

DATA SECURITY and INFORMATION THEORY for the XXI CENTURY

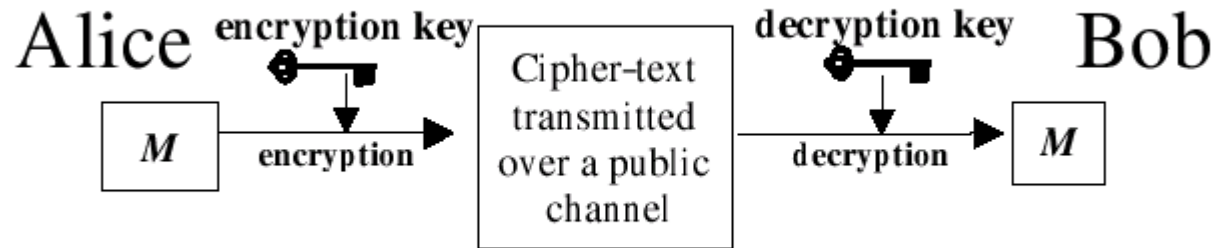
presented by

Mario Forcinito, Ph.D., P.Eng.

Definitions

- Cryptology: from the Greek words: *kryptos*, meaning “hidden,” and *ology*, meaning “science.” the science concerned with communications in secure and usually secret form. It encompasses both:
 - Cryptography (from the Greek *graphia* meaning “writing”) and ;
 - Cryptanalysis, or the art of extracting the meaning of a cryptogram.

The Basic Idea of Cryptography

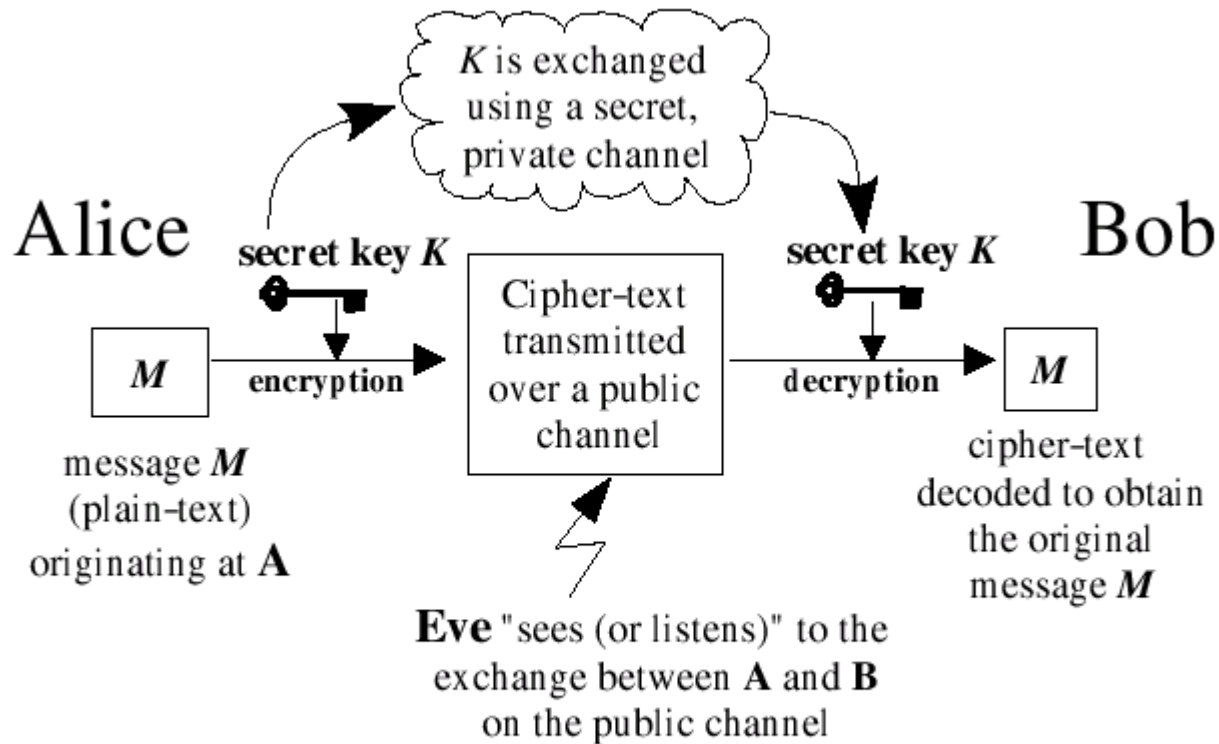


M : plain-text (the message)

