Cryptography I

General concepts and some classical ciphers

- Basic concepts
- Attack models
- Classic ciphers: mono-alphabetic
- Vigenere cipher
- One-time-pad cipher

Security Goals

- Confidentiality (secrecy, privacy)
  - Assure that data is accessible to only one who are authorized to know
- Integrity
  - Assure that data is only modified by authorized parties and in authorized ways
- Availability
  - Assure that resource is available for authorized users

General tools

- Cryptography
- Software controls
- Hardware controls
- Policies and procedures
- Physical controls

What is Crypto?

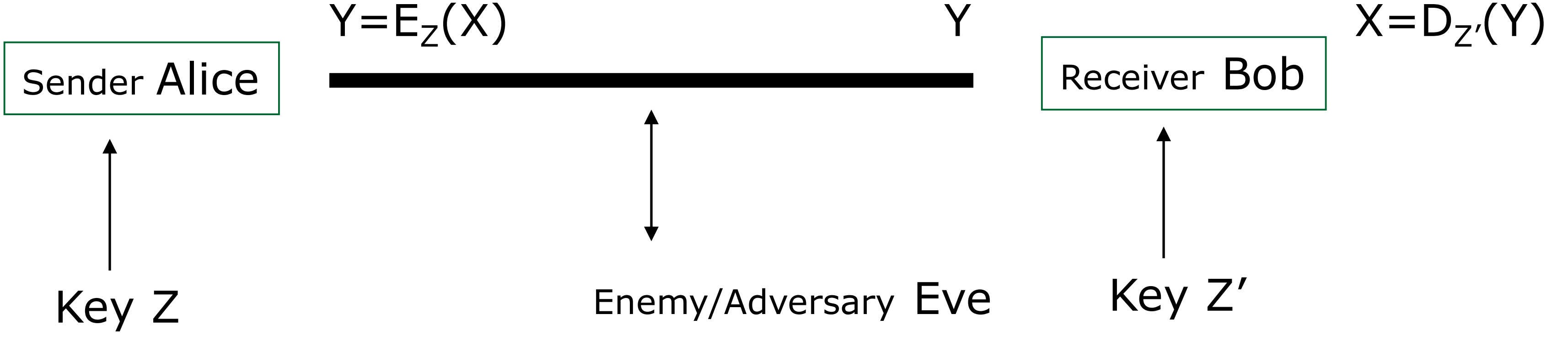
- Constructing and analyzing cryptographic protocols which enable parties to achieve security objectives
  - Under the present of adversaries.
- A protocol (or a scheme) is a suite of procedures that tell each party what to do
  - usually, computer algorithms
- Cryptographers devise and analyze protocols under Attack model
  - assumptions about the resources and actions available to the adversary
    - So, you need to think as an adversary

Terms

- **Cryptography:** the study of mathematical techniques for providing information security services.
- **Cryptanalysis:** the study of mathematical techniques for attempting to get security services breakdown.
- **Cryptology:** the study of cryptography and cryptanalysis.

Terms …

- plaintexts
- ciphertexts
- keys
- encryption
- decryption



Secret-key cryptography

- Also called: symmetric cryptography
- Use the same key for both encryption & decryption (Z=Z')
- Key must be kept secret
- Key distribution – how to share a secret between A and B very difficult

Public-key cryptography

- Also called: asymmetric cryptography
- Encryption key different from decryption key and
  - It is not possible to derive decryption key from encryption key
- Higher cost than symmetric cryptography

Is it a secure cipher system?

- **Why insecure**
  - **just break it under a certain reasonable attack model (show failures to assure security goals)**
- Why secure:
  - Evaluate/prove that under the considered attack model, security goals are assured
  - Provable security: Formally show that (with mathematical techniques) the system is as secure as a well-known secure one (usually simpler).

Breaking ciphers …

- There are different methods of breaking a cipher, depending on:
  - the type of information available to the attacker
  - the interaction with the cipher machine
  - the computational power available to the attacker

Breaking ciphers …

- **Ciphertext-only attack**:
  - The cryptanalyst knows **only the ciphertext**.
  - Goal: to find the plaintext and the key.
  - NOTE: such vulnerable is seen completely insecure
- **Known-plaintext attack**:
  - The cryptanalyst knows **one or several pairs of ciphertext and the corresponding plaintext.**
  - Goal: to find the key used to encrypt these messages
    - or a way to decrypt any new messages that use the same key (although may not know the key).

