2550 Intro to cybersecurity

L8: Perfect Secrecy

abhi shelat/Ran Cohen

Thanks to Gil Segev (HUJI) for sharing slides

Basic Notation

- $x \in X$ means: the element x is in the set X
- Universal quantifier
  - $\forall x \in X$ means: for all elements x in the set X
  - E.g., given $X = \{1,2,3\}$ we can say $\forall x \in X$ it holds that $x < 4$
- Existential quantifier
  - $\exists x \in X$ means: there exists an element x in the set X
  - E.g., given $X = \{1,2,3\}$ we can say $\exists x \in X$ such that x is even

Basic Probability

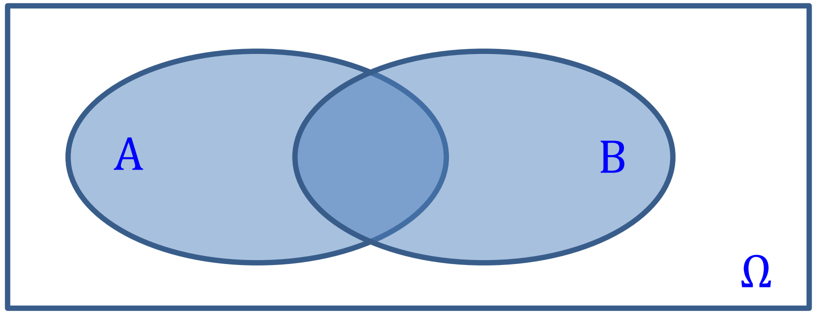
- A probability space Ω is a finite (or countable) set and a function $\Pr: \Omega \to [0,1]$ (the interval $0 \leq x \leq 1$) such that $\sum_{x \in \Omega} \Pr[x] = 1$
- An event is a subset of the probability space.
  The probability of an event $E \subseteq \Omega$ is defined as $\Pr[E] = \sum_{x \in E} \Pr[x]$
- Example: tossing a fair dice
  - Define $\Omega = \{1,2,3,4,5,6\}$ with the function $\Pr[x] = 1/6$
  - The probability of the event $E = \{2,4,6\}$ is $\Pr[E] = 1/2$
  - Unfair dice: define Ω as above with the function $\Pr[1] = 1/2$ and $\Pr[x] = 1/10$ for $x \in \{2,3,4,5,6\}$
  - In that case $\Pr[E] = 3/10$

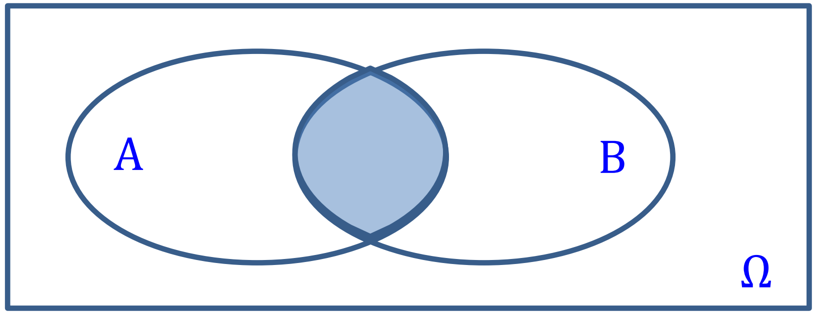
Basic Probability

A random variable is a function on the probability space $X: \Omega \to \mathbb{R}$

- Fair dice example: we can define random variable X as the result of the dice
  - $\Pr[X = 3] = 1/6$
  - $\Pr[X < 3] = 1/3$
- We can also define the random variable Y to be 0 if the result is even and 1 if it is odd. In this case
  - $\Pr[Y = 3] = 0$
  - $\Pr[Y < 3] = 1$
  - $\Pr[Y = 0] = \Pr[Y = 1] = 1/2$

Basic Probability





Given events A and B we can define

- Their union $A \cup B = \{x \mid x \in A \text{ } or \text{ } x \in B\}$
- Their intersection $A \cap B = \{x \mid x \in A \text{ } and \text{ } x \in B\}$

Conditional probability

- Fair dice: $\Pr[X = 3] = 1/6$ and $\Pr[Y = 1] = 1/2$
- What is the probability of getting 3 given that the result is odd?
- For events A, B with $\Pr[B] > 0$ we define the conditional probability of A given B as

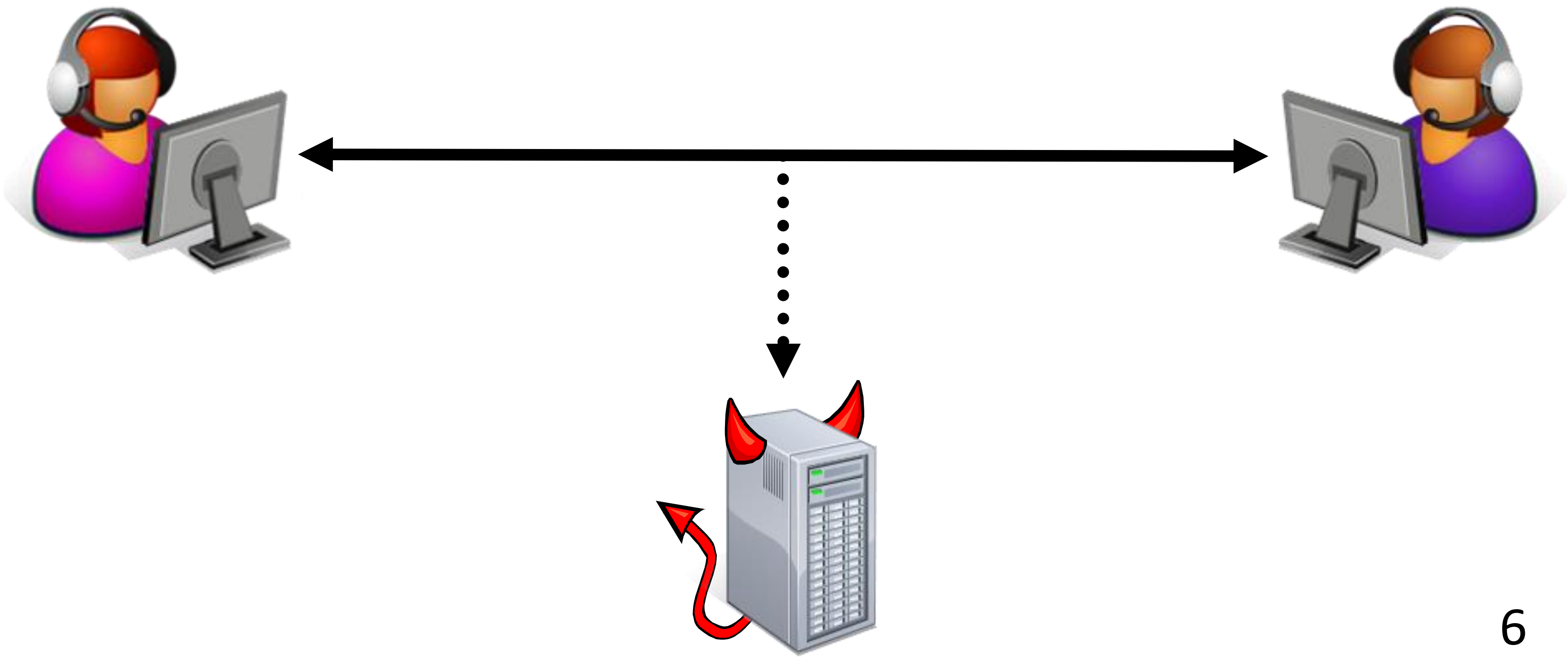
$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}$$

