Classical Ciphers

Sushmita Ruj

Crypto Timeline

1900 BC
Earliest known
Instance of Cryptography

100 BC
Julius Caesar
Used ciphers

Classical Ciphers

1920's
Mechanical Machines:
Enigma

1939
World War II

1970
Modern Cryptography

Now

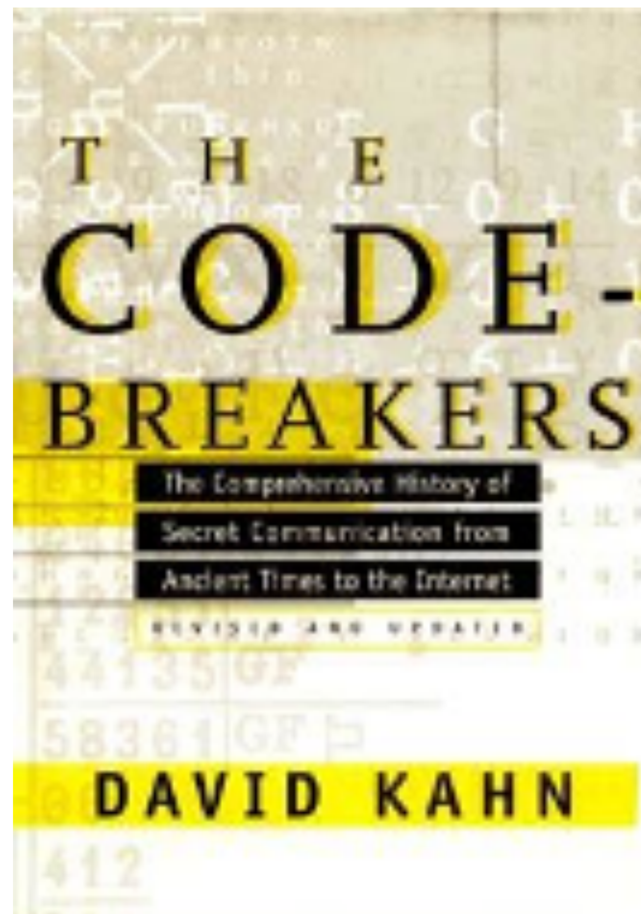
1970
IBM: DES
Data Encryption Standards

1976
Diffie Hellman
Key exchange

1977
RSA

2001
AES Standardized

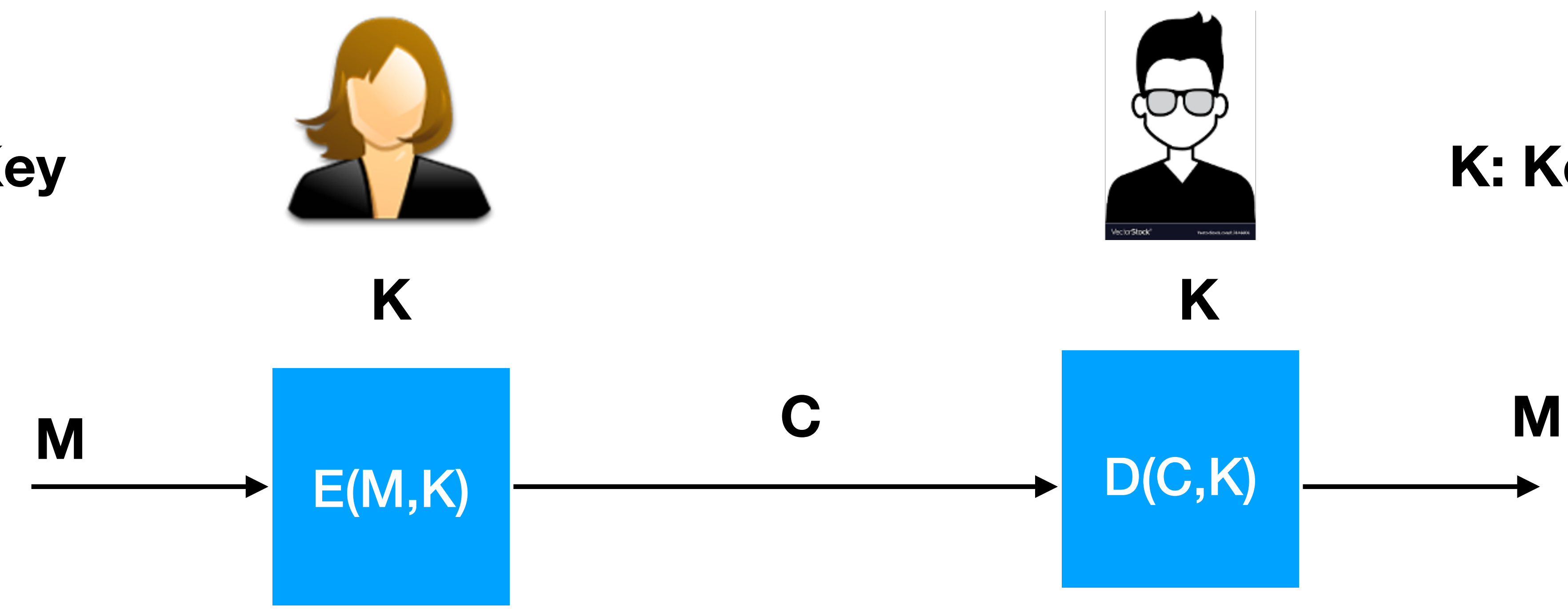
The Codebreakers



Classical Ciphers

K: Key

K: Key



E: Encryption Algorithm

D: Decryption Algorithm

E and D Very simple functions
Simple substitution, permutation

Simple Encryption: XOR

K=1001 K=1001

M = 1101 → [1001 / 1101] → C=0100 → [0100 / 1001] → M=1101

E: Encryption Algorithm : $M \oplus K$

D: Decryption Algorithm: $C \oplus K$

E and D Very simple functions
Simple substitution, permutation

Modular Arithmetic

- $E(M, K) = (M + K) \bmod n$
- $D(C,K) = (C - K) \bmod n$
- $M = 5, K = 10, n = 13, C = 2$
- $D(2, 10) = -8 \bmod 13 = 5$

Message Space

Previous Example