Breaking ciphers …

- **Chosen-plaintext attack**
  - The cryptanalyst **can choose a number of messages and obtain the ciphertexts for them**
  - Goal: deduce the key used in the other encrypted messages or decrypt any new messages (using that key).
- **Chosen-ciphertext attack**
  - Similar to above, but the cryptanalyst **can choose a number of ciphertexts and obtain the plaintexts.**
- Both can be **adaptive**
  - The choice of ciphertext may depend on the plaintext received from previous requests.

Models for Evaluating Security

- **Unconditional (information-theoretic) security**
  - **Assumes that the adversary has unlimited computational resources**.
  - Plaintext and ciphertext modeled by their distribution
  - Analysis is made by using probability theory.
  - For encryption systems: **perfect secrecy**, observation of the ciphertext provides no information to an adversary.

Models for Evaluating Security

- **Provable security:**
  - Prove security properties based on assumptions that it is difficult to solve a well-known and supposedly difficult problem (NP-hard …)
    - E.g.: computation of discrete logarithms, factoring
- **Computational security (practical security)**
  - Measures the amount of computational effort required to defeat a system using the best-known attacks.
  - Sometimes related to the hard problems, but no proof of equivalence is known.

Models for Evaluating Security

- **Ad hoc security (heuristic security):**
  - Variety of convincing arguments that every successful attack requires more resources than the ones available to an attacker.
  - Unforeseen attacks remain a threat.
  - **THIS IS NOT A PROOF**

Classic ciphers

Shift cipher (additive cipher)

- Key Space: [1 .. 25]
- Encryption given a key K:
  - each letter in the plaintext P is replaced with the K'th letter following corresponding number (shift right):
  - Another way: $Y = X \oplus K$ ➔ additive cipher
- Decryption given K:
  - shift left

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

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

P = CRYPTOGRAPHYISFUN

K = 11

C = NCJAVZRCLASJTDQFY

Shift Cipher: Cryptanalysis

- Easy, just do exhaustive search
  - key space is small (<= 26 possible keys).
  - once K is found, very easy to decrypt

We can think of similar ciphers

- Multiplicative Cipher

$Y = X \times Z \; mod \; 26$ OR $Y \equiv_{26} X \times Z$

How many possible keys?

- Affine Cipher

$$Y \equiv_{26} X \times Z + K$$

How many possible keys?

- Can you think of other ciphers?

General Mono-alphabetical Substitution Cipher

- The key space: all permutations of $\Sigma = \{A, B, C, \ldots, Z\}$
- Encryption given a key π:
  - each letter X in the plaintext P is replaced with $\pi(X)$
- Decryption given a key π :
  - each letter Y in the cipherext P is replaced with $\pi^{-1}(Y)$

