers

Polyalphabetic Cipher

A simple example with L=2:

Map μ_0

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

Map μ_1

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

Encrypting HELLO yields the ciphertext SHLJV

Substitution Ciphers

Vigenère Cipher (1586): a famous *hand cipher*, which is a particular case of the Polyalphabetic cipher

- An ordered alphabet $\mathcal{A}$ is assumed as both plaintext and ciphertext alphabets, while ptx and ctx messages are **letter sequences**: *words*
- The cipher employs **L cyclic shifts of $\mathcal{A}$ as maps for the encryption**
- The cipher key is given as a sequence of L letters (*keyword*). Each of them denotes the 1st letter of a cyclic shift of the alphabet
- The key space is now limited to: $|\mathcal{K}| \leq 26^L$, $L \geq 1$
- Operatively, the definition of the maps is as follow: each letter is denoted with a number: A$=0$, B$=1$,...,Z$=25$, and the cipher key is given by: $\mathbf{k}=(\mathbf{k_0}, \mathbf{k_1}, \ldots, \mathbf{k_{L-1}})$ $\mathbf{k_j} \in \{\mathbf{0}, \ldots, \mathbf{25}\}$

Given $\mathbf{p} \in \mathcal{M}$, $\mathbf{p} = \mathbf{p_0 p_1} \ldots, \mathbf{p_i}, \ldots$

$$\mathbf{c_i} = (\mathbf{p_i} + \mathbf{k}_{\mathbf{i} \bmod \mathbf{L}}) \bmod \mathbf{26} \qquad \mathbf{p_i} = (\mathbf{c_i} - \mathbf{k}_{\mathbf{i} \bmod \mathbf{L}}) \bmod \mathbf{26}$$

Substitution Ciphers

Vigenère Cipher - Cryptanalysis

The cipher is still easy to break with a *Ciphertext Only Attack* using statistics on the letter frequencies

- Find the length of the keyword **L** through the **Kasisky Test** (see next slide)
- Split up the ciphertext into L sequences of letters (one sequence for each keyword letter): each sequence is computed by aligning letters corresponding to the same “shift ciphertext”
- Apply frequency analysis to each “shift ciphertext”
- Use the retrieved *shift cipher* keys to derive the value of the *Vigenère keyword*

Substitution Ciphers

Kasisky Test for Finding Vigenère Key Length

- In 1853 Major F. Kasisky proposed a method to find the length of the keyword employed to encrypt messages with the Vigenère cipher
- **Key observation**: Two identical segments of $2 \leq l \leq L$ plaintext letters (l-gram), will be encrypted to the same sequence of l ciphertext letters
- **Deduction**: The distance d between two repeated sequences of l chars in the ciphertext may suggest a multiple of the key length L
- Operatively we need to:
  - Search for pairs of identical segments of length at least 3 (i.e., pair of 3-grams in the ctx message)
  - Record the distances between the occurrences of two identical segments
  - Compute the keyword length L as the *greatest common divisor* of the distances
  - note: some segments may be identical only out of chance

Cryptanalysis of Vigenère Cipher (1/5)

Given the following Vigenère ciphertext:

```
CJ UT WFCN LTTF VF AAHGKEE DNH VYC IPSPKGTMV EVLINFA, NC HXS SLGIX QNVYGEM VYYI
MVG ZLUHFORRXHB-RIMRXGUZLV TBF KCAXQQD- KJGWERRXHBU ICKHZWKGDGG PFU JGRGIUPR
KKCJ RHBVZLJX JKXMGHIUCW XGHQ KFT MKGERN-YWTJR. LX WPKCGTQV RLS MFCEQPVH DP
BXKSEKGCZ TNFAZL CH UGVBHCC NPVYGKQ IHKCIBH XOEY MIAST KFGHIIY ANUSTJNPVS,
ERPGRWPX JDOS PFRTL, RKXGITZ ERQW, TBF JCRKSV TMGICTRRT WCELKTGHU. FSG ISTJMCTZ
CEB TVCPFKXV ZKMCH KSNP KDKS CEB BHFG FL DNF CSGABHA KM AXH ULAW XHJVPTTZ ERPG-
BST GGVXCPJ KTWWCKC PM O FZQITBEV UWTH YV SHXR VF BD PWVY DPVS-VF-DPVS OVCIBBIJ,
NPIST UMRNAGERH, TBF R DXKA JRLSLVCBC. JGTQIRJGOVVJN, MVG KCRABKTYA PWBRPSKM
GEYQEWPX PTFCVV ADEZCSMGTHKFLH BG HFSCWSF FL QKCCUAPLHKEE TOSTPRWBBI RQ HX-
EWVLRXG QW XTKCU RLS HBGJ RWTH QECH HKP UMV PCWCBCM FGTMVGWBV. UWTH KJ RD
WWUKGCZIKJF P WWIZRPE RQCJPK KJVL XM WU RQ TTGKCW GXDTFBJVWDCC PL HJV QEHYGE
UDKR? JFU SH KG TMCOSTJC EKWXRRTEM YYCC XJGIW HRZNRZAX WU SMJGQGU MUY O
URRTEZKKC PGR UDCPKSF FTTK OP VLIBFG TMCMWPVLI?
```