- $\Pr[X \cap Y] = 1/6 \Rightarrow \Pr[X \mid Y] = \frac{1/6}{1/2} = 1/3$

What is Cryptography?

Cryptography is an ancient art

- For many centuries focused exclusively on secret communication
- Consumers were military and intelligence organizations
- Relied on creativity and personal skill
- 500BC – 20th century: Design → break → repair → break → repair → ···



What is Cryptography?

Cryptography is an ancient art

- For many centuries focused exclusively on secret communication
- Consumers were military and intelligence organizations
- Relied on creativity and personal skill
- 500BC – 20^{th} century: Design → break → repair → break → repair → ⋯

Modern Cryptography: Cryptography as a science

- Radical change in the late 20^{th} century
- Much more than secret communication
- Used everywhere & consumed by everyone!
- Relies on rigorous threat models, firm foundations & proofs!

Outline

- **Symmetric-key encryption**
- **Some historical ciphers**
- **The basic principles of modern cryptography**
- **Perfect secrecy and its limitations**

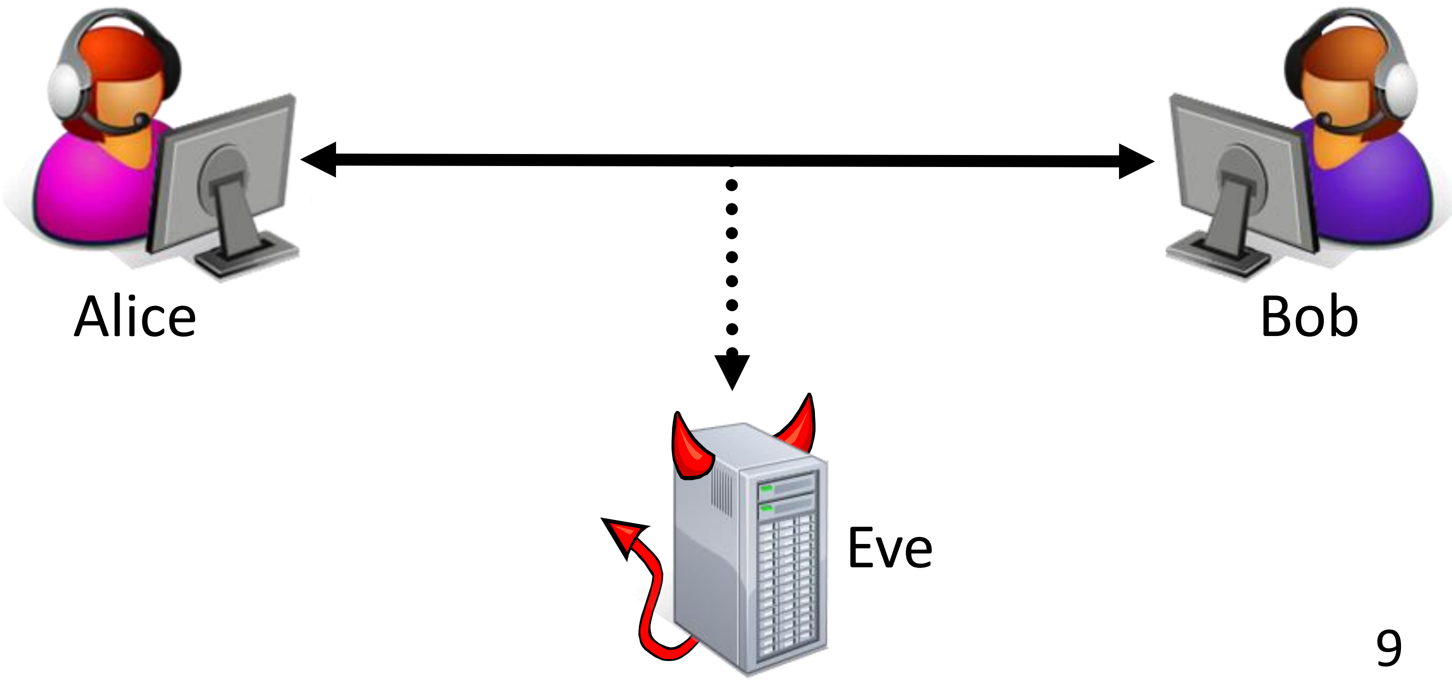
Symmetric-Key Encryption

Alice and Bob wish to communicate secretly

- Eve observes the communication

Assumption: Alice and Bob share a secret key

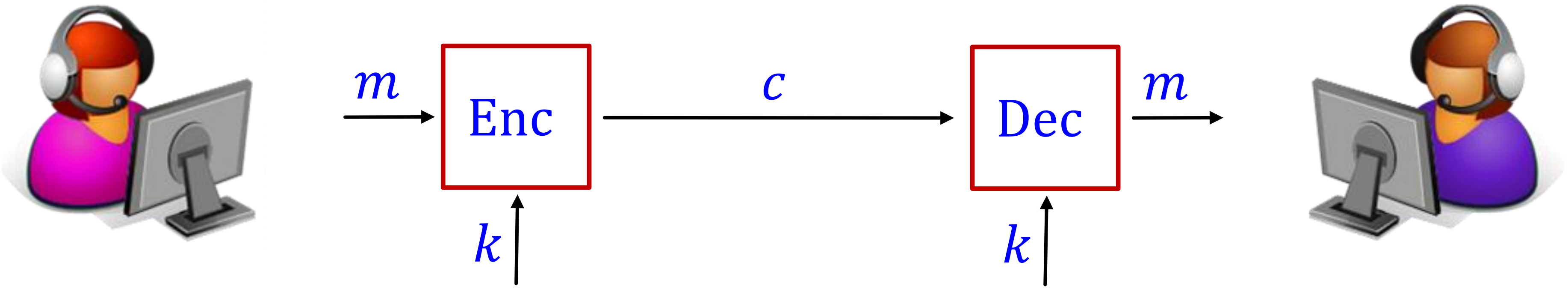
- The key is not known to Eve
- Same key used for both encryption and decryption