- Message Space: Set of all possible messages **$\{0,1\}^*$**
- Key Space: Set of all possible keys **$\{0,1\}^*$**
- Ciphertext Space: Set of all possible cipher texts **$\{0,1\}^*$**

Simple Ciphers: Shift Ciphers

- Message space $\mathscr{M}$, Key Space $\mathscr{K}$: Set of 26 English Alphabets. correspondence between alphabetic characters and residues modulo 26 as follows: $A \leftrightarrow 0, B \leftrightarrow 1, \cdots Z \leftrightarrow 25$
- $\mathscr{M} = \mathscr{C} = \mathscr{K} = \mathbb{Z}_{26}$
- K = 4 (say)
- C= E (M, K) = (M + K) mod 26
- D (C, K) = (C - K) mod 26
- CRYPTO -> ?

Caesar Cipher, K=3

Simple Ciphers: Substitution Ciphers

- Message space, Key Space: Set of 26 English Alphabets

K is the permutation π

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**...**
B	C	J	F	I	G	A	D	E	H	...

π^{-1}

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**...**
G	A	B	H	I	D	F	J	E	C	...

E: π **D: π^{-1}**

- M = “HEAD”

- C = ? **DIBF**

E(H,π) =π(H)=D

D(I,π) =π^{-1}(I)= E

Size of the key space = ?

