CISC 3325 - Information Security

Toolbox: Cryptography

Topics for today

- Cryptography:
  - Problems encryption is designed to solve
  - Encryption tools categories, strengths, weaknesses
    - applications of each
  - Certificates and certificate authorities

Cryptography

https://www.tripwire.com/state-of-security/security-data-protection/cryptography/ordinary-people-need-cryptography/

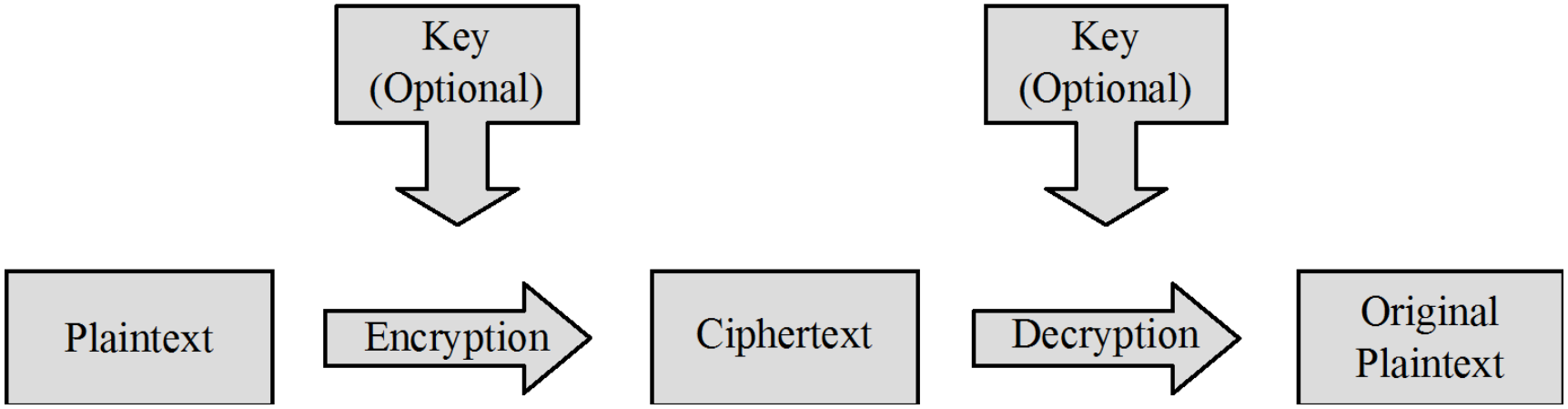
Problems Addressed by Encryption

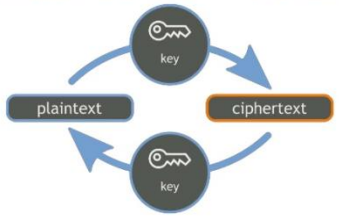
- Suppose a sender wants to send a message to a recipient. An attacker may attempt to:
  - Block the message
  - Intercept the message
  - Modify the message
  - Fabricate an authentic-looking alternate message

Encryption Terminology

- Sender
- Recipient
- Transmission medium
- Interceptor/intruder
- Encrypt, encode, or encipher
- Decrypt, decode, or decipher
- Cryptosystem
- Plaintext
- Ciphertext

Encryption/Decryption Process





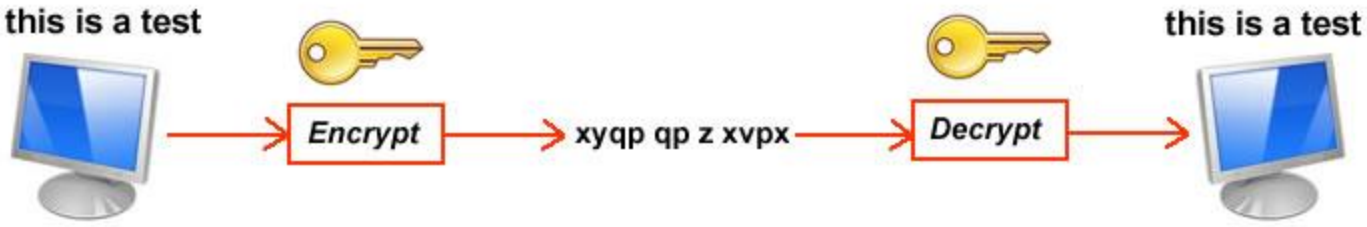
Symmetric Cryptosystem

- Scenario
  - Alice wants to send a message (plaintext P) to Bob.
  - The communication channel is insecure and can be eavesdropped
  - If Alice and Bob have previously agreed on a symmetric encryption scheme and a secret key K, the message can be sent encrypted (ciphertext C)
- Issues
  - What is a good symmetric encryption scheme?
  - What is the complexity of encrypting/decrypting?
  - What is the size of the ciphertext, relative to the plaintext?

https://www.slideshare.net/TusharSwami1/cryptography-72969031

Basics

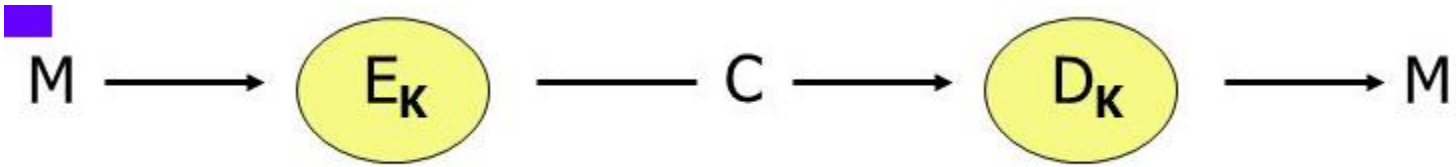
- Notation
  - Secret key K
  - Encryption function EK(P)
  - Decryption function DK(C)
  - Plaintext length typically same as ciphertext length
  - Encryption and decryption are permutation functions (bijections) on the set of all n-bit arrays



http://www.atksolutions.com/articles/basics_on_how_encryption_algorithm_works.html

Basics

- Efficiency
  - functions EK and DK should have efficient algorithms
- Consistency
  - Decrypting the ciphertext yields the plaintext
  - DK(EK(P)) = P



http://slideplayer.com/slide/10975387/

Attacks

http://www.hackerbulletin.com/types-cryptographic-attacks/

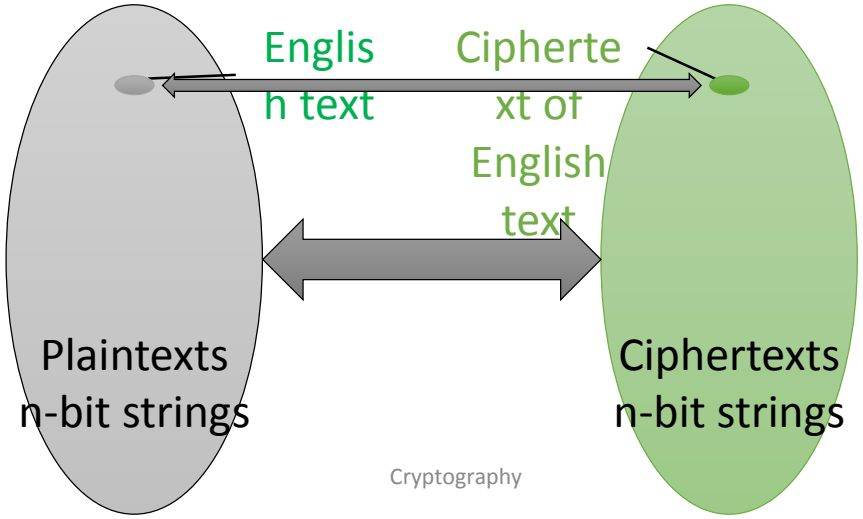
Brute-Force Attack

- Try all possible keys K and determine if DK(C) is a likely plaintext
  - Requires some knowledge of the structure of the plaintext (e.g., PDF file or email message)
- Key should be a sufficiently long random value to make exhaustive search attacks unfeasible

Image by Michael Cote from http://commons.wikimedia.org/wiki/File:Bingo_cards.jpg

Encrypting English Text

- English text typically represented with 8-bit ASCII encoding
- A message with t characters corresponds to an n-bit array, with n = 8t
- Redundancy due to repeated words and patterns
  - E.g., “th”, “ing”
- English plaintexts are a very small subset of all n-bit arrays



Entropy of Natural Language

- How much information can an alphabet with 8 characters carry?
  - 3 bits of information
  - $2^3 = 8$
- For the English language, each character can convey log (26) = 4.7 bits of information
- However, meaningful English text is only ~1.25 bits per char
- Therefore, plaintext redundancy = 4.7 – 1.25 = 3.45
- For example, if a word has 8 characters, what is the effective dictionary size?
  - How many words on average will we have?
  - $2^{8*1.25} = 2^{10} = 1024$

Entropy of English Language

- How do you statistically calculate the entropy of the next letter when the previous N - 1 letters are known?
  - In a word
- As N increases, the entropy approaches H, or the entropy of English

	F_0	F_1	F_2	F_3
26 letter	4.70	4.14	3.56	3.3

http://people.seas.harvard.edu/~jones/cscie129/papers/stanford_info_paper/entropy_of_english_9.htm

