Cryptography I

General concepts and some classical ciphers

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?

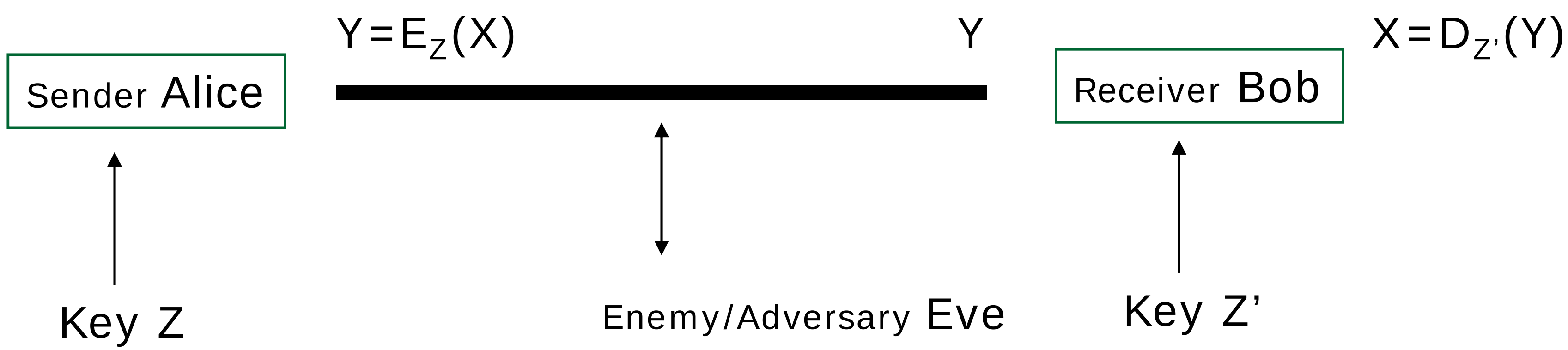
- Constructing and analyzing cryptographic protocols which enable parties to achieve security objectives
  - Under the present of adversaries.
- A protocol (or a scheme) is a suite of procedures that tell each party what to do
  - usually, computer algorithms
- Cryptographers devise and analyze protocols 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

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

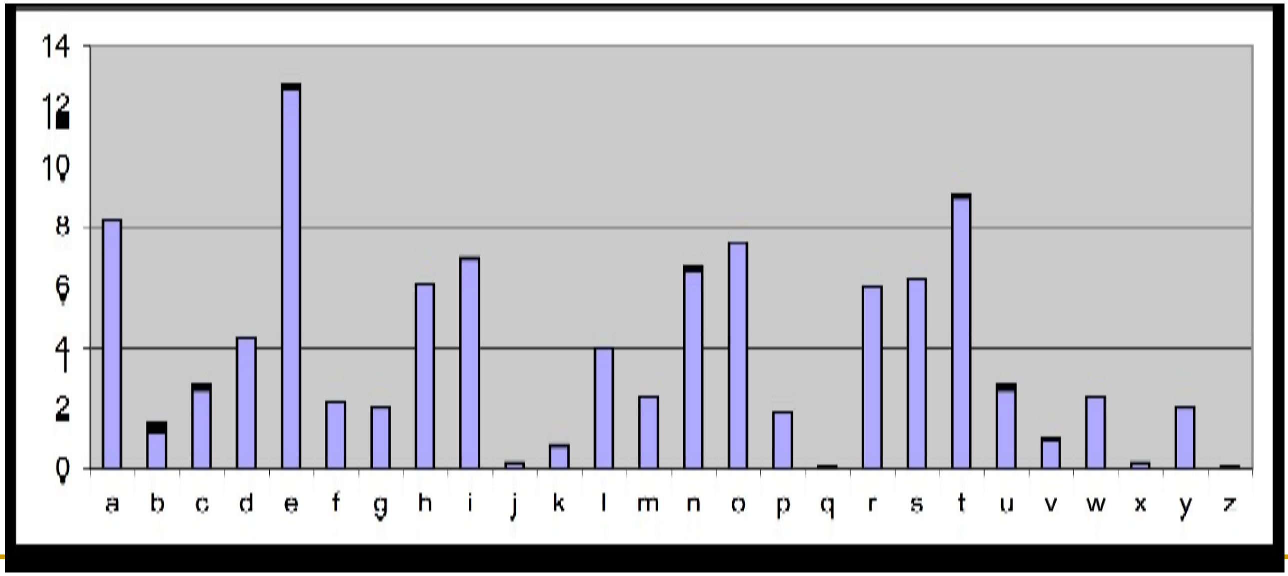
BECAUSE $\rightarrow$ AZDBJSZ

Looks secure, early days

- Exhaustive search is infeasible
  - key space size is 26! ≈ 4*1026
- 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 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

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*		

- $e \Rightarrow Z$

$f_j = 29, f_v = 27$

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

$h \Rightarrow C$

- $a \Rightarrow V$

  (article 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 } ➔ $n \Rightarrow B$

- Observations:
  - A cipher system should not allow statistical properties of plaintext to pass to the ciphertext.
  - The ciphertext ginerated by a "good" cipher systim should be satistically indistinguishable form 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 ➔ Polyalphabetic Substitution Ciphers
  - Developed into a practical cipher by Vigenère (published in 1586)

- **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.
- 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 $Y = (y_1\ y_2\ \ldots\ y_n)$

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

$d_k(Y) = (x_1-k_1\ \ x_2-k_2\ \ldots\ x_n-k_n) \bmod m$

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.