Cryptanalysis of Vigenère Cipher (2/5)

First step:

We need to derive the keyword length L, thus we need to look for repeated sequences of characters keeping in mind that the keyword will not be too short nor too long (it must be easy to memorize). For example, in English texts repetitions of the same bigrams is quite common; and these are likely to match up to the same two letters in the Vigenère keyword every so often. Subsequently by looking for the distance between two repeated sequences we can guess L

- Each distance should be a multiple of the keyword
- Taking the **gcd** of all distances between sequences should give L

From the first sentence of the ciphertext we have:

```
CJ UT WFCN LTTF VF AAHGKEE DNH VYC IPSPKGTMV EVLINFA, NC HXS SLGIX QNVYGEM VYYI

MVG ZLUHFORRXHB-RIMRXGUZLV TBF KCAXQQD- KJGWERRXHBU ICKHZWKGDGG PFU JGRGIUPR

KKCJ RHBVZLJX JKXMGHI- UCW XGHQ KFT MKGERN-YWTJR
```

Distance between occurrences of RR: 30
Distance between occurrences of KG: 96, 46
gcd(30,96)=6, **gcd**(30,46)=2, **gcd**(96,46)=2.

Unlikely to have a keyword with L=2...Guess $L=6$

Cryptanalysis of Vigenère Cipher (3/5)

Second step:

We take every L-th (sixth) letter starting from the first letter and look at the statistics just as we did for a shift cipher to deduce the 1st letter of the keyword. Then we repeat the same procedure starting from the second letter, to find the 2nd letter of the keyword. And so on...

Freq. of the 1st keyword letter					
A	1.49	J	3.73	S	0.75
B	1.49	K	8.96	T	7.46
C	8.96	L	0.00	U	8.21
D	1.49	M	0.00	V	8.21
E	6.72	N	2.99	W	2.24
F	4.48	O	1.49	X	0.75
G	11.19	P	8.21	Y	1.49
H	1.49	Q	4.48	Z	0.00
I	2.99	R	0.75		

Freq. of the 2nd keyword letter					
A	0.00	J	8.21	S	2.24
B	0.75	K	9.70	T	3.73
C	5.22	L	2.24	U	3.73
D	1.49	M	0.75	V	10.44
E	7.46	N	1.49	W	0.75
F	11.19	O	0.00	X	2.99
G	0.75	P	2.24	Y	4.48
H	0.00	Q	0.00	Z	3.73
I	4.48	R	11.94		

Cryptanalysis of Vigenère Cipher (4/5)

Third step:

We now have $L{=}6$ shift ciphers to break!!!

We need a statistical test to compare each of the frequency distributions of our subsequences (e.g, the ones in the previous slides) to the ones of the English alphabet obtained as a shift-rotation for every possible shift (from 0 to 25).

Chi-squared Statistic

It assesses the similarity of two categorical probability distributions

$$\chi^2(\mathcal{E}, \mathcal{C}) = \sum_{i=0}^{25} \frac{(\mathcal{E}_i - \mathcal{C}_i)^2}{\mathcal{C}_i}$$

where $\mathcal{E}_i$ is the frequency of the i-th letter of the English alphabet, and $\mathcal{C}_i$ is the the frequency of the i-th letter of the considered shift cipher.

If the two distributions are identical, the chi-squared statistic is 0**, if the distributions are very different, some higher number will result.**

Following the aforementioned procedure, we find: keyword=CRYPTO

Cryptanalysis of Vigenère Cipher (5/5)

The decrypted message with spaces, punctuation, and capitalization is:

AS WE DRAW NEAR TO CLOSING OUT THE TWENTIETH CENTURY, WE SEE QUITE CLEARLY THAT THE INFORMATION-PROCESSING AND TELECOMMUNICATIONS REVOLUTIONS NOW UNDERWAY WILL CONTINUE VIGOROUSLY INTO THE TWENTY-FIRST. WE INTERACT AND TRANSACT BY DIRECTING FLOCKS OF DIGITAL PACKETS TOWARDS EACH OTHER THROUGH CYBERSPACE, CARRYING LOVE NOTES, DIGITAL CASH, AND SECRET CORPORATE DOCUMENTS. OUR PERSONAL AND ECONOMIC LIVES RELY MORE AND MORE ON OUR ABILITY TO LET SUCH ETHEREAL CARRIER PIGEONS MEDIATE AT A DISTANCE WHAT WE USED TO DO WITH FACE-TO-FACE MEETINGS, PAPER DOCUMENTS, AND A FIRM HANDSHAKE. UNFORTUNATELY, THE TECHNICAL WIZARDRY ENABLING REMOTE COLLABORATIONS IS FOUNDED ON BROADCASTING EVERYTHING AS SEQUENCES OF ZEROS AND ONES THAT ONES OWN DOG WOULDNT RECOGNIZE. WHAT IS TO DISTINGUISH A DIGITAL DOLLAR WHEN IT IS A S EASILY REPRODUCIBLE AS THE SPOKEN WORD? HOW DO WE CONVERSE PRIVATELY WHEN EVERY SYLLABLE IS BOUNCED OFF A SATELLITE AND SMEARED OVER AN ENTIRE CONTINENT?