C : cipher-text (scrambled message)

Remarks

- **B** need to be the sole possessor of the decryption key (apart, maybe, from **A**)
- In general a key is a mathematical object or a binary string although keys can also be other objects such as the entire Navajo language adapted for encryption purposes during WWII.
- Enciphering and deciphering operations are usually mathematical procedures (algorithms).
- Using encryption for storing messages and files is another important application for encryption algorithms.



Symmetric Key Encryption

- Encryption and Decryption keys are identical.
- In many cases the algorithm for decryption is identical to the algorithm for encryption.

Algorithms for Symmetric Encryption

- DES (Data Encryption Standard)
1970's 56 bits key
- 3-DES (Triple DES)
- AES (Advanced Encryption Standard)
2000 - 128 and 256 bits key
- One Time Pad Perfect Secrecy
1920's - key is as long as the message
- Linear Feedback Shift Register (Stream Ciphers)

DES, AES as well as most block ciphers combines the two fundamental techniques for construction of ciphers advocated by Shannon in 1949, namely, **confusion** and **diffusion**.

Confusion

This technique blocks the cryptanalyst from obtaining statistical patterns and redundancies in the cipher-text arising from the plain-text.

The statistical dependency of the cipher text on the plain-text is obfuscated.

The easiest way to cause confusion is through the use of substitutions.

Diffusion

This technique dissipates the redundancy of the plain-text by spreading it over the cipher-text.

Diffusion implies that, if we change just one character in the plain text, we cause a big change in the cipher-text.

A cryptanalyst will need a large amount of cipher-text to capture redundancy in the plain text.

Diffusion doesn't simply means a permutation or rearrangement of the characters in the message string.

Shannon's recipe for diffusion also involves acting on a string with a diffusing function.

In general, the cipher texts in DES and AES will reveal some information about the message, the amount depending on the nature of the message.

Perfect Secrecy or Perfect Security

is achieved when knowledge of the cipher text reveals no information whatsoever about the message, as is the case with the Vernam cipher (One-time pad).