Entropy of English Language

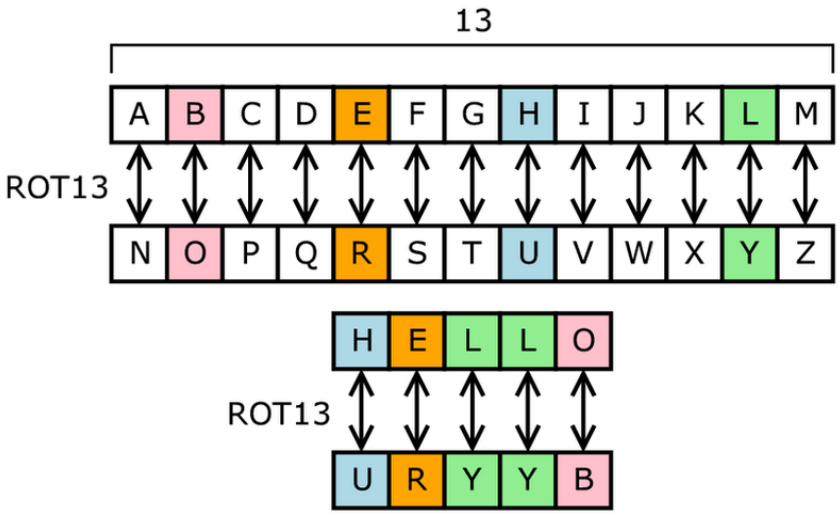
- Do we include the space?
  - The 27-letter sequences include the space as a letter
  - Onecanalmostalwaysfillinthespacesfromasequenceofwordswithnospaces
  - Therefore, spaces are basically redundant and will cause lower calculated entropies when taken into account
    - Only in the case where no statistics are taken into account, F0, is the entropy higher when the space is added
      - This simply adds another possible symbol, which means more uncertainty

	F_0	F_1	F_2	F_3
26 letter	4.70	4.14	3.56	3.3
27 letter	4.76	4.03	3.32	3.1

Substitution Ciphers

- Each letter is uniquely replaced by another.
- There are 26! possible substitution ciphers.
- There are more than 4.03 x 1026 such ciphers.

- One popular substitution "cipher" for some Internet posts is ROT13.



Public domain image from http://en.wikipedia.org/wiki/File:ROT13.png

Frequency Analysis

- Letters in a natural language, like English, are not uniformly distributed.
- Knowledge of letter frequencies, including pairs and triples can be used in cryptologic attacks against substitution ciphers.

a:	8.05%	b:	1.67%	c:	2.23%	d:	5.10%
e:	12.22%	f:	2.14%	g:	2.30%	h:	6.62%
i:	6.28%	j:	0.19%	k:	0.95%	l:	4.08%
m:	2.33%	n:	6.95%	o:	7.63%	p:	1.66%
q:	0.06%	r:	5.29%	s:	6.02%	t:	9.67%
u:	2.92%	v:	0.82%	w:	2.60%	x:	0.11%
y:	2.04%	z:	0.06%				

Letter frequencies in the book *The Adventures of Tom Sawyer*, by Twain.

Substitution Boxes

- Substitution can also be done on binary numbers.
- Such substitutions are usually described by substitution boxes, or S-boxes.

	00	01	10	11
00	0011	0100	1111	0001
01	1010	0110	0101	1011
10	1110	1101	0100	0010
11	0111	0000	1001	1100

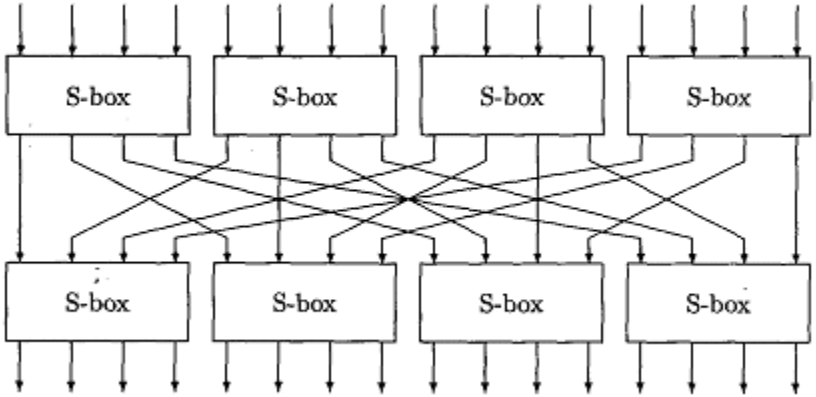
(a)

	0	1	2	3
0	3	8	15	1
1	10	6	5	11
2	14	13	4	2
3	7	0	9	12

(b)

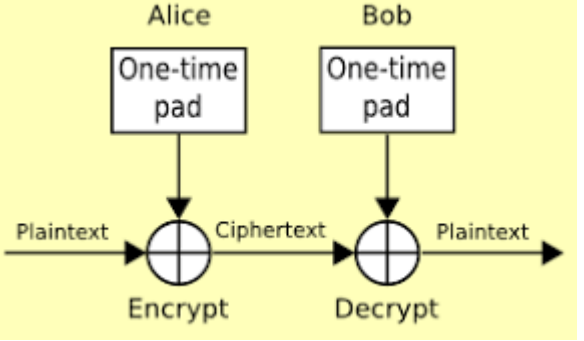
Figure 8.3: A 4-bit S-box (a) An S-box in binary. (b) The same S-box in decimal. This particular S-box is used in the Serpent cryptosystem, which

Substitution Boxes



https://www2.dmst.aueb.gr/dds/secimp/crypto/sp.htm

One-Time Pads

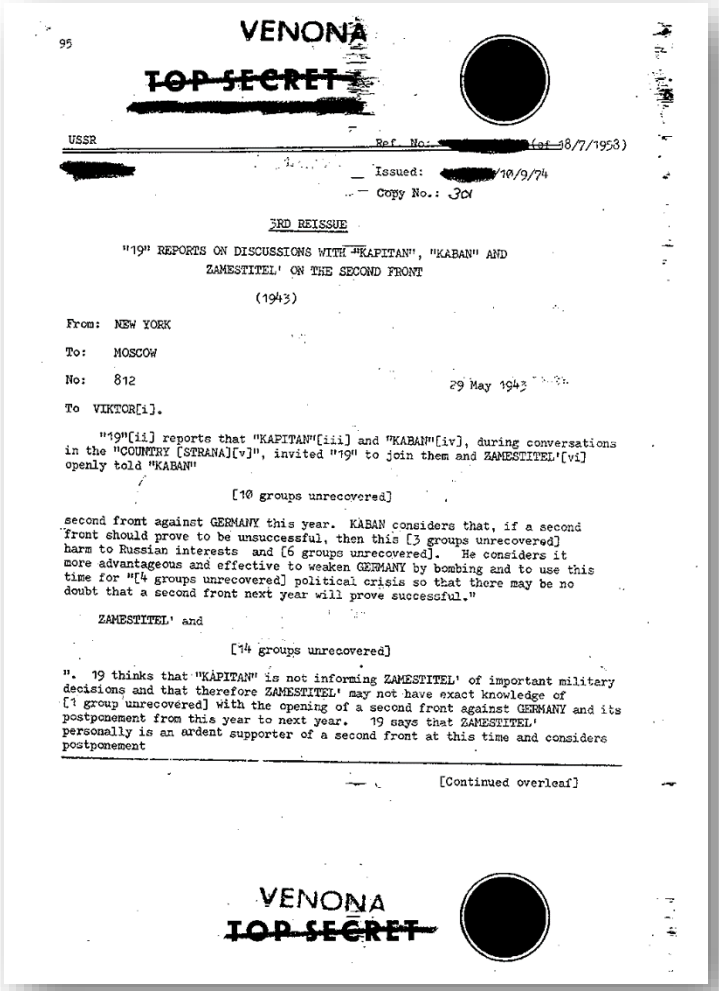


- There is one type of substitution cipher that is absolutely unbreakable.
  - The one-time pad was invented in 1917 by Joseph Mauborgne and Gilbert Vernam
  - We use a block of shift keys, (k1, k2, . . . , kn), to encrypt a plaintext, M, of length n, with each shift key being chosen uniformly at random.
- Since each shift is random, every ciphertext is equally likely for any plaintext.

https://programmingcode4life.blogspot.com/2015/10/one-time-pad-cipher.html

Weaknesses of the One-Time Pad

- In spite of their perfect security, one-time pads have some weaknesses
- The key has to be as long as the plaintext
- Keys can never be reused
  - Repeated use of one-time pads allowed the U.S. to break some of the communications of Soviet spies during the Cold War.



95

VENONA

~~TOP SECRET~~

USSR

Ref No: [redacted] (of 8/7/1953)

[redacted] Issued: [redacted] 10/9/74

Copy No.: 304

3RD REISSUE

"19" REPORTS ON DISCUSSIONS WITH "KAPITAN", "KABAN" AND ZAMESTITEL' ON THE SECOND FRONT

(1943)

From: NEW YORK

To: MOSCOW

No: 812

29 May 1943

To VIKTOR[i].

"19"[ii] reports that "KAPITAN"[iii] and "KABAN"[iv], during conversations in the "COUNTRY [STRANA][v]", invited "19" to join them and ZAMESTITEL'[vi] openly told "KABAN"

[10 groups unrecovered]

second front against GERMANY this year. KABAN considers that, if a second front should prove to be unsuccessful, then this [3 groups unrecovered] harm to Russian interests and [6 groups unrecovered]. He considers it more advantageous and effective to weaken GERMANY by bombing and to use this time for "[4 groups unrecovered] political crisis so that there may be no doubt that a second front next year will prove successful."