Substitution Ciphers

Beale Cipher - another quite famous *hand cipher*

- A well-known variant of the Vigenère cipher
  - is based on the clue that "Longer the key lengths, lower the possibility of re-enciphering the same *d*-grams in the same way"
- **Beale's cipher** (a.k.a. **book cipher**): the keyword is taken as the first few words of a book that is agreed upon by the cipher users
  - Preserves readability/memorizability of the key, but enlarges the keyspace effectively
  - After the keyphrase is chosen, everything else works like the Vigenère cipher

Questions: What is the resistance of the Vigenère and Beal ciphers against "known/chosen plaintext attacks"?

Permutation Ciphers

Permutation Cipher

- The encryption transformation of this cipher consists of a **permutation** of **the positions** of the plaintext letters
- It is also called the Transposition Cipher
- The cipher key is a random permutation with length L, e.g., $L = 5$; $\pi = (1243) \in S_5$, that is :

$$\pi = (1243) = (1243)(5) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 1 & 3 & 5 \end{pmatrix}$$

- The key length L is kept secret

Permutation Ciphers

Example

- Take: $L = 5$, $\pi = (1243)$, plaintext “fun crypto”
- Remove all the spaces: “funcrypto”
- Split into chunks of L letters long: “funcr yptoP”
  the blocks are padded up (e.g., with ’P’) to obtain a number of letters which is a multiple of L
- Apply the permutation π to the plaintext, blockwise:
  “nfcur tyopP”
- The ciphertext is composed aligning all the encrypted chunks: “nfcurtyopP”

Note: **a permutation cipher is not equivalent to a substitution map** as each plaintext letter may corresponds to multiple ciphertext letters depending on the position

Permutation Ciphers

Pros and Cons

Pros:

- The keyspace can be quite large: ($\mathbf{L}$!)
- The cipher **does alter** the **d-grams** ($d \geq 2 \wedge d \leq L$) **frequency distributions** between plaintext and ciphertext message spaces
- **A ciphertext-only analysis (COA) is not effective**

Cons:

- The cipher **does not alter the single letter frequency** distributions
- A *Known Plaintext Analysis* (KPA) can easily reveal the key, if L is reasonably small

Permutation Ciphers

Cryptanalysis

- KPA: obtain the permutation asking for the encryption of an ordered string
- Example: given the ciphertext:

  coenunpaoteitmheewralsiatetglicralldlsdnwohwiatheb

- Ask one of the parties to encrypt the message

  abcdefghijklmnopqrstuvwxyz

- Obtaining the ciphertext

  cadbehfigjmknlorpsqtwuxvyz

- Exploit the known order of the plaintext to derive the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & \dots \\ 2 & 4 & 1 & 3 & 5 & 7 & 9 & 6 & 8 & 10 & 12 & 14 & 11 & 13 & 15 & \dots \end{pmatrix}$$

- The sequence repeats modulo 5 (thus $L = 5$). The key is (1243)

Affine Ciphers

The Hill Cipher

The Hill Cipher is a polyalphabetic cipher invented in 1929 by the mathematician Lester S. Hill for ease of automation

- Each letter of the alphabet is first encoded as a number: $A = 0$, $B = 1, \ldots, Z = 25$
- The cipher key K is thought as an $m \times m$ invertible matrix of numbers $\mod 26$
- A block of m plaintext letters P is then considered as a column vector with m components
- Encryption: $C{=}K\ P \mod 26$
- Decryption: $P{=}K^{-1}\ C \mod 26$

Affine Ciphers

Hill Cipher Cryptanalysis

Pros:

- The keyspace can be quite large: $\approx(26^{m^2})!$, possible matrixes (singular ones cannot be used)
- The cipher alters the frequency distributions of texts in a complex way
- A ciphertext-only analysis (COA) is not effective

Cons:

- A *K*nown Plaintext Analysis (KPA) can easily reveal the key by solving "linear" equations
- Given a single pair of plaintext-ciphertext (P, C) messages, the key is computed as: $KP=C \Rightarrow K=C\,P^{-1}$

Historical Ciphers

Lesson Learned

- The cipher key should be long enough to withstand bruteforce attacks
  - **Rule of thumb**: the key-space should have a size, which is encoded with at least 80-bit (for currently off-the-shelf computing machines)
- The mapping between ptx and ctx letters, in the definition of the encryption/decryption transformation, should not be the same for every occurrence of the same ptx letter (i.e., it should be position dependent – a combination of substitution and permutation may be of help to avoid frequency based COAs)
- A "linear" mapping (or any other efficiently invertible relation) between ptx and ctx are threaten by KPAs

Frequency attacks exploit the redundancy of the English language

- lossless compression before encryption removes redundancy !!!
- employing an uncommon (or dead) natural language may also help !!!

Perfect Secrecy – Information Theoretic Security

Is it possible to achieve perfect secrecy?