- **Example:**

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

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

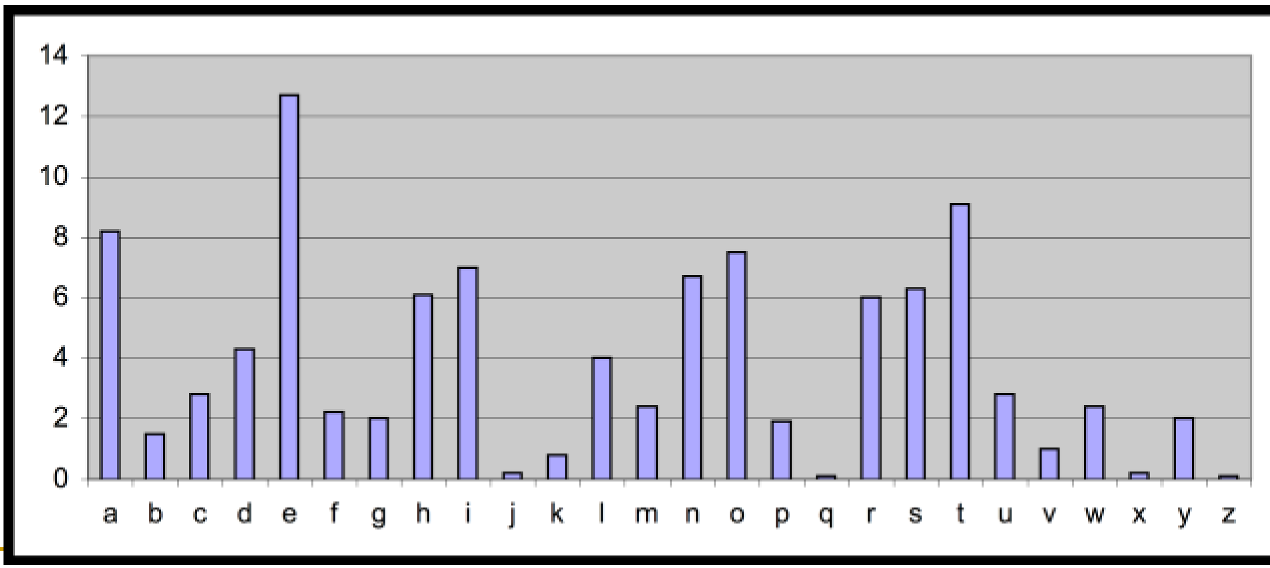
BECAUSE $\rightarrow$ AZDBJSZ

Looks secure, early days

- Exhaustive search is infeasible
  - key space size is $26! \approx 4*10^{26}$
- Dominates the art of secret writing throughout the first millennium A.D.
- Thought to be unbreakable by many back then

Cryptanalysis of Substitution Ciphers: Frequency Analysis

- Each language has certain features:
  - frequency of letters, or of groups of two or more letters.
- Substitution ciphers preserve the mentioned language features ➔ vulnerable to frequency analysis attacks



Substitution Ciphers: Cryptanalysis

- The number of different ciphertext characters or combinations are counted to determine the frequency of usage.
- The cipher text is examined for patterns, repeated series, and common combinations.
- Replace ciphertext characters with possible plaintext equivalents using known language characteristics.
- Example:

  THIS IS A PROPER SAMPLE FOR ENGLISH TEXT. THE FREQUENCIES OF LETTERS IN THIS SAMPLE IS NOT UNIFORM AND VARY FOR DIFFERENT CHARACTERS. IN GENERAL THE MOST FREQUENT LETTER IS **E** FOLLOWED BY A SECOND GROUP. IF WE TAKE A CLOSER LOOK WE WILL NOTICE THAT FOR BIGRAMS AND TRIGRAMS THE NONUNIFORM IS EVEN MORE.

  - Observations: f_X=1 và f_A=15.

- The letters in the English alphabet can be divided into 5 groups of similar frequencies

  I e

  II t,a,o,i,n,s,h,r

  III d,l

  VI c,u,m,w,f,g,y,p,b

  V v,k,j,x,q,z
- Some frequently appearing bigrams or trigrams

  Th, he, in, an, re, ed, on, es, st, en at, to

  The, ing, and, hex, ent, tha, nth, was eth, for, dth.

Example on cryptanalyzing a MS cipher

YKHLBA JCZ SVIJ JZB TZVHI JCZ VHJ DR IZXKHLBA VSS RDHEI DR YVJV LBXSKYLBA YLALJVS IFZZXC CVI LEFHDNZY EVBLRDSY JCZ FHLEVHT HZVIDB RDH JCLI CVI WZZB JCZ VYNZBJ DR ELXHDZSZXJHDBLXI JCZ XDEFSZQLJT DR JCZ RKBXJLDBI JCVJ XVB BDP WZ FZHRDHEZY WT JCZ EVXCLBZ CVI HLIZB YHVEVJLXVSST VI V HXXIKSJ DR JCLI HZXZBJ YZNZXDFEZBJ LB JZXCBDSDAT EVBT DR JCZ XLFCZH ITIJZEIJCVJ PZHZ DBXZ XDBILYXHZYIZKHZ VHZBDP WHZVMVWSZ

Letter:	A	B	C	D	E	F	G
Frequency:	5	24	19	23	12	7	0
Letter:	H	I	J	K	L	M	N
Frequency:	24	21	29	6	21	1	3
Letter:	O	P	Q	R	S	T	U
Frequyency:	0	3	1	11	14	8	0
Letter:	V	W	X	Y	Z		
Frequency:	27	5	17	12	45		

- $e \Rightarrow Z$

$f_j = 29$, $f_v = 27$

$f_{jcz} = 8 \rightarrow t \Rightarrow J$

$h \Rightarrow C$

- $a \Rightarrow V$

  (đứng riêng, mạo từ a)

J,V,B,H,D,I,L,C {t,a,o,i,n,s,h,r}

t,a *h*

JZB = te ? { teo, tei, ten, ter, tes } ➔ $n \Rightarrow B$

YKHLnA the SaIt ten TeaHI the aHt DR IeXKHLnA aSS RDHEI DR Yata LnXSKYLnA YLALtaS IFeeXh haI LEFHDNeY EanLRDSY the FHLEaHT HeaIDn RDH thLI haI Ween the aYNent DR ELXHDeSeXtHDnLXI the XDEFSeQLtT DR the RKnXtLDnI that Xan nDP We FeHRDHEeY WT the EaXhLne haI HLIen YHaEatLXaSST ***aI*** *a HXXIKSt DR thLI HeXent YeNeXDFEent Ln teXhnDSDAT EanT DR the XLFheH ITIteEIthat PeHe DnXe XDnILYXHeYIeKHe aHenDP WHeaMaWSe*