ZAMESTITEL' and

[14 groups unrecovered]

". 19 thinks that "KAPITAN" is not informing ZAMESTITEL' of important military decisions and that therefore ZAMESTITEL' may not have exact knowledge of [1 group unrecovered] with the opening of a second front against GERMANY and its postponement from this year to next year. 19 says that ZAMESTITEL' personally is an ardent supporter of a second front at this time and considers postponement

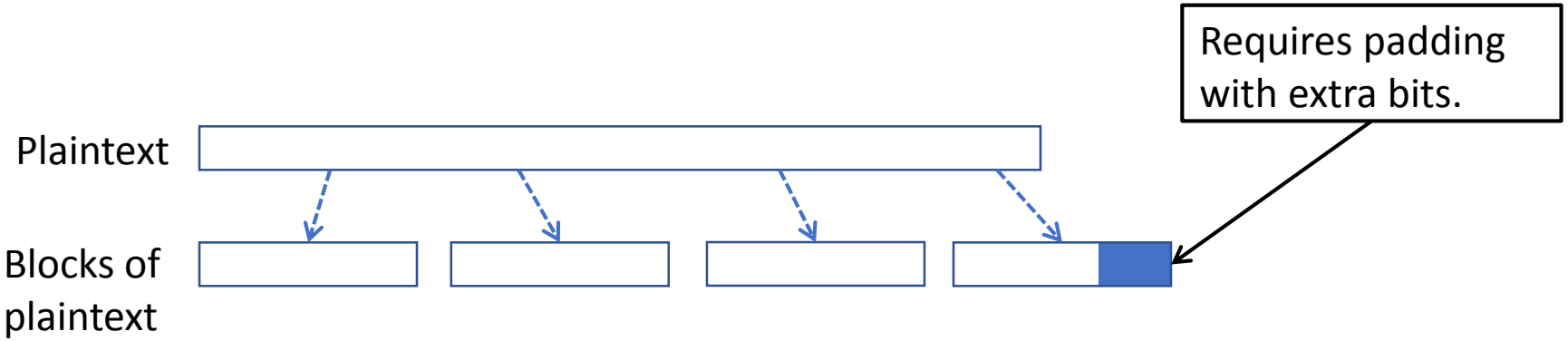
[Continued overleaf]

VENONA

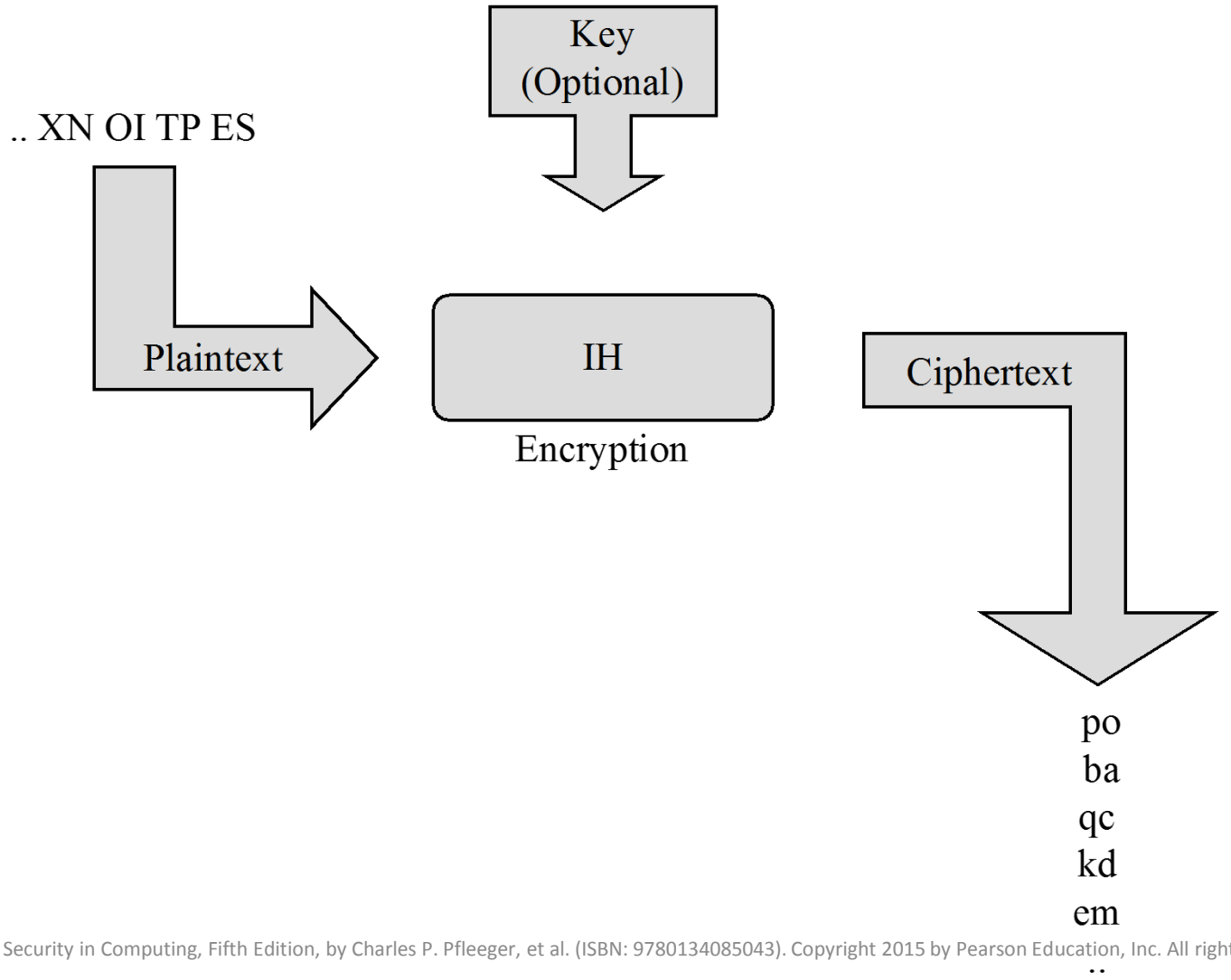
~~TOP SECRET~~

Block Ciphers

- Perhaps the most-used technique for encryption
  - Plaintext is partitioned into a sequence of blocks,
    - P = P[0], …, P[m−1]
  - Each lock is of fixed length b (e.g., 128 bits)
  - Each block is encrypted (more or less) separately, BlockCipher(key, plaintextBlock) → ciphertextBlock

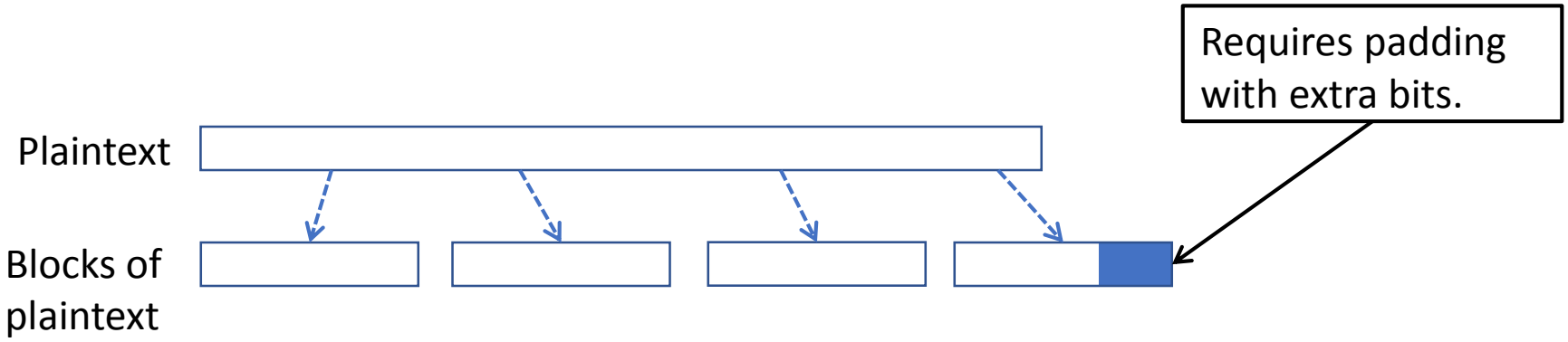


Block Ciphers



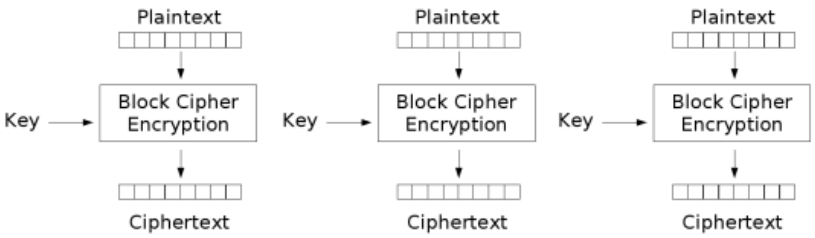
Padding

- Plaintext length must be a multiple of the block size
  - Otherwise we must pad last partial block to a full block
  - Must be able to tell when data ends, padding begins
- Example for block-size = 128 (16 bytes)
  - Plaintext: “Roberto” (7 bytes)
  - Padded plaintext: “Roberto999999999” (16 bytes)
    - Problem: cannot tell if plaintext was “Roberto” or “Roberto9”
  - Better: padded plaintext “Roberto099999999”