Symmetric-Key Encryption

Syntax: Three algorithms $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$

- Key-generation algorithm Gen outputs a key $k \in \mathcal{K}$
- Encryption algorithm Enc takes a key $k \in \mathcal{K}$ and a plaintext $m \in \mathcal{M}$, and outputs a ciphertext $c \in \mathcal{C}$
- Decryption algorithm Dec takes a key $k \in \mathcal{K}$ and ciphertext $c \in \mathcal{C}$, and outputs a plaintext $m \in \mathcal{M}$



$\mathcal{K}$ – key space

$\mathcal{M}$ – plaintext (=message) space

$\mathcal{C}$ – ciphertext space

Symmetric-Key Encryption

Syntax: Three algorithms $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$

- Key-generation algorithm Gen outputs a key $k \in \mathcal{K}$
- Encryption algorithm Enc takes a key $k \in \mathcal{K}$ and a plaintext $m \in \mathcal{M}$, and outputs a ciphertext $c \in \mathcal{C}$
- Decryption algorithm Dec takes a key $k \in \mathcal{K}$ and ciphertext $c \in \mathcal{C}$, and outputs a plaintext $m \in \mathcal{M}$

$$k \leftarrow \mathrm{Gen}() \qquad c \leftarrow \mathrm{Enc}_k(m) \qquad m = \mathrm{Dec}_k(c)$$

$\leftarrow$ randomized assignment
$=$ deterministic assignment

Correctness: $\forall k \in \mathcal{K}, m \in \mathcal{M}$

$$\mathrm{Dec}_k\big(\mathrm{Enc}_k(m)\big) = m$$

Kerckhoffs' principle

- Gen, Enc, and Dec are publicly known
- The only secret is the key k

JOURNAL

DES

SCIENCES MILITAIRES.

Janvier 1883.

LA CRYPTOGRAPHIE MILITAIRE.

« La cryptographie est un auxiliaire puissant de la tactique militaire. »
(Général Lewal, *Études de guerre.*)

I.

LA CRYPTOGRAPHIE DANS L'ARMÉE

A. Notions historiques.

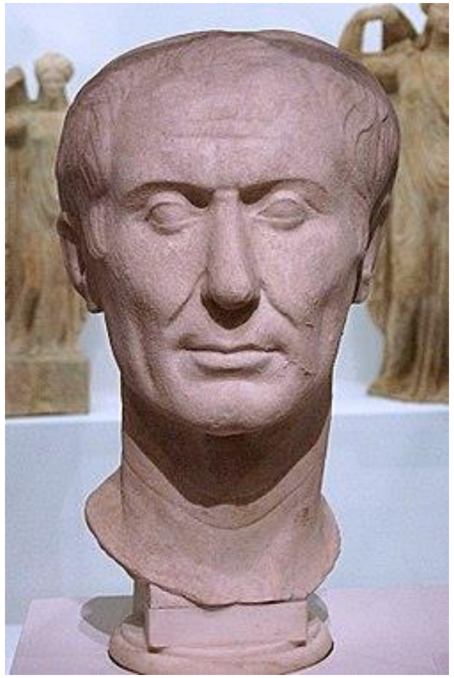
Wikipedia

Outline

- **Symmetric-key encryption**
- **Some historical ciphers**
- **The basic principles of modern cryptography**
- **Perfect secrecy and its limitations**

Shift Cipher (Caesar's Cipher)

- $\mathcal{M} = \{\mathrm{a}, \dots, \mathrm{z}\}^{\ell}$ and $\mathcal{C} = \{\mathrm{A}, \dots, \mathrm{Z}\}^{\ell}$
- Gen uniformly samples $k \leftarrow \{0, \dots, 25\}$
- Enc shifts each letter k positions forward (wrapping around from Z to A)
- Dec shifts backward



Wikipedia

Example with $k = 1$:

$$\mathrm{Enc}_k(\mathrm{welcometocryptocourse}) = \mathrm{XFMDPNFUPDSZQSPDPVSTF}$$

Is it "secure"?

- There are only 26 possible keys…
- $|\mathcal{K}|$ must not allow exhaustive search!

Substitution Cipher

- $\mathcal{M} = \{\mathrm{a}, \dots, \mathrm{z}\}^{\ell}$ and $\mathcal{C} = \{\mathrm{A}, \dots, \mathrm{Z}\}^{\ell}$
- Gen uniformly samples a permutation k over $\{\mathrm{a}, \dots, \mathrm{z}\}$
- Enc applies the permutation k to each letter
- Dec applied the inverse permutation k^{-1}