- A perfectly secret cipher should be unbreakable *regardless* of the (computational) effort thrown at it
- This implies that the ciphertext alone provides no information (no clue) to an attacker
- Claude Shannon proved the existence of such a cryptoscheme in its *Communication Theory of Secrecy Systems* paper (1945)
- Every other possible scheme claiming perfect secrecy is either a scam, or a (possibly unnecessary) complication of this one

A perfectly secure cipher is proven to be resistant to COAs, KPAs, and CCA(2)s

Information Theoretic Security

Given a generic symmetric cipher

$$\langle \mathcal{A}, \mathcal{M}, \mathcal{K}, \mathcal{C}, \{\mathbb{E}_k(), k \in \mathcal{K}\}, \{\mathbb{D}_k(), k \in \mathcal{K}\} \rangle$$

The attacker can analyze an arbitrary number of (chosen) ptx-ctx pairs
Basic Assumptions

- Each item in $\mathcal{M}, \mathcal{K}, \mathcal{C}$ is modeled as a random variable P, K, C, respectively, with certain probability distributions: $\Pr(P = m), \Pr(K = k), \Pr(C = c)$, with $m \in \mathcal{M}$, $k \in \mathcal{K}$ and $c \in \mathcal{C}$
- The two variables P **and** K **are statistically independent**, i.e.: the user does not employ any criterion in picking a key to encrypt a given message

Since K and P are independent random variables, the probability to observe a certain ctx can be written as follows (n.b. different keys may transform different ptx into the same ctx):

$$\Pr(C = c) = \sum_{k : c \in \{\mathbb{E}_k(m), \forall m \in \mathcal{M}\}} \Pr(K = k) \Pr(P = \mathbb{D}_k(c))$$

Information Theoretic Security

Consider the following toy symmetric-key cipher, everything about $\mathcal{M}$, $\mathcal{C}$, $\mathcal{K}$ is known.

$\mathcal{M}=\{a, b, c, d\}$; $\mathcal{C}=\{1, 2, 3, 4\}$; $\mathcal{K}=\{k_1, k_2, k_3\}$;

$\Pr(P=m)$, $m\in\mathcal{M}$

a	b	c	d
0.25	0.3	0.15	0.3

$\Pr(K=k)$, $k\in\mathcal{K}$

$\mathbf{k_1}$	$\mathbf{k_2}$	$\mathbf{k_3}$
0.25	0.5	0.25

$c=E_k(m)$, $k\in\mathcal{K}$, $m\in\mathcal{M}$

	a	b	c	d
$\mathbf{k_1}$	3	4	2	1
$\mathbf{k_2}$	3	1	4	2
$\mathbf{k_3}$	4	3	1	2

Prob. of observing a certain ctx:

$\Pr(C=1)=\Pr(K=k_1)\cdot\Pr(P=d)+\Pr(K=k_2)\cdot\Pr(P=b)+\Pr(K=k_3)\cdot\Pr(P=c)=0.2625$

$\Pr(C=2)=0.2625$, $\Pr(C=3)=0.2625$, $\Pr(C=4)=0.2125$

... ctxs are distributed almost uniformly

Prob. of observing a certain ctx, knowing "a priori" the ptx value:

$$\Pr(C=c|P=m)=\sum_{k:m=\mathbb{D}_k(c)}\Pr(K=k)$$

$\Pr(C=c|P=m)$

	a	b	c	d
1	0	0.5	0.25	0.25
2	0	0	0.25	0.75
3	0.75	0.25	0	0
4	0.25	0.25	0.5	0

Information Theoretic Security

When we try to break a cipher we are interested in the conditional probability of guessing the ptx value, knowing the value of the ctx, i.e.: $\Pr(P{=}m|C{=}c)$

$$\Pr(P = m|C = c) = \frac{\Pr(P = m)\Pr(C = c|P = m)}{\Pr(C = c)}$$

$\Pr(P = m|C = c)$

	1	**2**	**3**	**4**
a	0	0	0.71	0.29
b	0.57	0	0.29	0.35
c	0.14	0.14	0	0.35
d	0.29	0.86	0	0

For the toy cipher in this small example, we note that the knowledge of the ciphertext reveals a lot of information:

- if we know $C{=}1$, then the ptx is $\neq a$
- if we know $C{=}2$, then the ptx is $\neq a$, $\neq b$
- if we know $C{=}3$, then the ptx is $\neq c$, $\neq d$
- if we know $C{=}4$, then the ptx is $\neq d$

Perfect Secrecy

Definition

A symmetric-key cryptosystem is *Perfectly Secure* if the ciphertext does not reveal any information about the plaintext

$$\Pr(P = m | C = c) = \Pr(P = m), \quad \forall m \in \mathcal{M}, c \in \mathcal{C}$$

Lemma

A symmetric-key cryptosystem is *Perfectly Secure* if the plaintext does not reveal any information about the ciphertext:

$$\Pr(C = c | P = m) = \Pr(C = c), \quad \forall m \in \mathcal{M}, c \in \mathcal{C}$$

Proof.

$$\Pr(C{=}c|P{=}m){=}\Pr(C{=}c) \;\Leftrightarrow\; \frac{\Pr(P{=}m|C{=}c)\Pr(C{=}c)}{\Pr(P{=}m)}{=}\Pr(C{=}c) \;\Leftrightarrow\; \Pr(P{=}m|C{=}c){=}\Pr(P{=}m)$$