Cryptanalyzing Substitution Cipher

- Most common letters in English
- E, T, A, I, N…
- From the given text find the frequency of each alphabet
- Map with English Alphabet
- Try this
- ZRTFT IH PQFTHZ IQ ZRT XBGBOZIO HTQBZT. HTWTFBG ZRLPHBQV HLGBF HYHZTSH RBWT VTOGBFTV ZRTIF IQZTQZILQH ZL GTBWT ZRT FTEPKGIO. ZRIH HTEBFBZIHZ SLWTSTQZ, PQVTF ZRT GTBVTFHRIE LD ZRT SYHZTFILPH OLPQZ VLLAP, RBH SBVT IZ VIDDIOPGZ DLF ZRT GISIZTV QPSKTF LD CTVI AQIXRZH ZL SBIQZBIQ ETBOT BQV LFVTF IQ ZRT XBGBJY. HTQBZLF BSIVBGB, ZRT DLFSTF NPTTQ LD QBKLL, IH FTZPFQIQX ZL ZRT XBGBOZIO HTQBZT ZL WLZT LQ ZRT OFIZIOBG IHHPT LD OFTBZIQX BQ BFSY LD ZRT FTEPKGIO ZL BHHIHZ ZRT LWTFMRTGSTV CTVI

Cryptanalyzing Substitution Cipher

- Or consider pairs of letters (diagrams)
- Or triples of letters….

Vigener Cipher

K = (2, 8, 15, 7, 4, 17)

T H I S C R Y P T O S Y S T E M I S N O T S E C U R E

C I P H E R C I P H E R C I P H E R C I P H E R C I P H

V P X Z G I A X I V W P U B T T M J P W I Z I T W Z T

Vigener Cipher

- Define $\mathscr{M} = \mathscr{C} = \mathscr{K} = (\mathbb{Z}_{26})^m$
- Let $(x_1, x_2, \cdots, x_m) \in \mathscr{M}$, $(y_1, y_2, \cdots, y_m) \in \mathscr{C}$
- For a key $K = (K_1, K_2, \cdots, K_m)$
- $E((x_1, x_2, \cdots, x_m), K) = (x_1 + k_1, x_2 + k_2, \cdots, x_m + k_m)$

  $=(y_1, y_2, \cdots, y_m)$
- $D((y_1, y_2, \cdots, y_m), K) = (y_1 - k_1, y_2 - k_2, \cdots, y_m - k_m)$

Affine Ciphers

$\mathscr{P} = \mathscr{C} = \mathscr{K} = \mathbb{Z}_{26}$ and let

$\mathscr{K} = \{(a, b) \in \mathbb{Z}_{26} \times \mathbb{Z}_{26} : gcd(a, 26) = 1\}$

For $K = (a, b) \in \mathscr{K}$, define

$y = E(x, K) = (ax + b) \bmod 26$

$D(y, K) = a^{-1}(y - b) \bmod 26$ where, $x, y \in \mathbb{Z}_{26}$

Eg: K = (7,3) , verify that this is correct.

Reading

- Stinson-Paterson, Chapter 2
- Extra Reading : Hill Cipher, Permutation Cipher

Mechanical Ciphers

- Rotor Machines: Enigma Machine
- To read on your own

Probability And Shannon's Theory

Security Notions

- Computational security means that specific algorithms to attack the cryptosystem are computationally infeasible (this requires knowing how much computational resources are available to the adversary)

- Provable security means that breaking the cryptosystem can be reduced (in a complexity-theoretic sense) to solving some underlying (assumed difficult) mathematical problem or breaking an underlying cryptographic primitive

- Unconditional security means that the cryptosystem cannot be broken, even with unlimited computational resources (because the adversary does not have enough information available to attack the system)

Notations & Definitions

Let U: finite set (e.g. $U = \{0,1\}^n$)

Def: **Probability distribution** P over U is a function $P: U \longrightarrow [0,1]$

such that $\sum P(x) = 1$

Examples:

1. Uniform distribution: for all $x \in U$: $P(x) = 1/|U|$
2. Point distribution at x_0: $P(x_0) = 1$, $\forall x \neq x_0$: $P(x) = 0$