Example with $k =$

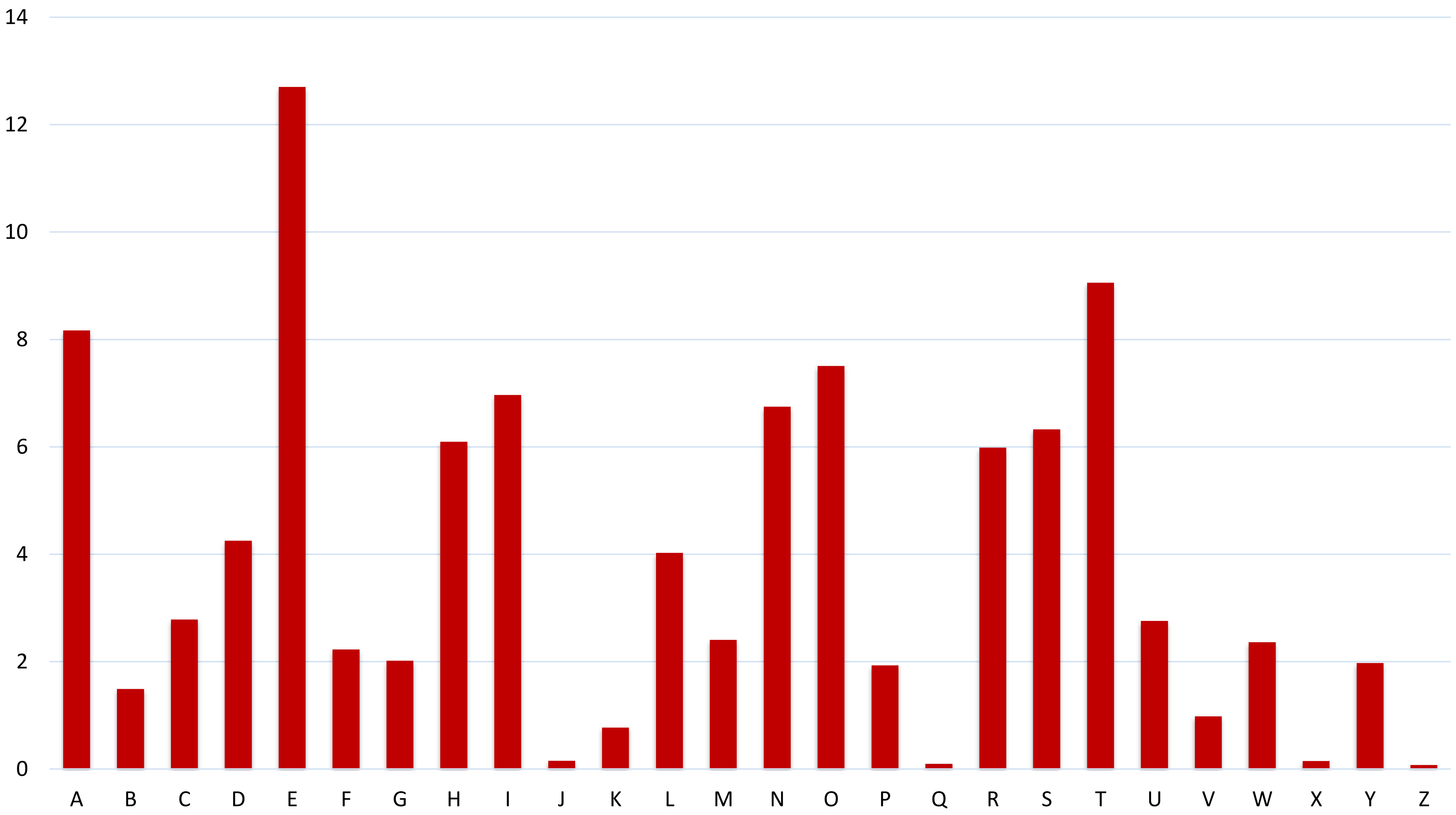
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
X	E	U	A	D	N	B	K	V	M	R	O	C	Q	F	S	Y	H	W	G	L	Z	I	J	P	T

$$\mathrm{Enc}_k(\mathrm{tellhimaboutme}) = \mathrm{GDOOKVCXEFLGCD}$$

Is it "secure"?

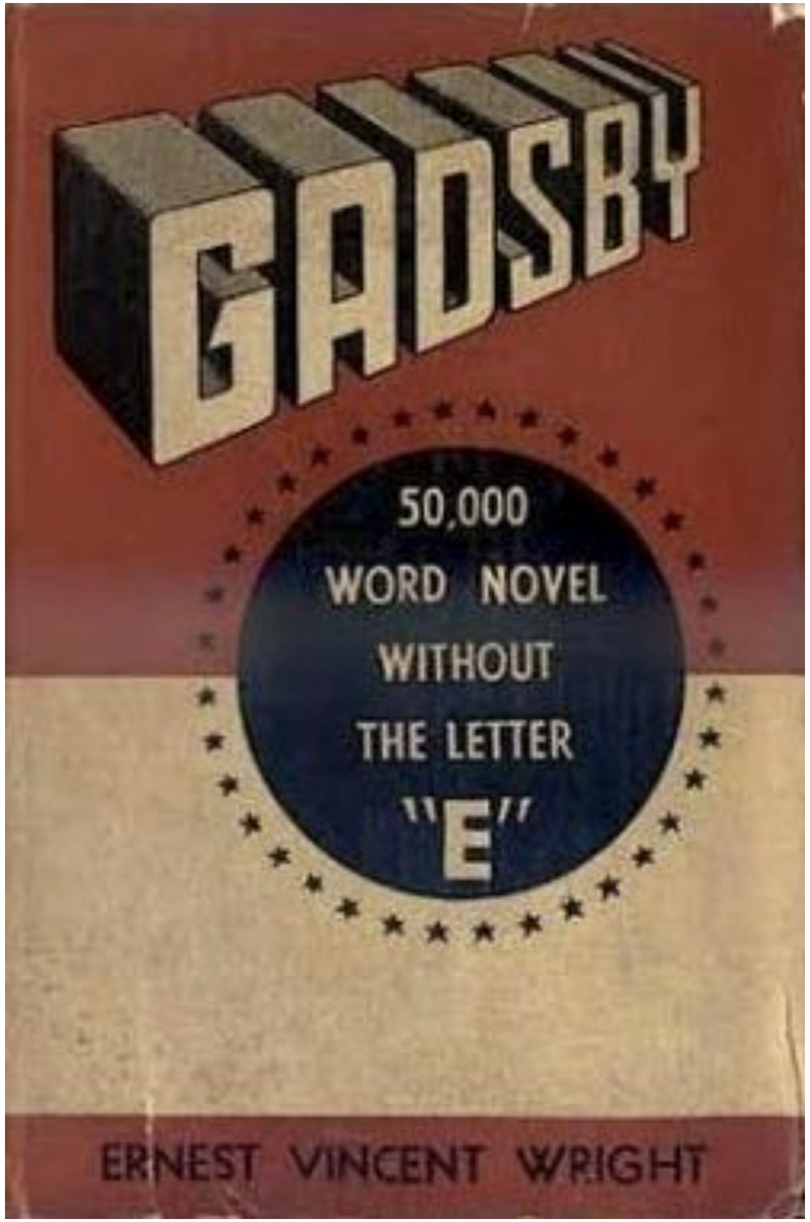
- There are many keys $(26! \approx 2^{88})$
- But can use statistical patterns of the English language…