Perfect Secrecy

Lemma

Given a Perfectly Secure symmetric key cryptosystem, the following conditions hold

$$|\mathcal{K}| \geq |\mathcal{C}| \geq |\mathcal{M}|$$

Proof

1. $|\mathcal{C}| \geq |\mathcal{M}|$ since the encryption must be injective (i.e. distinct ptxs map into distinct ctxs)
2. $\Pr(C = c) > 0 \, \forall c \in \mathcal{C}$, otherwise we can alter the definition of the ctx space; A perfectly secure cipher means $\Pr(C{=}c){=}\Pr(C{=}c|P{=}m) > 0$ therefore $\forall\, m \in \mathcal{M}, c \in \mathcal{C}$ there must be at least one $k \in \mathcal{K}$ s.t. $\mathbb{E}_k(m) = c$ which in turn means that $|\mathcal{K}| \geq |\mathcal{C}|$

Perfect Secrecy

Theorem (C. Shannon)

Let

$$\langle \mathcal{A}, \mathcal{M}, \mathcal{K}, \mathcal{C}, \{\mathbb{E}_k(), k \in \mathcal{K}\}, \{\mathbb{D}_k(), k \in \mathcal{K}\} \rangle$$

denote a symmetric key cryptosystem where the keys are picked independently of plaintexts values and $|\mathcal{K}| = |\mathcal{C}| = |\mathcal{M}|$

The cryptosystem is perfectly secure *iff*

(*i*) *every key is used with probability* $\frac{1}{|\mathcal{K}|}$

(*ii*) $\forall\,(m, c) \in \mathcal{M} \times \mathcal{C}$ *there is a* **unique** *key* $k \in \mathcal{K}$ *s.t.* $\mathbb{E}_k(m) = c$

Perfect Secrecy (Shannon's Theorem)

Proof. *if* part

Hp: $|\mathcal{K}| = |\mathcal{C}| = |\mathcal{M}|$, *(i)* every key is s.t. $\Pr(K = k) = 1/|\mathcal{K}|$,

$$\textit{(ii)}\ \forall\, m \in \mathcal{M}, c \in \mathcal{C}\, \exists\, !\, k \in \mathcal{K} \text{ s.t. } c = \mathbb{E}_k(m)$$

Th: The system is *perfectly secure*: $\Pr(P = m|C = c) = \Pr(P = m)$

Considering a given ctx c and keeping in mind that P and K are independent

$$\Pr(C = c) = \sum_{k : c \in \{\mathbb{E}_k(m), \forall m \in \mathcal{M}\}} \Pr(K = k) \Pr(P = \mathbb{D}_k(c)) =$$

$$\overset{\text{using}\textit{(i)}}{=} \frac{1}{|\mathcal{K}|} \sum_k \Pr(P = \mathbb{D}_k(c)) \overset{\text{using}\textit{(ii)}}{=} \frac{1}{|\mathcal{K}|} \sum_m \Pr(P = m) = \frac{1}{|\mathcal{K}|}$$

Now, the following equalities hold:

$$\Pr(P = m|C = c) = \frac{\Pr(C = c|P = m) \Pr(P = m)}{\Pr(C = c)} = \frac{1/|\mathcal{K}| \Pr(P = m)}{1/|\mathcal{K}|} \quad \textbf{cvd.}$$

Perfect Secrecy (Shannon's Theorem)

Proof. *only if* part

Hp: The system is *perfectly secure*: $|\mathcal{K}| = |\mathcal{C}| = |\mathcal{M}|$, $\Pr(P = m|C = c) = \Pr(P = m)$
Th: *(i)* every key is s.t. $\Pr(K = k) = 1/|\mathcal{K}|$
(ii) $\forall\, m \in \mathcal{M}, c \in \mathcal{C}\, \exists\,!\, k \in \mathcal{K}$ s.t. $c = \mathbb{E}_k(m)$

- Given an arbitrary pair (m, c): at least one k^* for mapping $m \mapsto c$ exists, otherwise $\Pr(P{=}m|C{=}c){=}0$, $\Pr(P{=}m){\neq}0$, thus denying the Hp.
  At most one k^* for mapping $m \mapsto c$ exists. Indeed, if two keys k_1, k_2 satisfy $\mathbb{E}_{k_1}(m){=}\mathbb{E}_{k_2}(m){=}c$, then since there are only as many keys as ctxs, i.e.: $|\mathcal{K}|{=}|\mathcal{C}|$, there must be another ctx $c'{\in}\mathcal{C}$ with $\mathbb{E}_k(m){\neq}c'$ $\forall k$ and the mapping $m \mapsto c'$ would be impossible $\forall k{\in}\mathcal{K}$. So this verifies *(ii)*.
- Given a ctx c, label as k_i the key values employed for $m_i \mapsto c$, $1{\leq}i{\leq}|\mathcal{K}|$. From the Hp. of perfect secrecy

$$\Pr(\mathbf{P} = \mathbf{m_i}) = \Pr(P = m_i|C = c) = \frac{\Pr(C = c|P = m_i)\Pr(P = m_i)}{\Pr(C = c)} =$$

$$\overset{\exists!k \text{ s.t. } m_i \text{ is mapped to } c}{=} \frac{\Pr(\mathbf{K} = \mathbf{k_i})\Pr(\mathbf{P} = \mathbf{m_i})}{\Pr(\mathbf{C} = \mathbf{c})} \Leftrightarrow$$

simplifying the first and last members of this chain of equalities...