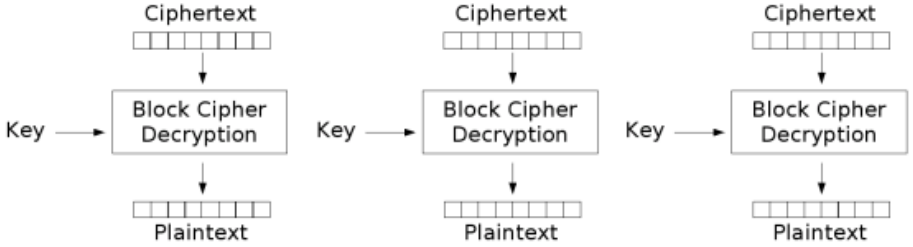


Block Cipher Modes

- A block cipher mode describes the way a block cipher encrypts and decrypts a sequence of message blocks.
- Electronic Code Book (ECB) Mode (is the simplest):
  - Block P[i] encrypted into ciphertext block C[i] = EK(P[i])
  - Block C[i] decrypted into plaintext block M[i] = DK(C[i])



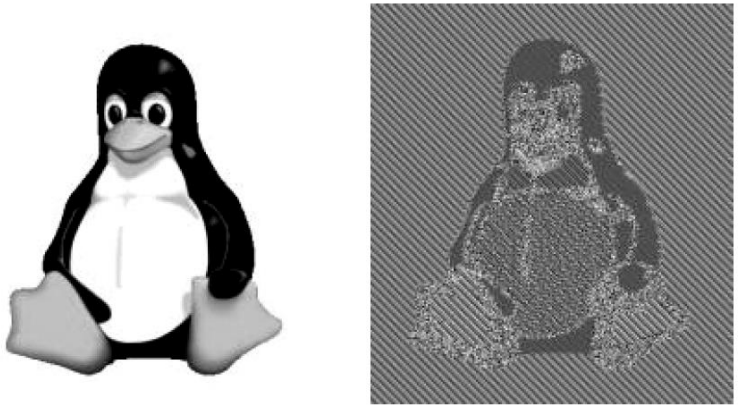
Electronic Codebook (ECB) mode encryption



Electronic Codebook (ECB) mode decryption

Strengths and Weaknesses of ECB

- Strengths:
  - Is very simple
  - Allows for parallel encryptions of the blocks of a plaintext
  - Can tolerate the loss or damage of a block
- Weakness:
  - Documents and images are not suitable for ECB encryption since patters in the plaintext are repeated in the ciphertext:

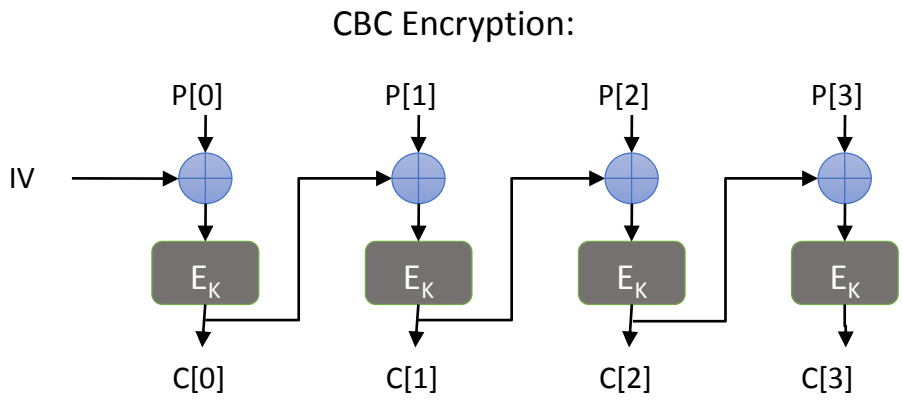


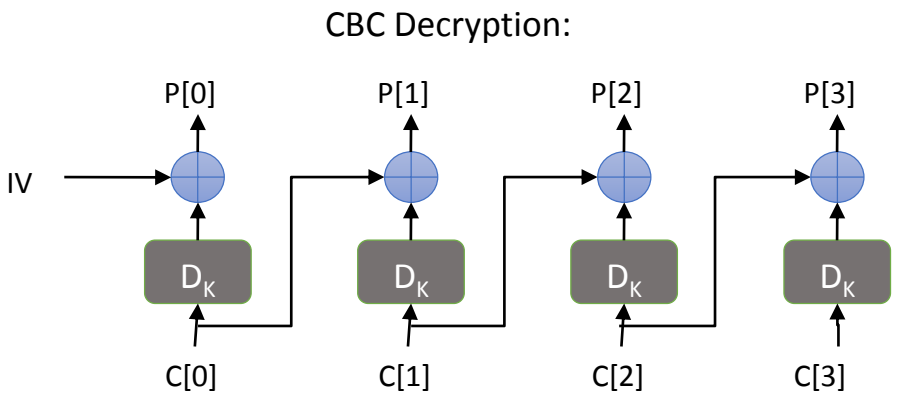
(a) (b)

Figure 8.6: How ECB mode can leave identifiable patterns in a sequence of blocks: (a) An image of Tux the penguin, the Linux mascot. (b) An encryption of the Tux image using ECB mode. (The image in (a) is by Larry Ewing, lewing@isc.tamu.edu, using The Gimp; the image in (b) is by Dr. Juzam. Both are used with permission via attribution.)

Cipher Block Chaining (CBC) Mode

- In Cipher Block Chaining (CBC) Mode
  - The previous ciphertext block is combined with the current plaintext block C[i] = EK (C[i –1] ⊕ P[i])
  - C[–1] = IV, a random block separately transmitted encrypted (known as the initialization vector)
  - Decryption: P[i] = C[i –1] ⊕ DK (C[i])





Strengths and Weaknesses of CBC

- Strengths:
  - Doesn’t show patterns in the plaintext
  - Is the most common mode
  - Is fast and relatively simple
- Weaknesses:
  - CBC requires the reliable transmission of all the blocks sequentially
  - CBC is not suitable for applications that allow packet losses (e.g., music and video streaming)

Block Ciphers in Practice

- Data Encryption Standard (DES)
  - Developed by IBM and adopted by NIST in 1977
  - 64-bit blocks and 56-bit keys
  - Small key space makes exhaustive search attack feasible since late 90s
- Triple DES (3DES)
  - Nested application of DES with three different keys KA, KB, and KC
  - Effective key length is 168 bits, making exhaustive search attacks unfeasible
  - C = EKC(DKB(EKA(P))); P = DKA(EKB(DKC(C)))
  - Equivalent to DES when KA=KB=KC (backward compatible)

Block Ciphers in Practice

- Advanced Encryption Standard (AES)
  - Selected by NIST in 2001 through open international competition and public discussion
  - 128-bit blocks and several possible key lengths: 128, 192 and 256 bits
  - Exhaustive search attack not currently possible
  - AES-256 is the symmetric encryption algorithm of choice

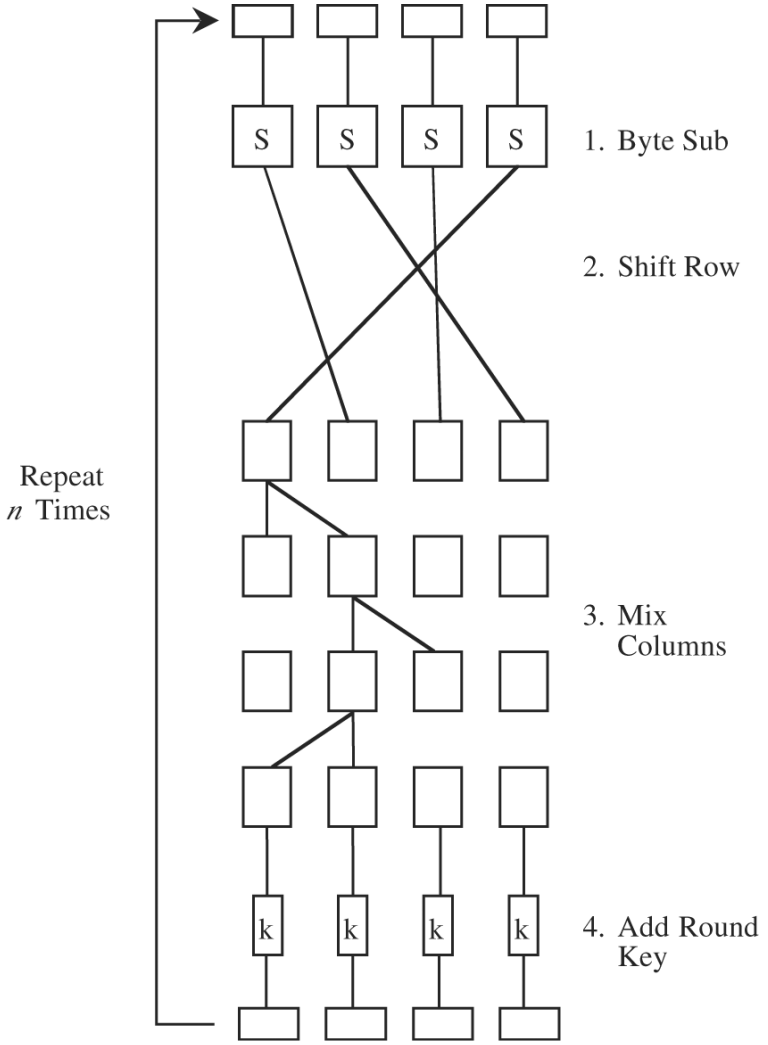
DES: The Data Encryption Standard

- Symmetric block cipher
- Developed in 1976 by IBM for the US National Institute of Standards and Technology (NIST)