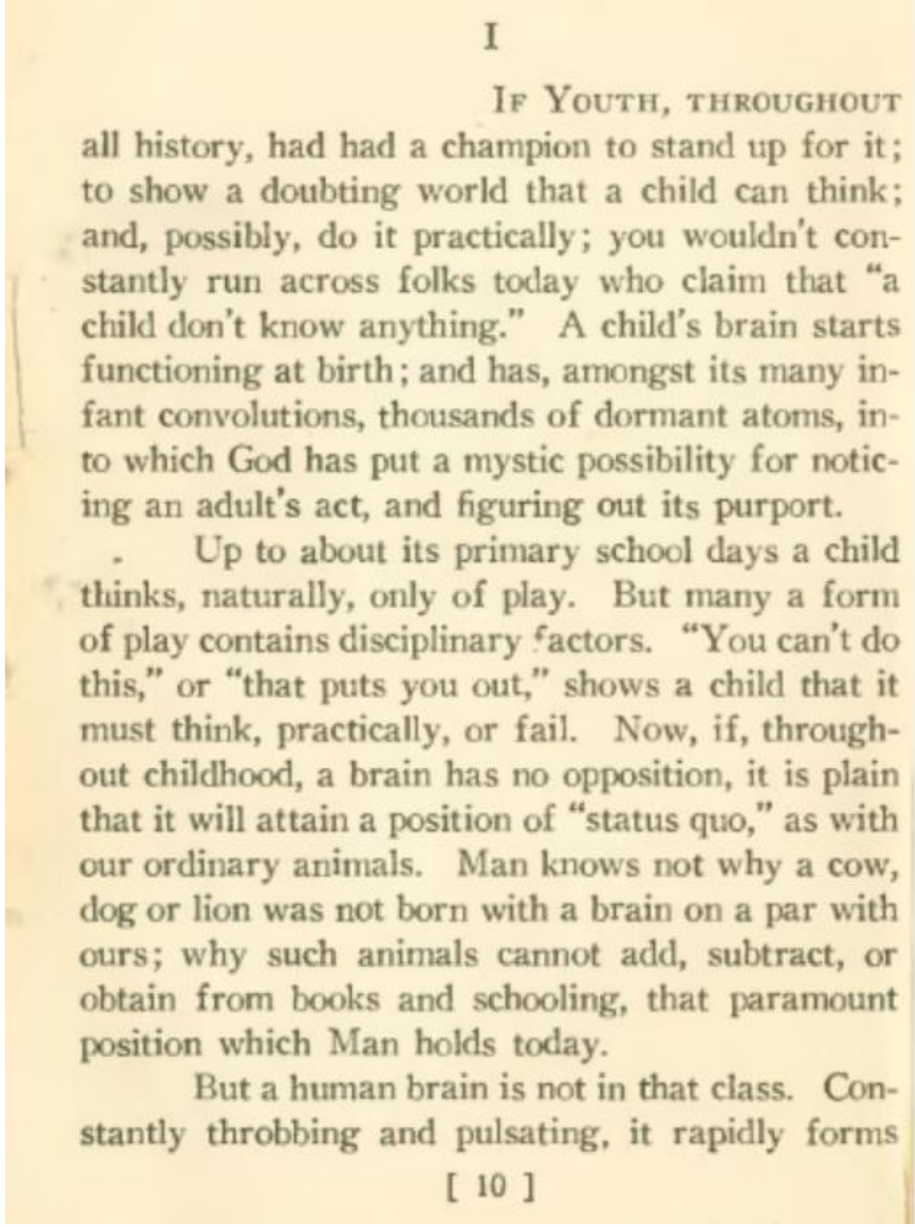
English Letter Frequencies



http://en.wikipedia.org/wiki/Letter_frequency

English Letter Frequencies





I

IF YOUTH, THROUGHOUT all history, had had a champion to stand up for it; to show a doubting world that a child can think; and, possibly, do it practically; you wouldn't constantly run across folks today who claim that "a child don't know anything." A child's brain starts functioning at birth; and has, amongst its many infant convolutions, thousands of dormant atoms, into which God has put a mystic possibility for noticing an adult's act, and figuring out its purport.

Up to about its primary school days a child thinks, naturally, only of play. But many a form of play contains disciplinary factors. "You can't do this," or "that puts you out," shows a child that it must think, practically, or fail. Now, if, throughout childhood, a brain has no opposition, it is plain that it will attain a position of "status quo," as with our ordinary animals. Man knows not why a cow, dog or lion was not born with a brain on a par with ours; why such animals cannot add, subtract, or obtain from books and schooling, that paramount position which Man holds today.

But a human brain is not in that class. Constantly throbbing and pulsating, it rapidly forms

[10]

Vigenère Cipher

- Gen uniformly samples $k = k_0 \dots k_{t-1} \leftarrow \{0, \dots, 25\}^t$
- $\mathcal{M} = \{\mathrm{a}, \dots, \mathrm{z}\}^\ell$ and $\mathcal{C} = \{\mathrm{A}, \dots, \mathrm{Z}\}^\ell$
- Enc shifts the ith letter $k_{i \bmod t}$ positions forward
- Dec shifts backward


Wikipedia

Example with $k = 123$:

$$\mathrm{Enc}_k(\underset{1231231}{\mathrm{tellhim}}) = \mathrm{UGOMJLN}$$

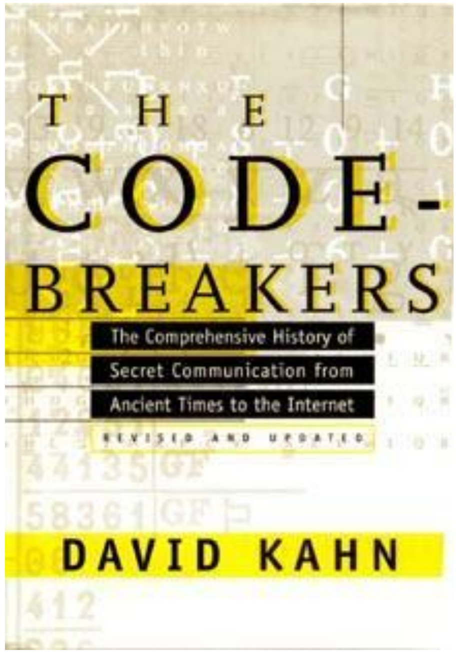
Is it "secure"?

- Trickier than breaking the shift and substitution ciphers
- But can still use statistical patterns

Historical Ciphers

Fascinating history

- Interesting & creative ideas (almost all broken by now)
- Influenced world history (e.g., cryptanalysis of the German Enigma in World War II)



It’s hard to design secure encryption schemes…

- What does “secure” mean?
- Can we avoid the “break → repair → break → repair → ⋯” cycle?
- Can we prove “security”?

Outline

- **Symmetric-key encryption**
- **Some historical ciphers**
- **The basic principles of modern cryptography**
- **Perfect secrecy and its limitations**

Modern Cryptography

Analyzing the security of a cryptographic system consists of

1. Formalizing a precise definition of security

If you don't understand what you want to achieve, how can you possibly know when (or if) you have achieved it?

Modern Cryptography

Analyzing the security of a cryptographic system consists of

1. Formalizing a precise definition of security
2. Stating the underlying assumptions

Others will attempt to validate (or invalidate) your assumptions

Modern Cryptography

Analyzing the security of a cryptographic system consists of

1. Formalizing a precise definition of security
2. Stating the underlying assumptions
3. Proving that the definition is satisfied given the assumptions

Can schemes still get "broken"?

YES!

- *If the definition does not capture real-world attacks*
- *If the assumptions turn out invalid*

Modern Cryptography

Analyzing the security of a cryptographic system consists of

1. Formalizing a precise definition of security
2. Stating the underlying assumptions
3. Proving that the definition is satisfied given the assumptions

Can schemes still get "broken"?