$$\Leftrightarrow \Pr(\mathbf{C} = \mathbf{c}) = \Pr(\mathbf{K} = \mathbf{k_i}) \overset{\text{being true } \forall\, \mathbf{i},\ \mathbf{1} \leq \mathbf{i} \leq |\mathcal{K}|}{\Longrightarrow} \Pr(\mathbf{K} = \mathbf{k_i}) = \mathbf{1}/|\mathcal{K}|, \forall\, \mathbf{i}, \qquad \textbf{cvd.}$$

Vernam Cipher

- In 1919 Eng. Gilbert S. Vernam patented a telegraphic device able to encrypt the symbols (in Baudot encoding – i.e., 5-bit ASCII) typed in by an operator
- The ciphertext was composed through a bitwise eXclusive-OR with a sequence of symbols provided on a paper tape (the key) having the same length of the input message (i.e., the plaintext)
- US army Gen. Joseph Mauborgne proposed to employ a distinct paper tape (key value) for each ptx (condition (ii) of Shannon's Th.) containing random information (condition (i) of Shannon's Th.).

  This idea combined with the Vernam's `xor`-ing machine became known as the *One-Time-Pad* (OTP) enciphering machine

Vernam Cipher

The aformentioned OTP system employed with binary keys and messages is the most effective implementation of a perfectly secure cryptoscheme

- The premises of Shannon's theorem hold:
  1. the ptx space $\mathcal{M}$, the ctx space $\mathcal{C}$ and the key space $\mathcal{K}$ all have the same size. Even better: $\mathcal{M}=\mathcal{C}=\mathcal{K}=\{0,1\}^L$, $L>0$.
  2. the key value is independent from the ptx value
- the necessary and sufficient conditions of the theorem hold:
  1. The key value is randomly generated. Thus, every key is actually used with probability $1/|\mathcal{K}|$
  2. The XOR operation is a **modulo 2 addition** between elements in $\{0,1\}$; the operation defines a trivial *algebraic group* over $\{0,1\}$, then for each ptx-ctx pair (m,c) there is a **unique** key s.t. $m \oplus k=c$

Vernam Cipher

OTP

The key management of a OTP cipher is a pain!
Choosing a key with uniform probability over $\mathcal{K}$, $\Pr(K{=}k){=}\frac{1}{|\mathcal{K}|}$, means that

- each key must be used only once
- each key must have the same length of the ptx
- each key must be truly random (i.e., unpredictable)

Moreover, the key pad must be available to both communication parties and used taking care of employing the keys in the same order

- If the sender re-uses always the same key, an attacker can ask to cipher a known message m and recover the key as: $k{=}c{\oplus}m$ (Known Ptx Attack)
- if the same key is used twice for encrypting two different messages, some infos about the ptxs is leaked from the ctxs as:
  if $c_1 = m_1 \oplus k$, $c_2 = m_2 \oplus k$ then $(c_1 \oplus c_2){=}(m_1 \oplus m_2)$

Vernam Cipher

Modified Shift Cipher

Consider the following (less efficient) variant of OTP cipher

- $\mathcal{M}, \mathcal{C}, \mathcal{K}$: strings composed by 5 letters of the English alphabet
- Identify each letter with a number: A=0, B=1, C=2, ...
- Given a ptx $(m_0, m_1, m_2, m_3, m_4)$ and a randomly chosen key $(k_0, k_1, k_2, k_3, k_4)$ with $0 \leq k_i \leq 25$, the encryption transformation computes the ctx as: $\mathbf{c_i = m_i + k_i \mod 26}$

This system is *perfectly secure* ($\mathcal{M}=\mathcal{C}=\mathcal{K}=\{0, \ldots, 25\}^5$; ptxs and keys are independent):

- each key is chosen with uniform probability $\Rightarrow \Pr(K = k) = \frac{1}{26^5}$
- for each ptx-ctx pair $\langle m, c \rangle$ there is a unique key: $k_i = (c_i - m_i) \mod 26, \forall\, i$

If the encrypted messages are no more than 26^5 (i.e., each key is used only once), then the attacker cannot learn anything as the decryption of a given ctx, through employing an exhaustive search of the key, will lead to recover all the meaningful plaintext strings composed by 5 letters

Computationally secure ciphers

A practical approach

- The One-Time-Pad is used in practice only in scenarios where secrecy is paramount (military/diplomatic communications)
- The fallback is to design a computationally secure symmetric cipher
  - Note that, public-key schemes were introduced after Shannon's work ... and by construction they are *only* computationally secure
- Modern computationally secure ciphers are able to reuse a "small key" (128-256 bits) while:
  - Avoiding the disclosure of the plaintext from the ciphertext
  - Avoiding the disclosure of the key from plaintext-ciphertext pairs
- Both these conditions are warranted provided the attacker has not an unbounded computational power
- Practically, the computational limit is made so high that no realistic attacker is able to break the cipher (e.g., computing 2^{256} encryptions)

Computationally secure ciphers

We need to develop some elements of the information theory related to *computationally secure* ciphers.
The main results are due (again) to Shannon and his idea to measure the **Information** through the concept of **Entropy**.