Distribution vector: $\big($ P(000), P(001), P(010), … , P(111) $\big)$

Probability

- For a set $A \subseteq U : Pr[A] = \sum_{x \in A} P(x) \in [0,1]$
- The set A is called an **event**

Example: U = $\{0,1\}^8$

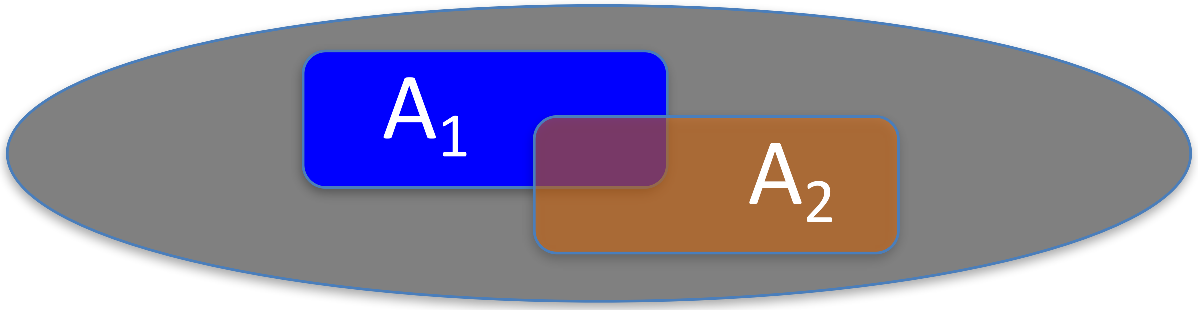
- A = $\{$ all x in U such that $lsb_2(x)=11$ $\}$ $\subseteq$ U

for the uniform distribution on $\{0,1\}^8$: Pr[A] = ?

Union Bound

- For events A_1 and A_2 $Pr[A_1 \bigcup A_2] \leq P[A_1] + P[A_2]$,
- if $A_1 \bigcap A_2 = \phi, Pr[A_1 \bigcup A_2] = P[A_1] + P[A_2]$

Example:



$A_1 = \{$ all x in $\{0,1\}^n$ s.t $lsb_2(x)=11$ $\}$;

$A_2 = \{$ all x in $\{0,1\}^n$ s.t. $msb_2(x)=11$ $\}$

$$Pr\left[\, lsb_2(x)=11 \text{ or } msb_2(x)=11 \,\right] = Pr\left[A_1 \cup A_2\right] \leq \tfrac{1}{4}+\tfrac{1}{4} = \tfrac{1}{2}$$

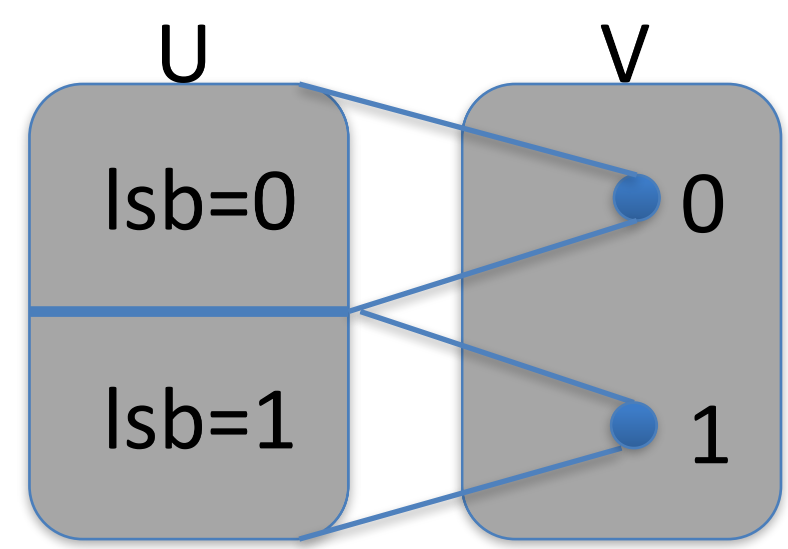
Random Variables

Def: A random variable X is a function $X : U \to V$

Eg: $X : \{0,1\}^n \to \{0,1\}$, X(y) = lsb(y) $\in\{0,1\}$

For the uniform distribution on U:

Pr[X=0] = 1/2 , Pr[X=1] = 1/2



More generally:

rand. var. X induces a distribution on V: Pr[X=v] := Pr$\left[\ X^{-1}(v)\ \right]$

Example

Let r be a uniform random variable on $\{0,1\}^2$

Define the random variable $X = r_1 + r_2$

Then $\Pr[X=2] = ¼$

Hint: $\Pr[X=2] = \Pr[\, r=11 \,]$

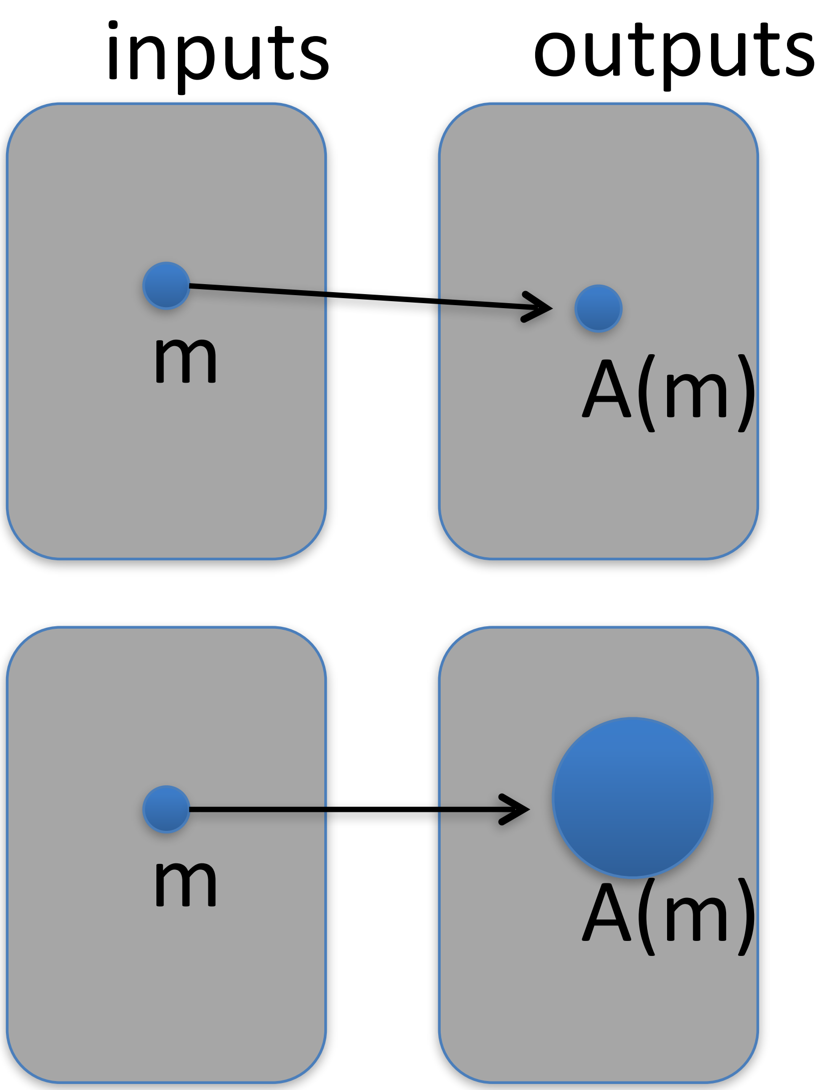
Randomised Algorithm

- Deterministic algorithm: $y \leftarrow A(m)$
- Randomized algorithm

  $y \leftarrow A(\, m \,;\, r\,)$ where $r \leftarrow \{0,1\}^n$

  output is a random variable

  $y \leftarrow A(\, m\,)$



Example: $A(m\,;\,k) = E(k, m)$, $y \leftarrow A(\, m\,)$

Independence

Def: Events A and B are **independent** if $\Pr[\text{ A and B }] = \Pr[A] \cdot \Pr[B]$