$e \Rightarrow Z,\ t \Rightarrow J,\ h \Rightarrow C,\ a \Rightarrow V,\ n \Rightarrow B$

{H,D,I,L} can be {o,i,s,r}

aI = a? {ao, ai, as, ar}

➔ *s => I*

Note:
UPPERCASE ~ cipher text
lowercase ~ plain text

key																										
	-	n	h	-	-	-	-	-	-	t	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	e
	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

YKHLnA the Sast ten TeaHs the aHt DR seXKHLnA aSS RDHEs DR Yata LnXSKYLnA YLALtaS sFeeXh has LEFHDNeY EanLRDSY the FHLEaHT HeasDn RDH thLs has Ween the aYNent DR ELXHDeSeXtHDnLXs the XDEFSeQLtT DR the RKnXtLDns that Xan nDP We FeHRDHEeY WT the EaXhLne has HLsen YHaEatLXaSST as a HXXsKSt DR ***thLs*** *HeXent YeNeXDFEent Ln teXhnDSDAT EanT DR the XLFheH sTsteEsthat PeHe DnXe XDnsLYXHeYseKHe aHenDP WHeaMaWSe*

{H,D,L} can be {o,i,r}

thLs = th?s {thos, this, thrs}

➔ *i => L*

key	-	n	h	-	-	-	-	-	s	t	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	e
	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

YKHinA the Sast ten TeaHs the **aHt** *DR seXKHinA aSS RDHEs DR Yata inXSKYinA YiAitaS sFeeXh has iEFHDNeY EaniRDSY the FHiEaHT HeasDn RDH this has Ween the aYNent DR EiXHDeSeXtHDniXs the XDEFSeQitT DR the RKnXtiDns that Xan nDP We FeHRDHEeY WT the EaXhine has* **Hisen** *YHaEatiXaSST as a HXXsKSt DR this HeXent YeNeXDFEent in teXhnDSDAT EanT DR the XiFheH sTsteEsthat PeHe DnXe XDnsiYXHeYseKHe aHenDP WHeaMaWSe*

{H,D} can be {o,r}

aHt = a?t {aot, art}

➔ *r => H, o => D*

key	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
	-	n	h	-	-	-	-	-	s	t	-	i	-	-	-	-	-	-	-	-	-	a	-	-	-	e

YKrinA the Sast ten Tears the art oR seXKrinA aSS RorEs oR Yata inXSKYinA YiAitaS sFeeXh has iEFroNeY EaniRoSY the FriEarT ***reason Ror this has Ween*** *the aYNent oR EiXroeSeXtroniXs the XoEFSeQitT oR the RKnXtions that Xan noP We FerRorEeY WT the EaXhine has risen YraEatiXaSST as a rXXsKSt oR* ***this reXent*** *YeNeXoFEent in teXhnoSoAT EanT oR the XiFher sTsteEsthat Pere onXe XonsiYXreYseKre arenoP WreaMaWSe*

reason Ror this has Ween ➔ *reason for this has been*

this reXent ➔ *this recent*

➔ *f => R, b => W, c => X*

key

-	n	h	o	-	-	-	r	s	t	-	i	-	-	-	-	-	-	-	-	-	a	-	-	-	e
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

YKrinA the Sast ten Tears the art of secKrinA aSS forEs of Yata incSKYinA YiAitaS sFeech has iEFroNeY EanifoSY the FriEarT reason for this has been the aYNent of EicroeSectronics the coEFSeQitT ***of the fKnctions*** *that can noP be FerforEeY bT the Eachine has risen YraEaticaSST as a rccsKSt of this recent YeNecoFEent in technoSoAT EanT* ***of the ciFher*** *sTsteEsthat Pere once consiYcreYseKre arenoP breaMabSe*

of the fKnctions → *of the functions*

of the ciFher → *of the cipher*

→ *u => K, p => F*

key																										
	-	n	h	o	-	-	-	r	s	t	-	i	-	-	-	-	-	f	-	-	-	a	b	c	-	e
	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

YurinA the Sast ten Tears the art of securinA aSS forEs of Yata incSuYinA YiAitaS speech has iEproNeY EanifoSY the priEarT reason for this has been the aYNent of EicroeSectronics the coEpSeQitT of the functions that can noP be perforEeY bT the Eachine has risen YraEaticaSST as a rccsuSt of this recent YeNecopEent in technoSoAT EanT of the cipher sTsteEsthat Pere once consiYcreYseure arenoP breaMabSe