- We will use the term "Information" as a synonym for *Uncertainty*:
  - if you are uncertain (or unaware) about the meaning of something, then revealing the meaning gives you fresh knowledge and hence information
- From the point of view of a crypto-analyst you want to find out the meaning of a ciphertext:
  - you could guess the plaintext
  - the level of uncertainty you have about either the (correct) plaintext or the (correct) key quantifies the amount of information leaked by the ctx

Entropy – Intuitions

- Consider a set with two elements $\mathcal{S}$={"Yes", "No"} including the answers Alice can give to the questions of Bob.
- Model any answer given by Alice as a sample of the random variable X taking values over $\mathcal{S}$.
- The level of uncertainty that Bob has about the value of the received answer is called *entropy of* X, denoted as $H(X)$, and measured in *bit*

- if Bob knows "a priori" that the answer of Alice to any question is "Yes", then there is no uncertainty about what he will hear as answers, thus the answers reveal to him no information, i.e., $H(X)=0$
- if Bob has no idea of the answer he will hear from Alice, then the probability of hearing "Yes" is the same of the one of hearing "No", thus he will learn 1 bit of information, i.e., $H(X)=1$

Entropy

Definition

Let X be a random variable which takes values in $\{x_1, x_2, \ldots x_n\}$ with probability distribution $p_i{=}\Pr(X = x_i)$, $\forall 1{\leq}i{\leq}n$. The Entropy of X is defined as:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

assuming conventionally that $p_i \log_2 p_i{=}0$, if $p_i{=}0$.

- $H(X){\geq}0$, and $H(X){=}0$ only if $p_j{=}1$, $p_i{=}0$ $\forall i$, $j{\neq}i$ $(\sum_i p_i{+}p_j{=}1)$
- if $p_i{=}\frac{1}{n}$ $\forall i$, then $H(X){=}\log_2 n$
- (Theorem) If X is a random variable over $\{x_1, \ldots, x_n\}$, it is alway true that $0{\leq}H(X){\leq}\log_2 n$

Another way to look at the Entropy is that it assesses by how much one can compress the representation of the information

Entropy

Considering the previous example of the toy symmetric-key cipher:

$\mathcal{M}=\{a, b, c, d\}$; $\mathcal{C}=\{1, 2, 3, 4\}$; $\mathcal{K}=\{k_1, k_2, k_3\}$;

$\Pr(P=m)$, $m\in\mathcal{M}$

a	**b**	**c**	**d**
0.25	0.3	0.15	0.3

$\Pr(K=k)$, $k\in\mathcal{K}$

$\mathbf{k_1}$	$\mathbf{k_2}$	$\mathbf{k_3}$
0.25	0.5	0.25

$c=E_k(m)$, $k\in\mathcal{K}$, $m\in\mathcal{M}$

	a	**b**	**c**	**d**
$\mathbf{k_1}$	3	4	2	1
$\mathbf{k_2}$	3	1	4	2
$\mathbf{k_3}$	4	3	1	2

$\Pr(C=c)$, $c\in\mathcal{C}$

1	**2**	**3**	**4**
0.26	0.26	0.26	0.22

$H(P)\approx 1.95$, $H(K)\approx 1.5$, $H(C)\approx 2.0$

- a ctx leaks about 2 bits of information about the key and the ptx
- it is interesting to find a method to assess how much information is about the key and how much of it is about the ptx

Entropy

Notable definitions

- Given two random variables X, Y over $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ the joint entropy considers their joint distribution:

$$H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} \Pr(X = x_i, Y = y_j) \log_2 \Pr(X = x_i, Y = y_j)$$

  and defines the amount of information you get observing a pair of values (x, y)

- Given two random variables X, Y over $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ the entropy of X given one observation $Y = y$ is:

$$H(X, Y = y) = -\sum_{i=1}^{n} \Pr(X = x_i | Y = y) \log_2 \Pr(X = x_i | Y = y)$$

- Given two random variables X, Y over $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ the conditional entropy of X given Y is:

$$H(X, Y) = -\sum_{j=1}^{m} \Pr(Y = Y_j) H(X | Y = y_j) =$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} \Pr(Y = Y_j) \Pr(X = x_i, Y = y_j) \log_2 \Pr(X = x_i | Y = y_j)$$

  defines the amount of information you get after a value of Y has been revealed

Entropy

Notable statements (...easy to prove)

- $H(X,Y) \leq H(X) + H(Y)$, the equality holds if X, Y are independent
- $H(X,Y) = H(Y) + H(X|Y)$;
- $H(X|Y) \leq H(X)$, the equality holds if X, Y are independent