random variables X,Y taking values in V are **independent** if

$\forall a,b \in V$: $\Pr[\text{ X=a and Y=b}] = \Pr[X=a] \cdot \Pr[Y=b]$

Example: $U = \{0,1\}^2 = \{00, 01, 10, 11\}$ and $r \xleftarrow{R} U$

Define R.V. X and Y as: $X = \text{lsb}(r)$, $Y = \text{msb}(r)$

$\Pr[\text{ X=0 and Y=0 }] = \Pr[\text{ r=00 }] = 1/4 = \Pr[X=0] \cdot \Pr[Y=0]$

Property

<u>Thm</u>: Y is a R.V. over $\{0,1\}^n$, X an indep. uniform var. on $\{0,1\}^n$

Then $Z := Y \oplus X$ is uniform var. on $\{0,1\}^n$

If n =1, Pr[Z=0] = Pr[X=0,Y=0] + Pr[X=1,Y=1] = P0/2 +P1/2= 1/2

One-time-pad

Symmetric Ciphers

Def: A **Cipher** defined over a message space, key space and Ciphertext space is a pair of efficient algorithms (E,D), where,

$E : \mathscr{M} \times \mathscr{K} \to \mathscr{C}$ and $D : \mathscr{C} \times \mathscr{K} \to \mathscr{M}$,

Such that,

$$\forall m \in \mathscr{M} \text{ **and** } k \in \mathscr{K}, D(E(m,k)) = m$$

E is often randomised, D is always deterministic

One Time Pad

$\mathscr{M} = \mathscr{C} = \{0,1\}^n$ $\mathscr{K} = \{0,1\}^n$

$C = E(M, K) = M \oplus K, D(C, K) = C \oplus K$

$D = C \oplus K$

Eg: M = 0111100101

K = 1100100100

C = 1011000001

One Time Pad

$\mathscr{M} = \mathscr{C} = \{0,1\}^n$ $\mathscr{K} = \{0,1\}^n$ **Vernam 1971**

$$C = E(M, K) = M \oplus K, \quad D(C, K) = C \oplus K$$

Eg: M = 0111100101

K = 1100100100

M XOR K = C = 1011000001

Advantages =?

Simple, Fast,

Disadvantages=?

Key as large as message

What is a secure cipher?

Attacker's abilities: **CT only attack** (for now)

Possible security requirements:

attempt #1: **attacker cannot recover secret key**

attempt #2: **attacker cannot recover all of plaintext**

Shannon's idea: CT should reveal no "info" about plaintext

•

Information Theoretic Security (Shannon 1949)

- A cipher (***E*, *D***) over $(\underline{\mathscr{K},\ \mathscr{M},\ \mathscr{C}})$ has **<u>perfect secrecy</u>** if $\forall m_0, m_1 \in \mathscr{M}, len(m_0) = len(m_1)$ and $\forall c \in \mathscr{C}$
- $P[E(m_0, k) = c] = P[E(m_1, k) = c]$
- Where, k is chosen uniformly at random from $\mathscr{K}$ (meaning $k \xleftarrow{R} \mathscr{K}$)

OTP has perfect secrecy

- $\forall m, c, P[E(m,k) = c] = \dfrac{\text{no.of keys } k \in \mathscr{K}, \text{ such that } E(m,k) = c}{|\mathscr{K}|}$
- This number is constant (=1).
- Therefore OTP has perfect secrecy

Thank you