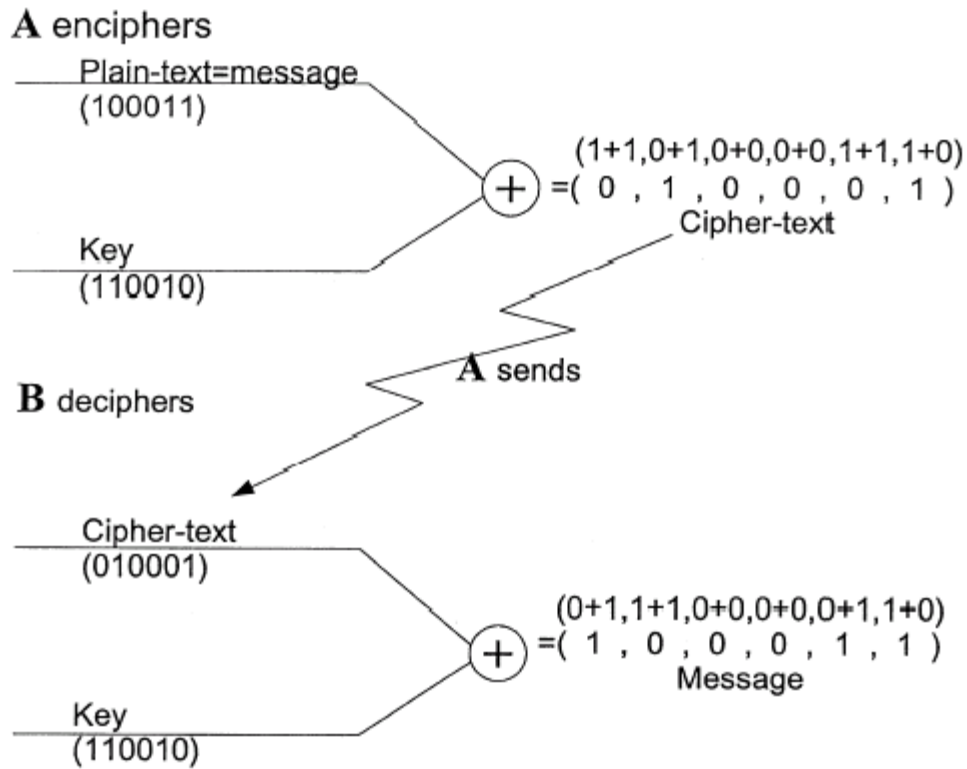
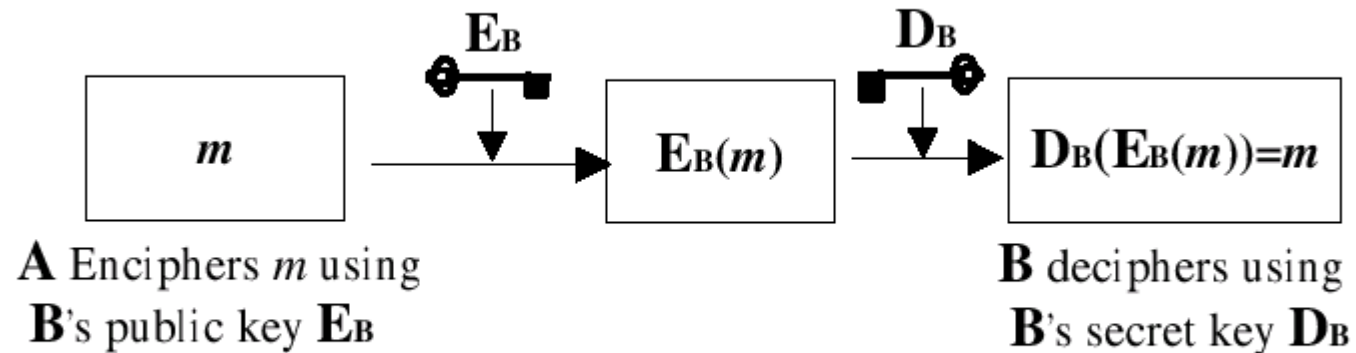


Figure 4.1: One-Time Pad



Asymmetric or Public Key Encryption

- Encryption and Decryption keys are different.
- Encryption key is public knowledge.
- Decryption key must remain secret.

The RSA Algorithm.

1. Bob chooses in secret two large primes p, q with $p \neq q$ and sets $N = pq$.
2. Bob chooses e bigger than 1 with e relatively prime to $p - 1$ and to $q - 1$ and with $e < (p - 1)(q - 1)$.
3. Bob calculates the decryption index d where $d < (p - 1)(q - 1)$ is such that the remainder of de on division by $(p - 1)(q - 1)$ is 1. More generally, Bob calculates a decryption index d where $d < t \leq (p - 1)(q - 1)$ is such that the remainder of de on division by t is 1. Here, t is any number divisible by both $p - 1$ and $q - 1$.
4. Bob announces his public key $[N, e]$ and keeps his private key d secret.
5. Alice wishes to send a secret message M and represents M as a number between 0 and $N - 1$. Alice then encrypts the message M as the remainder C of M^e upon division by N and transmits C to Bob.
6. Bob decrypts C by calculating the remainder of C^d upon division by N : this gives the original secret message M .

Example: RSA on a calculator

Encryption algorithm: “multiply M by itself e times and take the remainder of this number when divided by N to be the cipher text C ”.

- Example $N = p * q = 11 * 5 = 55$, encryption exponent $e = 7$, message $M = 6$
 - calculate $6^7 = 279, 936$.
 - $279, 936 = 55y + z$ (where z is one of $0, 1, 2, 3, 4, \dots, 53, 54$).
- We are not really interested in the value of y ; we just need z .

$$279, 936 / 55 = 5089.74545454$$

$$y = 5089 \text{ and } z/55 = 0.745454$$

$$\text{so } z = 55(0.745455) = 41.0000$$

$$C=41$$

Example: RSA Decryption

Decryption algorithm: “Calculate the remainder of C^d upon division by N ”

- We need the unique integer d_1 , between 1 and 20, such that $7d_1$ gives a remainder of 1 when divided by 20.

Why 20, instead of $(p-1)(q-1) = 10*4 = 40$?