Form	Operation	Properties	Strength
DES	Encrypt with one key	56-bit key	Inadequate for high-security applications by today's computing capabilities
Double DES	Encrypt with first key; then encrypt result with second key	Two 56-bit keys	Only doubles strength of 56-bit key version
Two-key triple DES	Encrypt with first key, then encrypt (or decrypt) result with second key, then encrypt result with first key (E-D-E)	Two 56-bit keys	Gives strength equivalent to about 80-bit key (about 16 million times as strong as 56-bit version)
Three-key triple DES	Encrypt with first key, then encrypt or decrypt result with second key, then encrypt result with third key (E-E-E)	Three 56-bit keys	Gives strength equivalent to about 112-bit key about 72 quintillion ($72*10^{15}$) times as strong as 56-bit version

AES: Advanced Encryption System

- Symmetric block cipher
- Developed in 1999 by independent Dutch cryptographers
- Still in common use



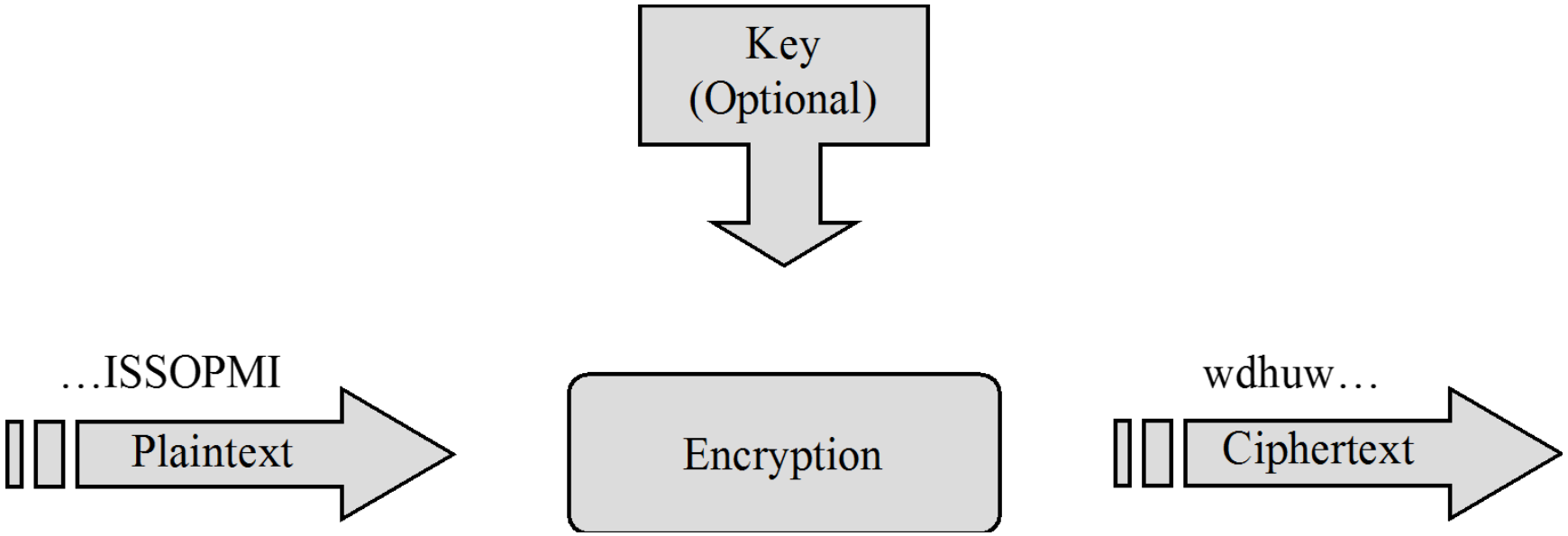
DES vs. AES

	DES	AES
Date designed	1976	1999
Block size	64 bits	128 bits
Key length	56 bits (effective length); up to 112 bits with multiple keys	128, 192, 256 (and possibly more) bits
Operations	16 rounds	10, 12, 14 (depending on key length); can be increased
Encryption primitives	Substitution, permutation	Substitution, shift, bit mixing
Cryptographic primitives	Confusion, diffusion	Confusion, diffusion
Design	Open	Open
Design rationale	Closed	Open
Selection process	Secret	Secret, but open public comments and criticisms invited
Source	IBM, enhanced by NSA	Independent Dutch cryptographers

Stream Ciphers

- An alternative approach to using block cipher
- Key stream
  - Pseudo-random sequence of bits S = S[0], S[1], S[2], …
  - Can be generated on-line one bit (or byte) at the time
  - Successive elements of the keystream generated based on an internal state
- Stream cipher
  - XOR the plaintext with the key stream $C[i] = S[i] \oplus P[i]$
  - Suitable for plaintext of arbitrary length generated on the fly, e.g., media stream

Stream Ciphers



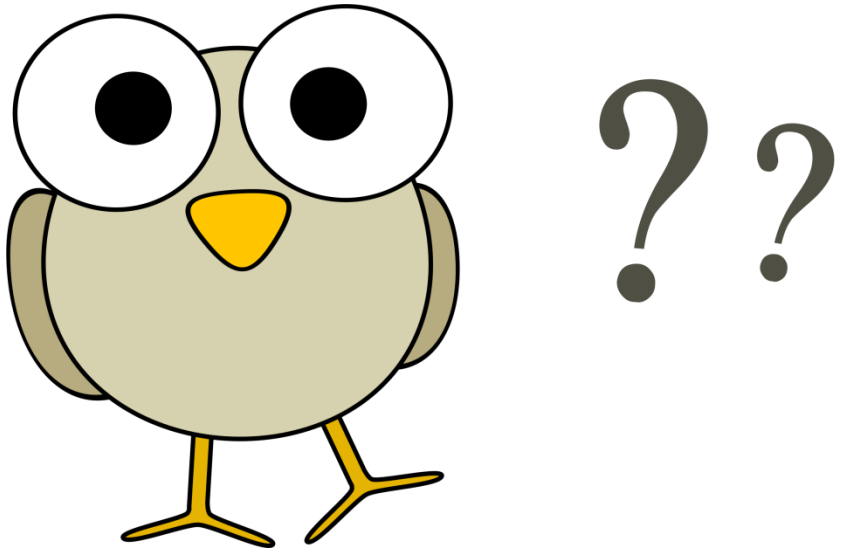
Key Stream Generation

- RC4
  - Designed in 1987 by Ron Rivest for RSA Security
  - Trade secret until 1994
  - Uses keys with up to 2,048 bits
  - Simple algorithm
- Block cipher in counter mode (CTR)
  - Use a block cipher with block size b
  - The secret key is a pair (K,t), where K a is key and t (counter) is a b-bit value
  - The key stream is the concatenation of ciphertexts
  - EK (t), EK (t + 1), EK (t + 2), …
  - Can use a shorter counter concatenated with a random value
  - Synchronous stream cipher

Stream vs. Block

	Stream	Block
Advantages	• Speed of transformation • Low error propagation	• High diffusion • Immunity to insertion of symbol
Disadvantages	• Low diffusion • Susceptibility to malicious insertions and modifications	• Slowness of encryption • Padding • Error propagation

- Questions?

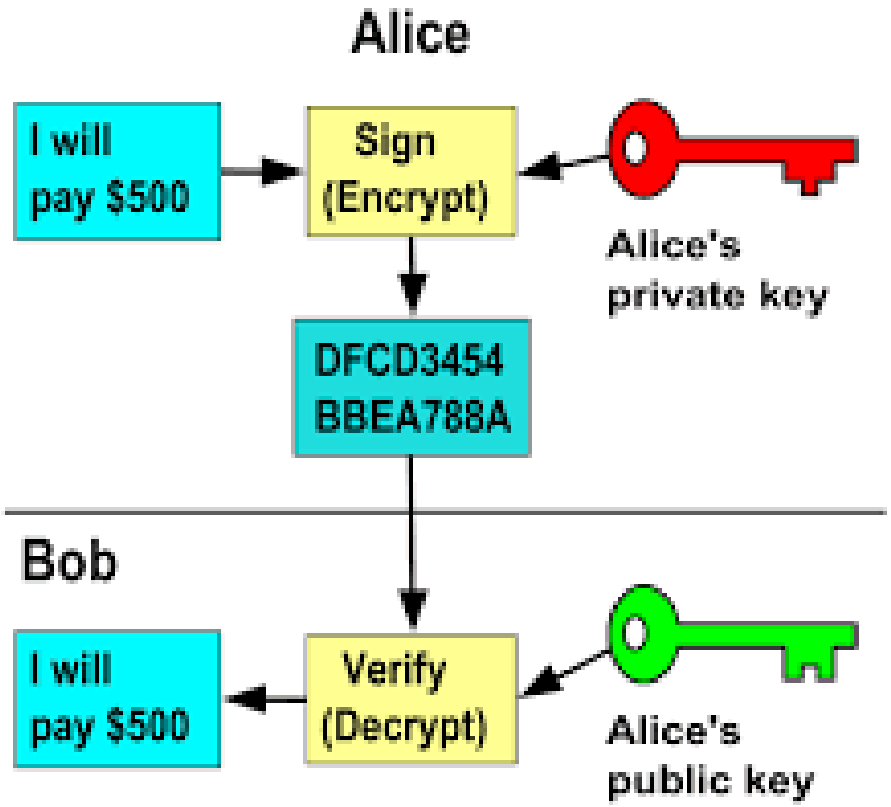


Public Key (Asymmetric) Encryption

Public Key (Asymmetric) Cryptography

- Instead of two users sharing one secret key, each user has two keys: one public and one private
- Messages encrypted using the user's public key can only be decrypted using the user's private key, and vice versa
  - P = D(Kpriv,E(Kpub,P)) or
  - P = D(Kpub,E(Kpriv,P))