Going back to a generic symmetric-key cipher

- $H(P|K,C) = 0$: if you know the ctx and the key you do not have any uncertainty in deriving the ptx
- $H(C|P,K) = 0$: if you know the ptx and the key you do not have any uncertainty in deriving the ctx
- $H(C,P,K) = H(P,K) + H(C|P,K) = H(P,K) = H(P) + H(K)$
- $H(C,P,K) = H(K,C) + H(P|K,C) = H(K,C)$
- the last two expressions enable us to write: $\mathbf{H}(\mathbf{K},\mathbf{C}) = \mathbf{H}(\mathbf{P}) + \mathbf{H}(\mathbf{K})$

Entropy

Key Equivocation

It defines the amount of information (uncertainty) about the key, that you got by the knowledge of a ctx: $H(K|C)$

- In a COA the goal is to find the correct key value looking at some ctxs
- $\mathbf{H}(\mathbf{K}|\mathbf{C}) = H(K, C) - H(C) = \mathbf{H}(\mathbf{P}) + \mathbf{H}(\mathbf{K}) - \mathbf{H}(\mathbf{C})$

Considering our previously introduced toy cipher:
$H(P){\approx}1.95$, $H(K){\approx}1.5$, $H(C){\approx}2.0$. $H(K|C){=}1.95{+}1.5{-}2.0{=}1.45$

The knowledge of any ctx leaves us with an uncertainty about the key of 1.45 bits, thus each ctx would allow us to learn 0.05 bits

- how to effectively use these 0.05 bits to rule out a certain subset of key values... is left to the attacker
- it would have been better to have a (theoretical) leakage from any ctx $<<0.05$ (possibly negligible)

Spurious Keys and Unicity Distance

- The redundancy of the natural language employed for the plaintext messages is of great help to the attackers
- For example, the following English sentence can be easily understood even if more than half of the characters is missing:

```
On** up** a t**e t**re **s a **rl **ll** Sn** Wh**e
```

- What about the Entropy per letter H_L of the English language?
  - $H_L \leq \log_2 26 = 4.70$ (the second member is the entropy per letter of a completely random string...)
  - keeping into account the actual frequencies of the English letters M $H_L \leq H(M) = 4.14$
  - Keeping into account the actual frequencies of English digrams M^2 $H_L \leq \frac{H(M^2)}{2} = 3.56$... an so on ... assessing $\frac{H(M^3)}{3}$, $\frac{H(M^4)}{4}$ etc...
- An approximation of the actual value is: $\mathbf{1.0 \leq H_L \leq 1.5}$

Spurious Keys and Unicity Distance

Definition of Language Redundancy

$$R_L = 1 - \frac{H_L}{\log_2 |\mathcal{M}|}$$

For English $|\mathcal{M}|$=26, $R_L \approx 0.75$.
This means that 10MB of English text may be encoded in 2.5MB.

Consider a cipher $\langle \mathcal{A}, \mathcal{M}, \mathcal{C}, \mathcal{K}, \{\mathbb{E}_e : e \in \mathcal{K}\}, \{\mathbb{D}_d : d \in \mathcal{K}\}\rangle$ where each ciphertext word is composed by n symbols, and $|\mathcal{M}|=|\mathcal{C}|$.
Taking a ctx $c \in \mathcal{C}$, let us denote as $\mathcal{K}_{\texttt{Meaningful}}(c)$ the set of keys which decrypt c in "meaningful" plaintexts.
Average number of Spurious keys:

$$\bar{s_n} = \sum_{c \in \mathcal{C}} \Pr(C = c)(\mathcal{K}_{\texttt{Meaningful}}(c) - 1)$$

Spurious Keys and Unicity Distance

As the length of the plaintext and ciphertext words increase ($n \to \infty$), after some approximations ... we can find that:

$$\bar{s_n} \geq \frac{|\mathcal{K}|}{|\mathcal{M}|^{nR_L}} - 1$$

- it is intuitive that the number of spurious keys decreases as the length of the ctx increases; while an attacker would like $\bar{s_n}=0$

Unicity Distance

It is the length of ciphertext words (i.e., the number of ctx) $n=n_0$ such that the number of spurious keys is equal to zero, i.e.: $\bar{s_n}=0$

$$n_0 \approx \frac{\log_2 |\mathcal{K}|}{R_L \log_2 |\mathcal{M}|}$$

Unicity Distance

- In a substitution cipher: $|\mathcal{M}|=26$, $|\mathcal{K}|=26!$, $R_L\approx 0.75$ the unicity distance is: $n_0\approx 25$
- In a generic modern symmetric-key cipher we may have that $|\mathcal{M}|=|\mathcal{C}|=|\mathcal{K}|=|\{0,1\}|^l$, for some bit length $l>>1$, while the ptx language is English.

  Assuming $R_L\approx 0.75$ (this is an under-estimate as we encode English letters with ASCII). The unicity distance is: $n_0\approx\frac{l}{R_L}=\frac{4l}{3}$.

  if we would be able to compress the plaintext in a perfect manner, to have $R_L\approx 0$, then $n_0\rightarrow\infty$
- Modern ciphers encrypt plaintexts with no redundancy ?
  The answer is no. As they usually add to the ptx some redundancy to counter active attacks, Known Plaintext Attacks or (Adaptive) Chosen Ciphertext Attacks.
- the considerations about the unicity distance are valid only if the threat model defines a passive attacker