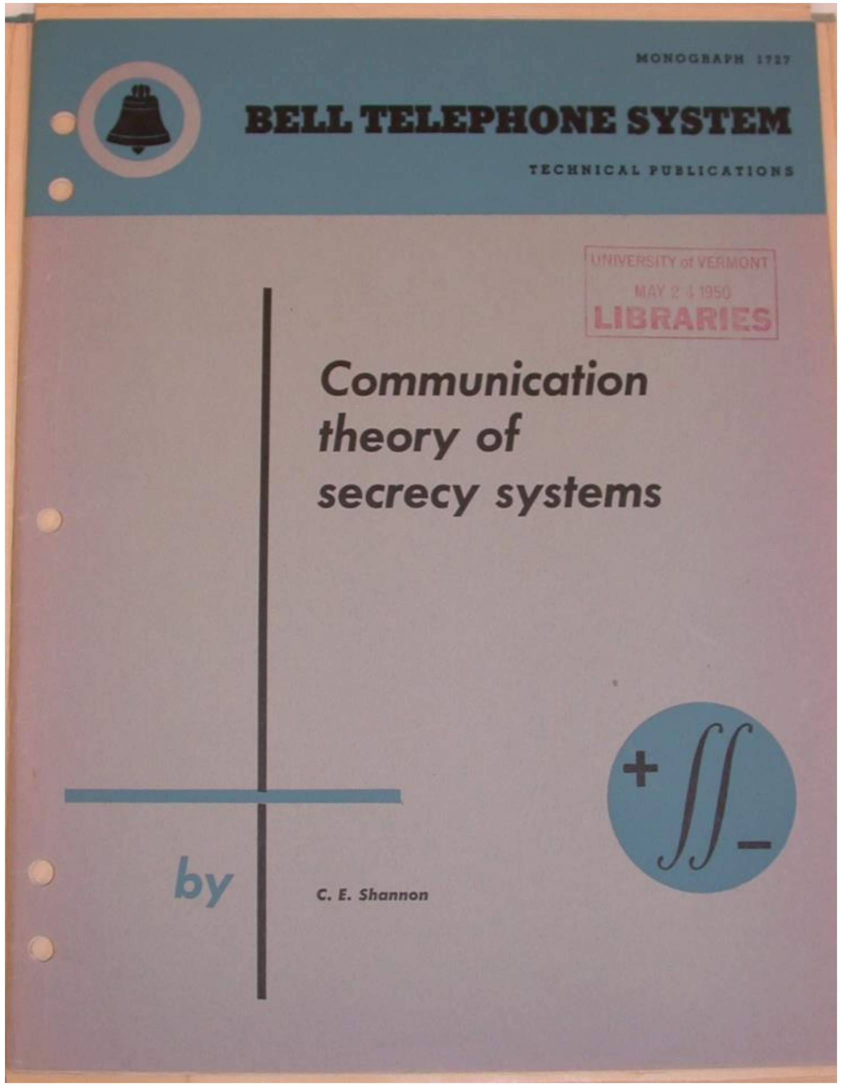
YES!

This does not detract from the benefits of having formal definitions and proofs!

Outline

- **Symmetric-key encryption**
- **Some historical ciphers**
- **The basic principles of modern cryptography**
- **Perfect secrecy and its limitations**

Perfect Secrecy (Shanon 1949)



Artwork by Bridgette Greenia

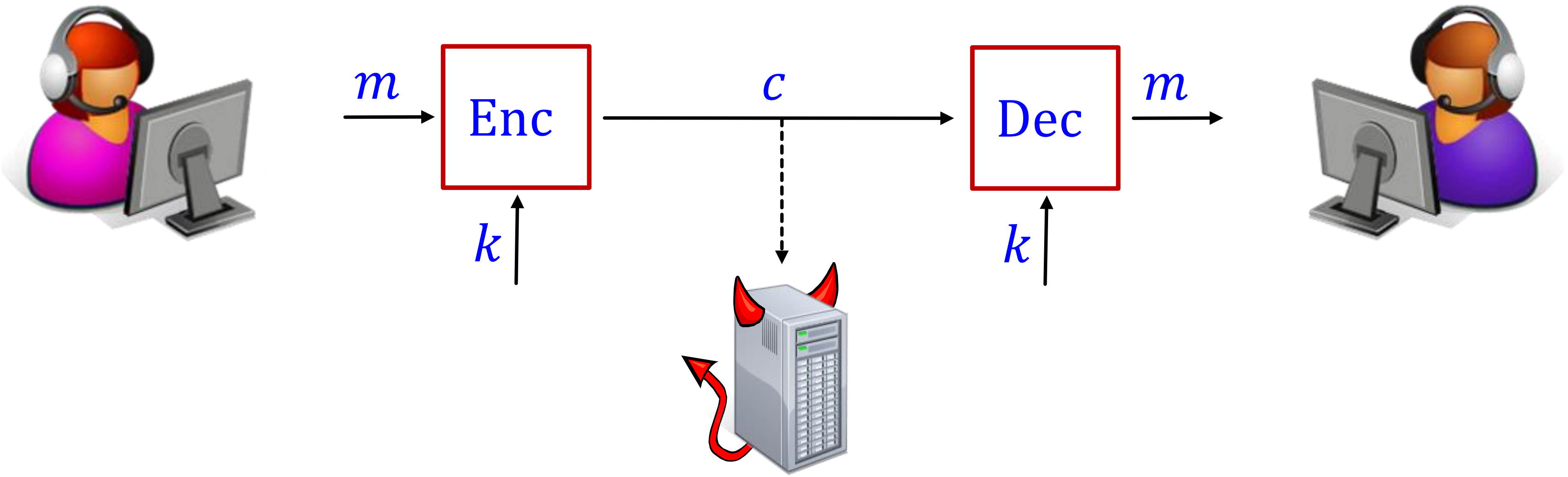
Perfect Secrecy

- Let $(\text{Gen}, \text{Enc}, \text{Dec})$ be a symmetric-key encryption scheme
- Alice and Bob share a key $k \leftarrow \text{Gen}()$

This defines a distribution K corresponding to the key

For example (shift cipher):

- Gen uniformly samples $k \leftarrow \{0, \ldots, 25\}$
- Then $\Pr[K = 6] = \Pr[K = 21] = \frac{1}{26}$