Public Key Encryption



https://commons.wikimedia.org/wiki/File:Public_key_signing.svg

Public Key Encryption

- A cryptographic system that uses pairs of keys
  - public keys which may be disseminated widely
    - Any person can encrypt the message
  - private keys which are known only to the owner
    - Message can only be decrypted with this key
- Analogous to a self-closing door
  - Need key to open it
  - Closing it shuts it automatically

Public Key Encryption

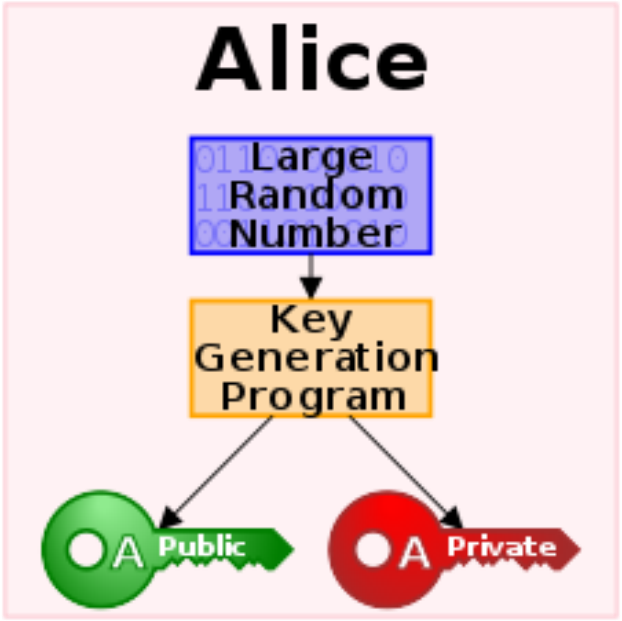
- Accomplishes two functions: authentication and encryption
- Authentication: the public key verifies that a holder of the paired private key sent message
  - Decryption will fail otherwise
- Encryption: only the paired private key holder can decrypt the encrypted message

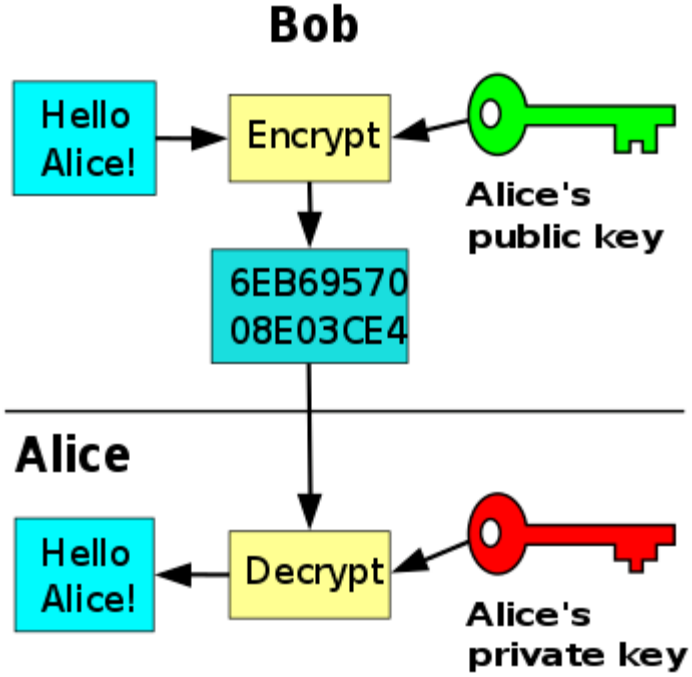
Public Key Encryption

- Why does it work?
- It is not feasible to compute the private key
  - From knowledge of its paired public key
- Therefore, only the private key is kept private
  - The public key can be openly distributed without compromising security
- Public key cryptography relies on math problems that have no efficient solution

Public Key Encryption

•





https://en.wikipedia.org/wiki/Public-key_cryptography

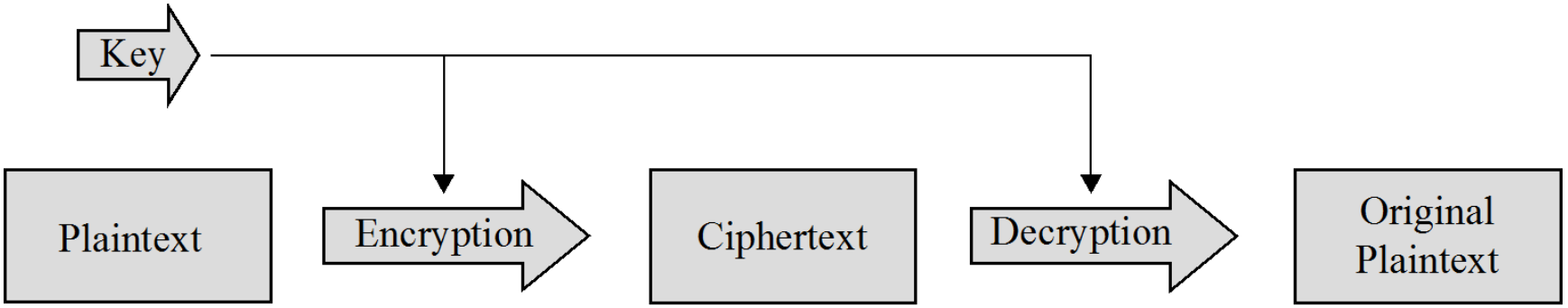
Public Key Encryption

- Often used to secure communication
  - Over the internet, open networks
  - Typically used for key exchange

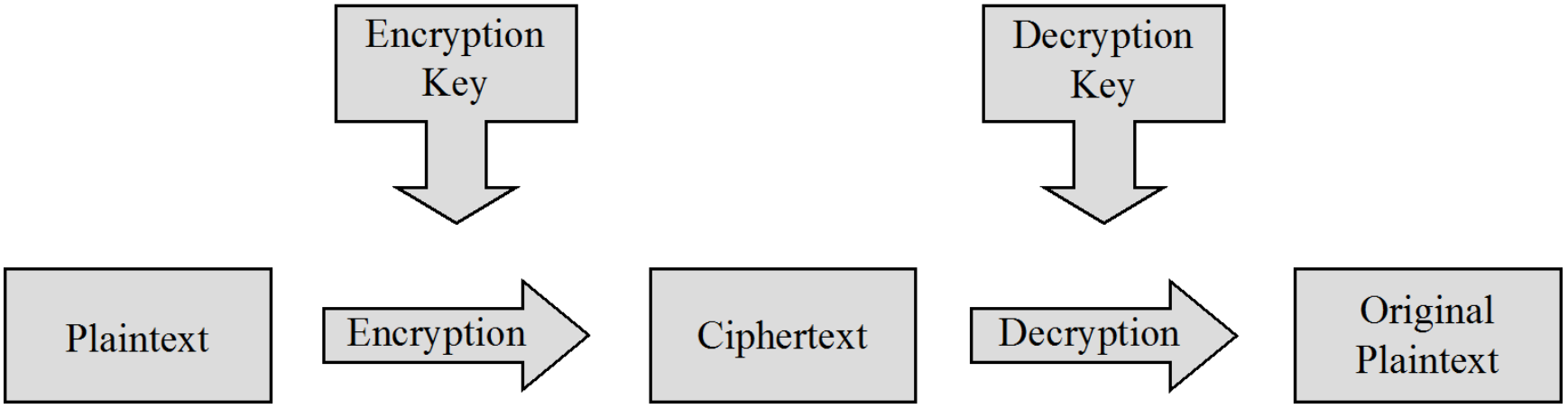
Secret Key vs. Public Key Encryption

	Secret Key (Symmetric)	Public Key (Asymmetric)
Number of keys	1	2
Key size (bits)	56–112 (DES), 128–256 (AES)	Unlimited; typically no less than 256; 1000 to 2000 currently considered desirable for most uses
Protection of key	Must be kept secret	One key must be kept secret; the other can be freely exposed
Best uses	Cryptographic workhorse. Secrecy and integrity of data, from single characters to blocks of data, messages and files	Key exchange, authentication, signing
Key distribution	Must be out-of-band	Public key can be used to distribute other keys
Speed	Fast	Slow, typically by a factor of up to 10,000 times slower than symmetric algorithms