YurinA the Sast ten Tears the art of securinA aSS ➔

during the last ten years the art of securing all

➔ *d => Y, g => A, l => S, y => T*

key																										
	-	n	h	o	-	p	-	r	s	t	u	i	-	-	-	-	-	f	-	-	-	a	b	c	-	e
	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

And the answer is

during the last ten years the art of securing all forms of data including digital speech has improved manifold the primary reason for this has been the advent of microelectronics the complexity of the functions that can now be performed by the machine has risen dramatically as a result of this recent development in technology many of the cipher systems that were once considered secure are now breakable

$f_P = 3$, $f_M = 1$

- P can be {j, k, q, z, w}
  - Pere = ?ere {jere, kere, qere, zere, were}. ➔ w => P
- M can be {j, k, q, z}
  - breaMable {breajable, breakable, breaqable, breazable} ➔ k => M

  $f_O = f_G = f_U = 0$ → *can not specify*

key																										
	g	n	h	o	m	p	-	r	s	t	u	i	k	v	-	-	x	f	l	y	-	a	b	c	d	e
	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

How can we design better ciphers?

- Observations:
  - A cipher system should not allow statistical properties of plaintext to pass to the ciphertext.
  - The ciphertext generated by a "good" cipher system should be statistically indistinguishable from random text.
- Idea for a stronger cipher (1460’s by Alberti)
  - use more than one cipher alphabet, and switch between them when encrypting different letters → Poly-alphabetic Substitution Ciphers
  - Developed into a practical cipher by Vigenère (published in 1586)

Vigenère cipher: a special Poly-alphabetic Substitution Cipher

- **Definition**:
  - Given m, a positive integer, $P = C = (Z_{26})^n$, and $K = (k_1, k_2, \ldots, k_m)$ a key, we define:
- **Encryption**:

  $e_k (p_1, p_2 \ldots p_m) = (p_1+k_1, p_2+k_2 \ldots p_m+k_m) \pmod{26}$
- **Decryption**:

  $d_k (c_1, c_2 \ldots c_m) = (c_1-k_1, c_2-k_2 \ldots c_m- k_m) \pmod{26}$
- **Example:**

  Plaintext: C R Y P T O G R A P H Y

  Key: L U C K L U C K L U C K

  Ciphertext: N L A Z E I I B L J J I

Vigenere Cipher: Cryptanalysis

- Find the length of the key
  - Kasiski method
  - Using Index of Coincidence (IC)
- Divide the message into that many shift cipher encryptions.
- Use frequency analysis to solve the resulting shift ciphers.

One-Time Pad

Key is chosen randomly

Plaintext $X = (x_1\ x_2\ \ldots\ x_n)$

Key $K = (k_1\ k_2\ \ldots\ k_n)$

Ciphertext $Y = (y_1\ y_2\ \ldots\ y_n)$

$e_k(X) = (x_1+k_1\ \ x_2+k_2\ \ldots\ x_n+k_n) \bmod m$

$d_k(Y) = (y_1-k_1\ \ y_2-k_2\ \ldots\ y_n-k_n) \bmod m$

Example

Plaintext space = Ciphtertext space =

Keyspace = $\{0,1\}^n$

Key is chosen randomly

For example:

Plaintext is	10001011
Key is	00111001
Then ciphertext is	10110010

Main points in One-Time Pad

- The key is never to be reused
  - Thrown away after first and only use
  - If reused ➔ insecure!
- One-Time Pad uses a very long key, exactly the same length as of the plaintext
  - In old days, some suggest choose the key as texts from, e.g., a book ➔ i.e. not **randomly chosen**
    - Not One-Time Pad anymore ➔ this does not have perfect secrecy as in true One-Time-Pad and can be broken
  - Perfect secrecy means key length be at least message length
    - **Difficult in practice**!

- Shift ciphers are easy to break using brute force attacks (eshautive key search)
- Substitution ciphers preserve language features (in N-gram frequency) and are vulnerable to frequency analysis attacks.
- Vigenère cipher are also vulnerable to frequency analysis once the key length is found.
  - In general poly-alphabetical substitution ciphers are not that secure
- OTP has perfect secrecy if the key is chosen randomly in the message length and is used only once.

Related course websites

https://users.soict.hust.edu.vn/vannk/AntoanThongtin/ComputerSecurity.htm