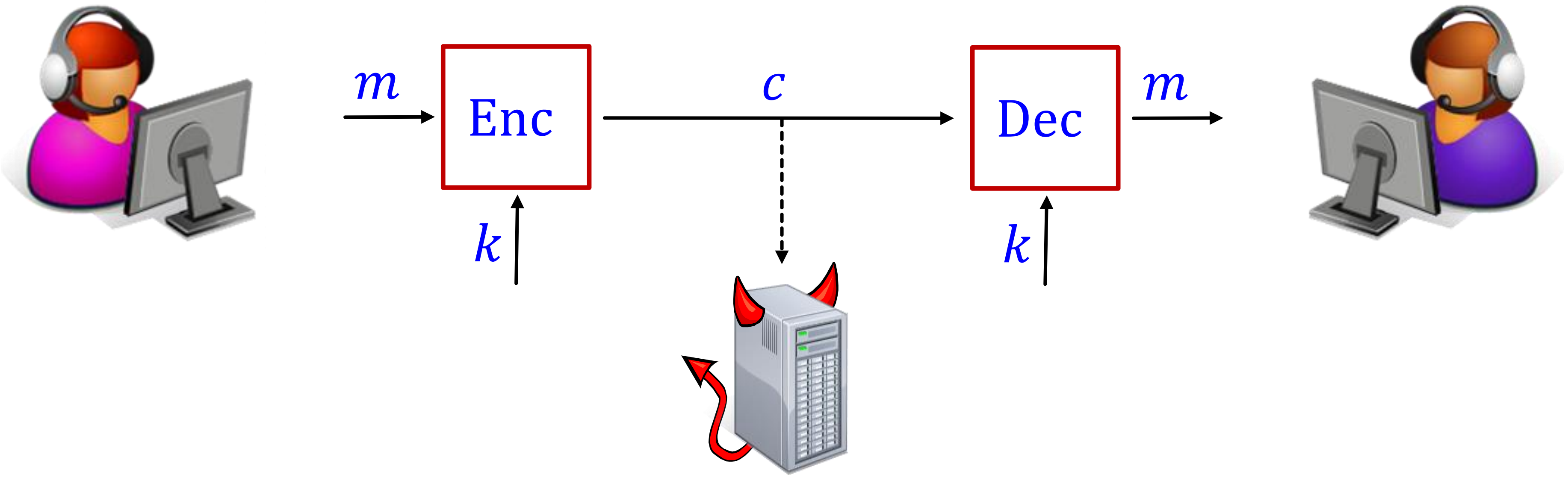


Perfect Secrecy

- Let $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ be a symmetric-key encryption scheme
- Alice and Bob share a key $k \leftarrow \mathrm{Gen}()$
- Eve knows an a-priori distribution M

K and M define a distribution $C = \mathrm{Enc}_K(M)$ corresponding to the ciphertext

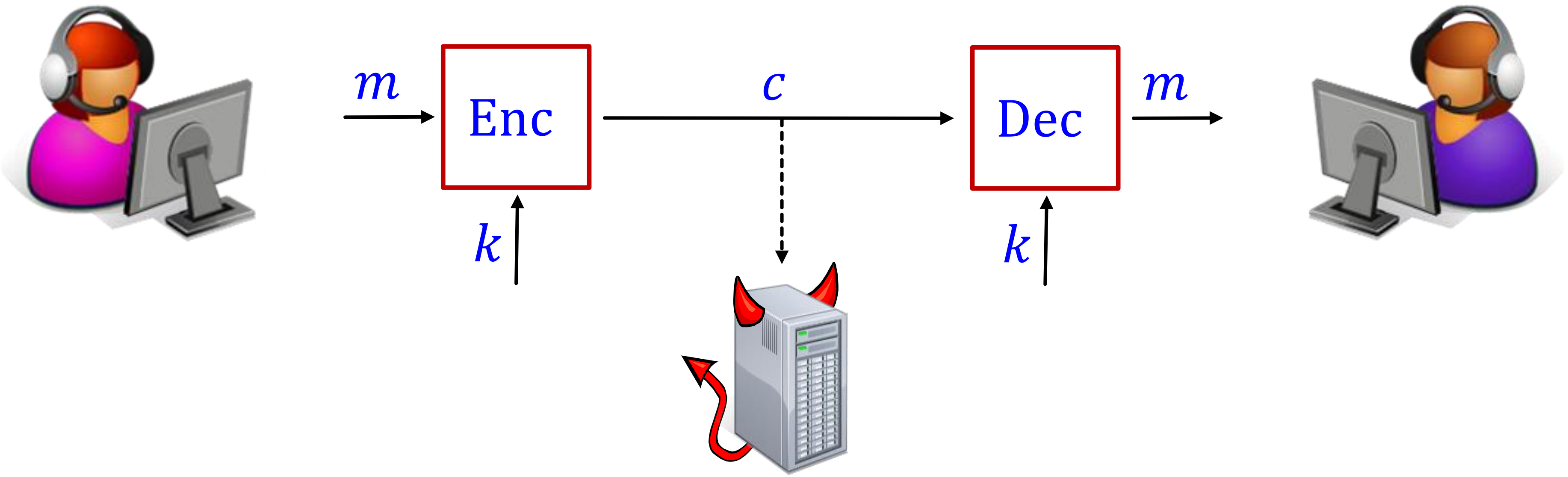
For example, Eve may know that
- $\Pr[M = "Attack\ now"] = 0.75$
- $\Pr[M = "Attack\ later"] = 0.25$



Perfect Secrecy

- Let $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ be a symmetric-key encryption scheme
- Alice and Bob share a key $k \leftarrow \mathrm{Gen}()$
- Eve knows an a-priori distribution M

Perfect secrecy (informal):
The ciphertext c should not reveal any additional information on m!!



Perfect Secrecy

Definition (Perfect secrecy):
A symmetric-key encryption scheme $(\text{Gen}, \text{Enc}, \text{Dec})$ is **perfectly secret** if for every distribution M over $\mathcal{M}$, for every $m \in \mathcal{M}$, and for every $c \in \mathcal{C}$ with $\Pr[C = c] > 0$ it holds that

$$\Pr[M = m \mid C = c] = \Pr[M = m]$$

Recall that

- Eve knows an a-priori distribution M
- K and M define a distribution $C = \text{Enc}_K(M)$

then perfect secrecy means that the distributions $\boldsymbol{M}$ and $\boldsymbol{C}$ are independent

Perfect Secrecy

Claim:
The shift and substitution ciphers **are not perfectly secret** for plaintexts of length $\ell > 1$

Shift cipher:

- $\mathcal{M} = \{\mathrm{a}, \dots, \mathrm{z}\}^{\ell}$ and $\mathcal{C} = \{\mathrm{A}, \dots, \mathrm{Z}\}^{\ell}$
- Gen uniformly samples $k \leftarrow \{0, \dots, 25\}$
- Enc shifts each letter k positions forward (wrapping around from Z to A)
- Dec shifts backward

Proof: To prove that a cipher is ***not*** perfectly secret we need to explicitly define a distribution M, a plaintext m and a ciphertext c (with $\Pr[C = c] > 0$) such that

$$\Pr[M = m \mid C = c] \neq \Pr[M = m]$$

Consider M defined by $\Pr[M = "aa"] = \Pr[M = "ab"] = 1/2$ and ciphertext $c = "AB"$.
On the one hand it holds that $\Pr[M = "aa" \mid C = "AB"] = 0$ (since a letter in the plaintext must be mapped to the same letter in the ciphertext).
On the other hand, by construction it holds that $\Pr[M = "aa"] = 1/2$.