Symmetric vs. Asymmetric



(a) Symmetric Cryptosystem



(b) Asymmetric Cryptosystem

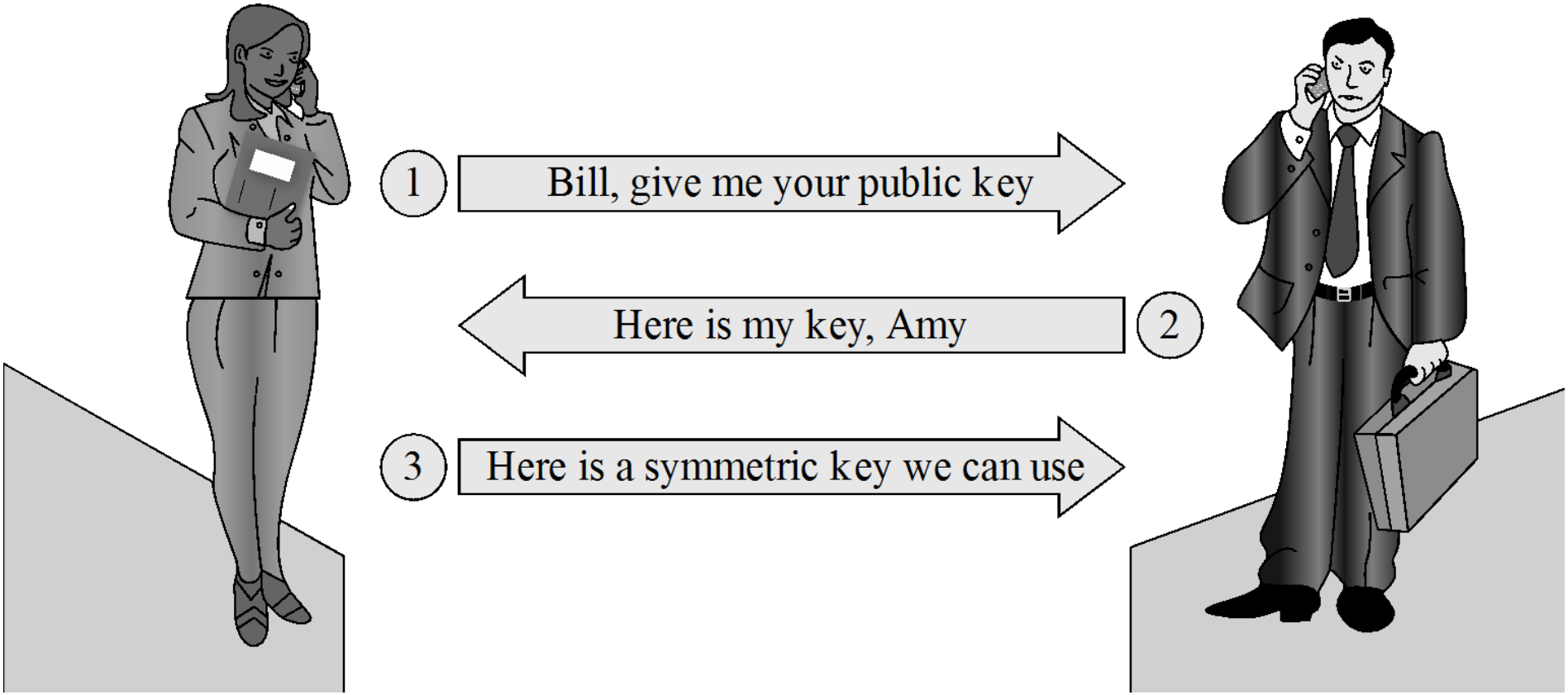
Rivest-Shamir-Adelman Public Key Encryption (RSA)

- The two keys can be used interchangeably:

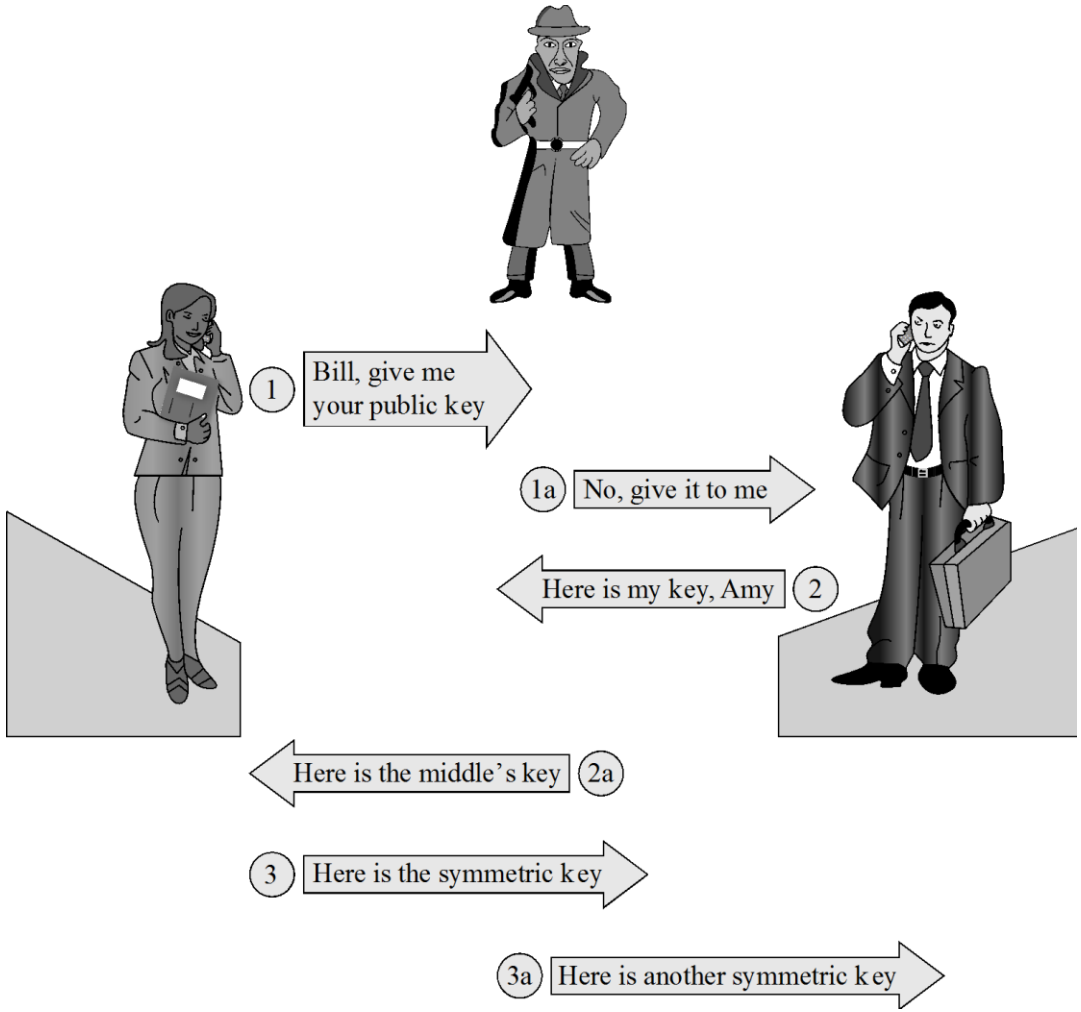
- C = RSA(P,e) where e can be either Kpriv or Kpub
- Then
- P = RSA(RSA(P,e),d) where d is the other key

- The keys can be 256 to 1000s of bits. The relationship between the keys relies on the fact that a large number is the product of two large prime numbers.

Public Key to Exchange Secret Keys



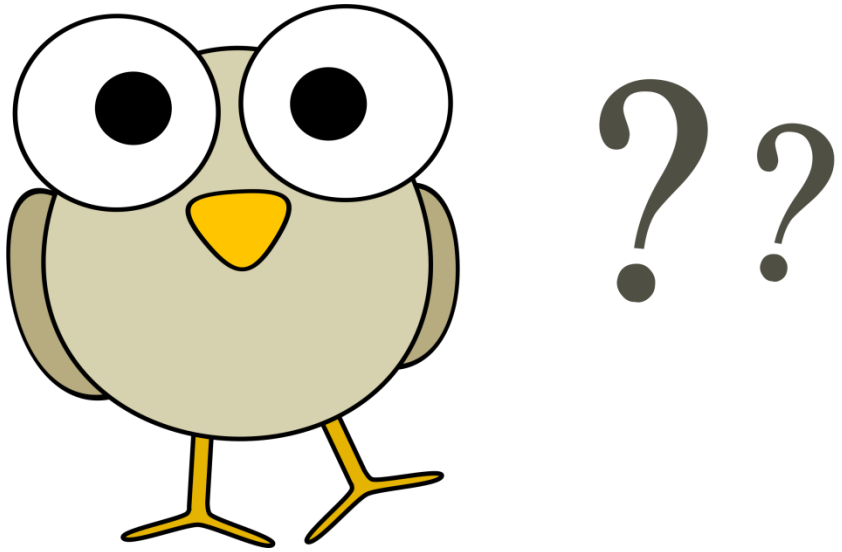
Key Exchange Man in the Middle



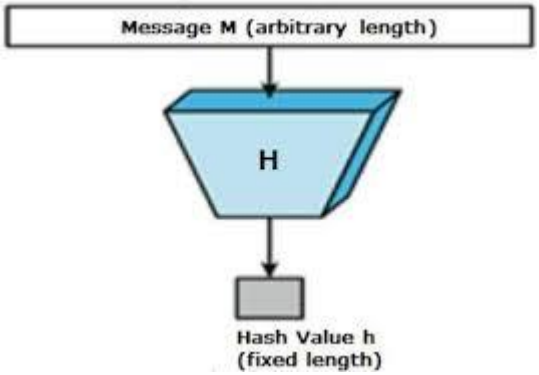
Revised Key Exchange Protocol

- Secure key exchange is complicated
- Many alternative secure protocols exist
  - Protocols need to be proven secure
  - New attacks introduced continuously

- Questions?