$\Pr[C = "AB"] > 0$

Perfect Indistinguishability

For any pair of messages $m_0, m_1 \in \mathcal{M}$, Eve cannot tell if c is an encryption of m_0 or m_1

Alternative Definition (Perfect secrecy):
A symmetric-key encryption scheme $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ is **perfectly secret** if for every $m_0, m_1 \in \mathcal{M}$ and for every $c \in \mathcal{C}$ it holds that

$$\Pr[\mathrm{Enc}_k(\boldsymbol{m_0}) = c] = \Pr[\mathrm{Enc}_k(\boldsymbol{m_1}) = c]$$

Where $k \leftarrow Gen()$.

Theorem: A symmetric-key encryption scheme is perfectly secret according to the first definition if and only if it is perfectly secret according to the second definition

The One-Time Pad (Vernam 1917)

Exclusive OR (XOR)

- The XOR of 2 bits $a, b \in \{0,1\}$ is defined as follows

a	b	$a \oplus b$
0	0	0
0	1	1
1	0	1
1	1	0

- The XOR of 2 strings in $\{0,1\}^\ell$ is defined bit-wise
  - E.g., $101 \oplus 011 = 110$
- Note that
  - $a \oplus a = 0$
  - $a \oplus 0 = a$

Wikipedia

The One-Time Pad

- $\mathcal{K} = \mathcal{M} = \mathcal{C} = \{0,1\}^\ell$
- Gen uniformly samples $k \leftarrow \{0,1\}^\ell$
- $\mathrm{Enc}_k(m) = m \oplus k$
- $\mathrm{Dec}_k(c) = c \oplus k$

$\Pr[K = k] = 2^{-\ell}$ for every $k \in \{0,1\}^\ell$

Correctness: $\forall k \in \mathcal{K}, m \in \mathcal{M}$

$$\mathrm{Dec}_k\big(\mathrm{Enc}_k(m)\big) = \mathrm{Dec}_k(m \oplus k) = m \oplus k \oplus k = m$$

Theorem:
The one-time pad is **perfectly secret** for plaintexts of any length ℓ.

The One-Time Pad

Theorem:
The one-time pad is **perfectly secret** for plaintexts of any length ℓ.

In general, for any $m \in \mathcal{M}$ and $c \in \mathcal{C}$

$$\Pr_k[\,\mathrm{Enc}_k(m) = c] = \frac{\#k \in \mathcal{K} \; s.t. \; \mathrm{Enc}_k(m) = c}{|\mathcal{K}|}$$

Proof:

For any $m, c \in \{0,1\}^\ell$ it holds that

$$\Pr_k[\,\mathrm{Enc}_k(m) = c] = \Pr_k[m \oplus k = c] = \frac{1}{2^\ell}$$

Therefore, for every $m_0, m_1 \in \mathcal{M}$ and for every $c \in \mathcal{C}$ it holds that

$$\Pr_k[\mathrm{Enc}_k(\boldsymbol{m_0}) = c] = \Pr_k[\mathrm{Enc}_k(\boldsymbol{m_1}) = c] = 2^{-\ell}$$

The One-Time Pad: Limitations

- Keys are as long as plaintexts
- "Two-time" insecurity:
  Given $c = \mathrm{Enc}_k(m)$ and $c' = \mathrm{Enc}_k(m')$ can learn $c \oplus c' = m \oplus m'$
- Insecurity against "known-plaintext attacks":
  From m and $c = \mathrm{Enc}_k(m)$ can recover $k = m \oplus c$

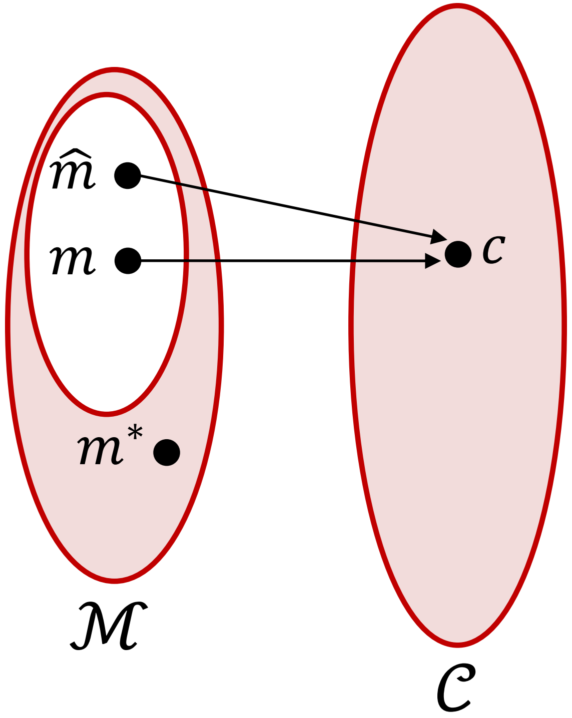
Theorem:
Let $\Pi = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ be a symmetric-key encryption scheme with key space $\mathcal{K}$ and message space $\mathcal{M}$. If Π is perfectly secret then $|\mathcal{K}| \geq |\mathcal{M}|$.

The One-Time Pad: Limitations

Proof:

Assume that $|\mathcal{K}| < |\mathcal{M}|$, and we show that the scheme is not perfectly secret.
Let M be the uniform distribution over $\mathcal{M}$, and fix some $m \in \mathcal{M}$.
Fix some $c \in \mathcal{C}$ which is a possible encryption of m.

Let $\mathcal{M}(c) \stackrel{\text{def}}{=} \{\hat{m} \mid \hat{m} = \text{Dec}_{\hat{k}}(c) \text{ for some } \hat{k} \in \mathcal{K}\}$, then $|\mathcal{M}(c)| \leq |\mathcal{K}|$.



Thus, the assumption $|\mathcal{K}| < |\mathcal{M}|$ implies that $|\mathcal{M}(c)| < |\mathcal{M}|$.
In particular, there exists some $m^* \in \mathcal{M}$ s.t. $m^* \notin \mathcal{M}(c)$.

This implies that

$$\Pr[M = m^* | C = c] = 0 \neq \frac{1}{|\mathcal{M}|} = \Pr[M = m^*]$$

and so the scheme is not perfectly secret.