Cryptographic Hash Functions



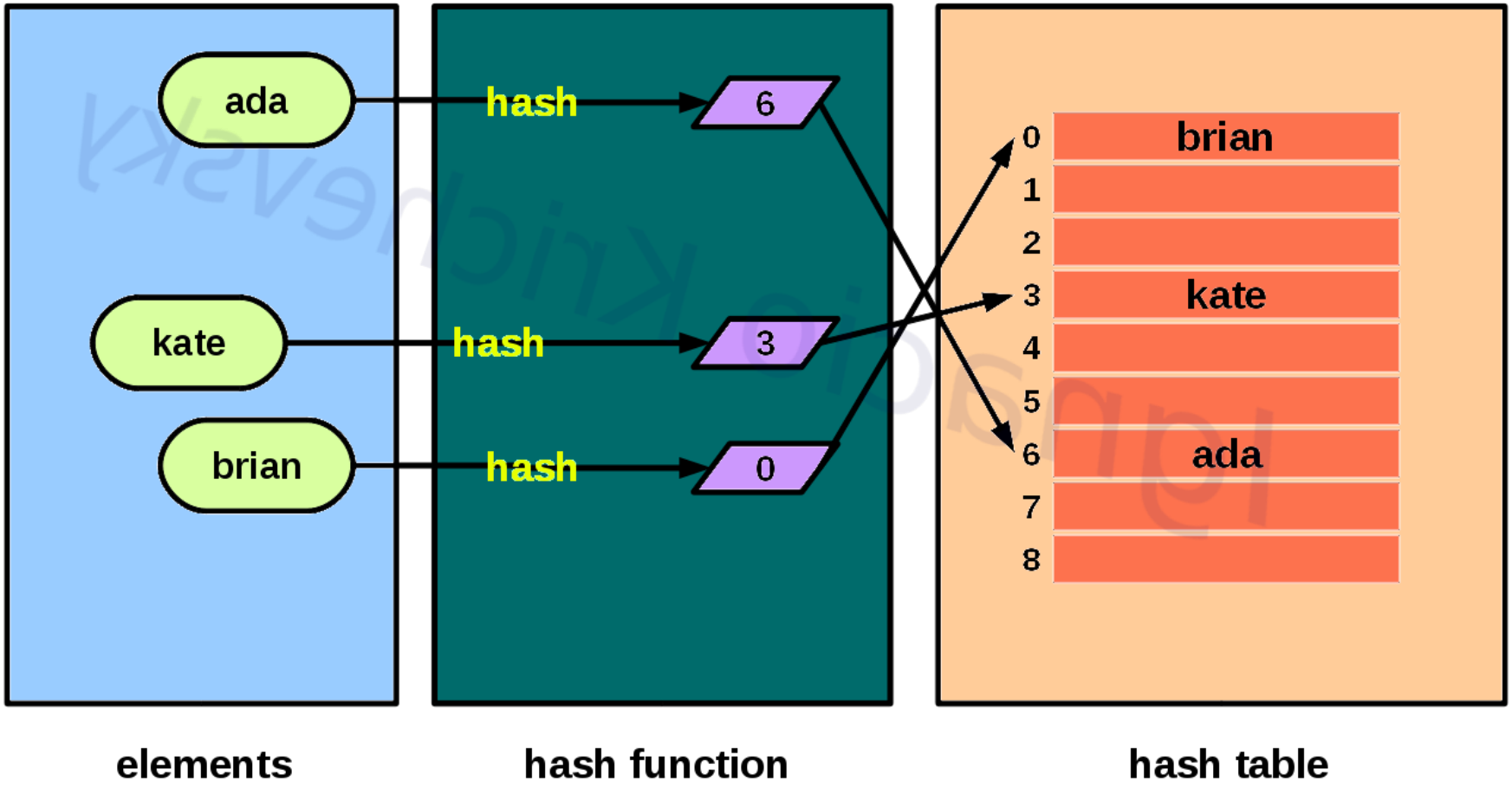
https://www.tutorialspoint.com/cryptography/cryptography_hash_functions.htm

Hash Functions

- A hash function h maps a plaintext x to a fixed-length value x = h(P) called hash value or digest of P
  - A collision is a pair of plaintexts P and Q that map to the same hash value, h(P) = h(Q)
  - Collisions are unavoidable
  - For efficiency, the computation of the hash function should take time proportional to the length of the input plaintext

Hash Functions

- Hash table
  - Search data structure based on storing items in locations associated with their hash value
  - Chaining or open addressing deal with collisions
  - Domain of hash values proportional to the expected number of items to be stored
  - The hash function should spread plaintexts uniformly over the possible hash values to achieve constant expected search time



http://www.kriche.com.ar/root/programming/data%20structures%20&%20algorithms/hashTable.html

Cryptographic Hash Functions

- A cryptographic hash function satisfies additional properties
  - Preimage resistance (aka one-way)
    - Given a hash value x, it is hard to find a plaintext P such that h(P) = x
  - Second preimage resistance (aka weak collision resistance)
    - Given a plaintext P, it is hard to find a plaintext Q such that h(Q) = h(P)
  - Collision resistance (aka strong collision resistance)
    - It is hard to find a pair of plaintexts P and Q such that h(Q) = h(P)
- Collision resistance implies second preimage resistance
- Hash values of at least 256 bits recommended to defend against brute-force attacks

Birthday Attack

- The brute-force birthday attack aims at finding a collision for a hash function h
  - Randomly generate a sequence of plaintexts X1, X2, X3,…
  - For each Xi compute yi = h(Xi) and test whether yi = yj for some j < i
  - Stop as soon as a collision has been found
- If there are m possible hash values, the probability that the i-th plaintext does not collide with any of the previous i -1 plaintexts is 1 - (i - 1)/m

Birthday Attack

- Probability F_k that the attack fails (no collisions) after k plaintexts is

$$F_k = (1 - 1/m)\,(1 - 2/m)\,(1 - 3/m)\,\ldots\,(1 - (k - 1)/m)$$

- Using the standard approximation $1 - x \approx e^{-x}$

$$F_k \approx e^{-(1/m + 2/m + 3/m + \ldots + (k-1)/m)} = e^{-k(k-1)/2m}$$

- The attack succeeds/fails with probability ½ when $F_k = ½$, that is,

$$e^{-k(k-1)/2m} = ½$$

$$k \approx 1.17\, m^{½}$$

  - We conclude that a hash function with b-bit values provides about b/2 bits of security

Secure Hash Algorithm (SHA)

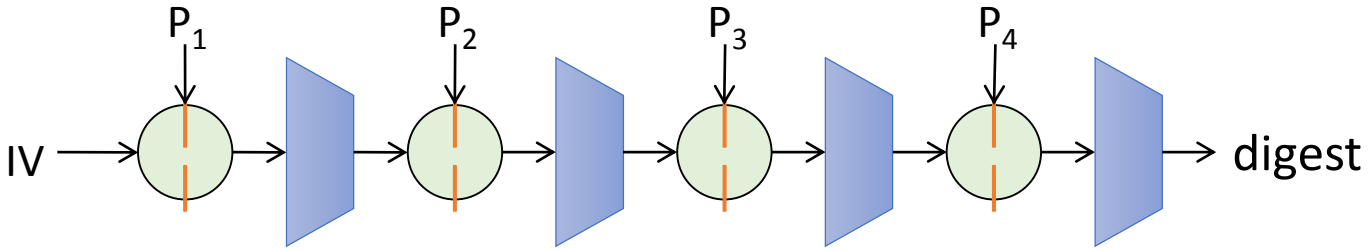
- Developed by NSA and approved as a federal standard by NIST
- SHA-0 and SHA-1 (1993)
  - 160-bits
  - Considered insecure
  - Still found in legacy applications
  - Vulnerabilities less severe than those of MD5
- SHA-2 family (2002)
  - 256 bits (SHA-256) or 512 bits (SHA-512)
  - Still considered secure despite published attack techniques

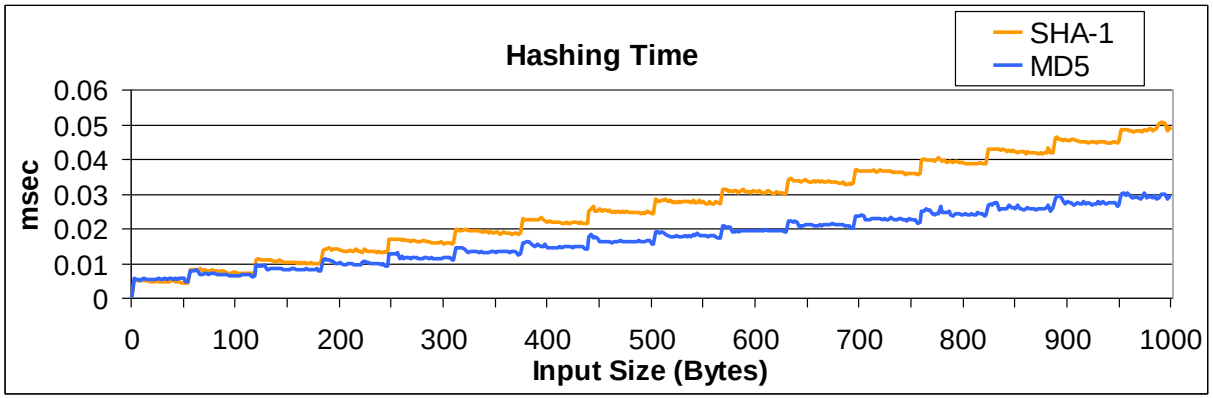
Secure Hash Algorithm (SHA)

- SHA-3 (2015)
  - More complex and secure hash algorithm
  - Faster than SHA-1 and SHA=2
- Older hash functions include MD4, MD5
  - Should never be used
  - Only SHA-2 AND SHA-3 should be used

Iterated Hash Function

- A compression function works on input values of fixed length
- An iterated hash function extends a compression function to inputs of arbitrary length
  - padding, initialization vector, and chain of compression functions
  - inherits collision resistance of compression function
- MD5 and SHA are iterated hash functions





- Questions?