- Because of the properties of the algorithm we can use any positive integer divisible by both $p - 1 = 10$ and $q - 1 = 4$.

Choose 20, and get $d_1 = 3$.

- It is much easier to use the decryption index 3 instead of the decryption index 23 ($7*23 = 40*4 + 1$)

$$41^3 = 68921$$

$$68921/55 = 1253.109090909$$

$$0.1090909*55 = 6 = M$$

Remarks on RSA

- Based on the unproven assumption that given N and e and C one cannot factor $N = p * q$ in a reasonable amount of time (Computationally secure)
- Keys are long (1024 bits is today's standard for minimum security).
- Long computation even for small messages.
- Used in many electronic signature schemes and to exchange keys for symmetric algorithms without the need of a secret channel.

Encrypting e-mail

- PGP : Pretty Good Privacy uses patented symmetric algorithms.
- GPG : Gnu Project Guard (free, open source software) uses public domain algorithms.
- SSL : Secure Socket Layer allows encryption on-line between unknown parties.
- Encryption algorithm for PGP, GPG and SSL are very similar, all based on RSA.

Encrypting e-mail

Procedure:

- generate a key pair
- publish public key (e-mail or key-server)
- select a symmetric algorithm
- generate a random key
- encrypt the key with public key of the recipient
- Send the message and the encrypted key by email to the recipient.

Asymmetric encryption is used for the key exchange and Symmetric encryption for the actual message.

Symmetric encryption is about 4000 times faster.

Signing / Authenticating

- A hash of the email message is encrypted with the sender's private key and appended to the end of the message.
- The recipient's computer may decrypt this message with the sender's public key.
- Check that the hash corresponds to the hash of the message.
- Authenticates the original message sender (the person in possession of the private key used to encrypt the message hash).
- Almost guarantees that the message wasn't altered since its signing.

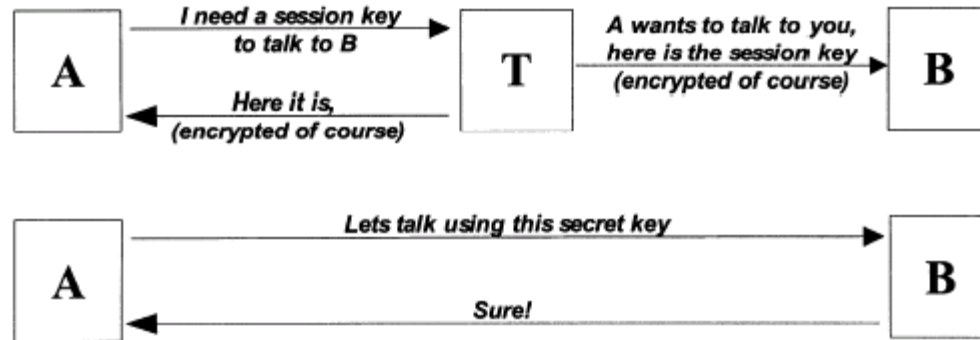


Figure 4.6: The basic Kerberos scheme for trusted server authentication.

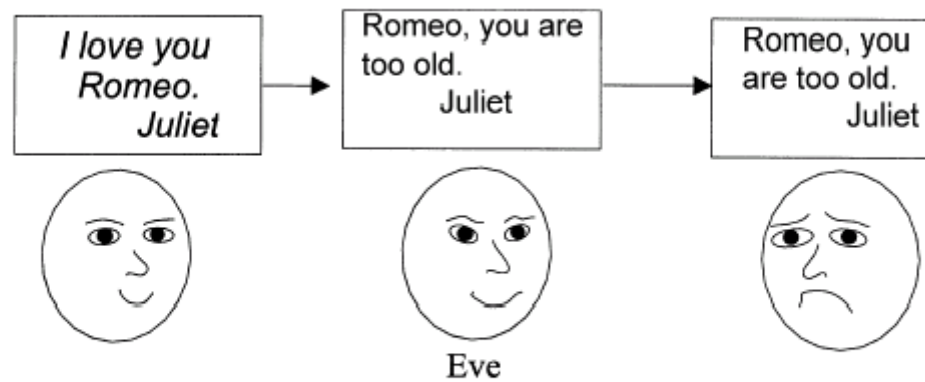


Figure 4.2: How does Romeo know that Juliet's message has not been altered?

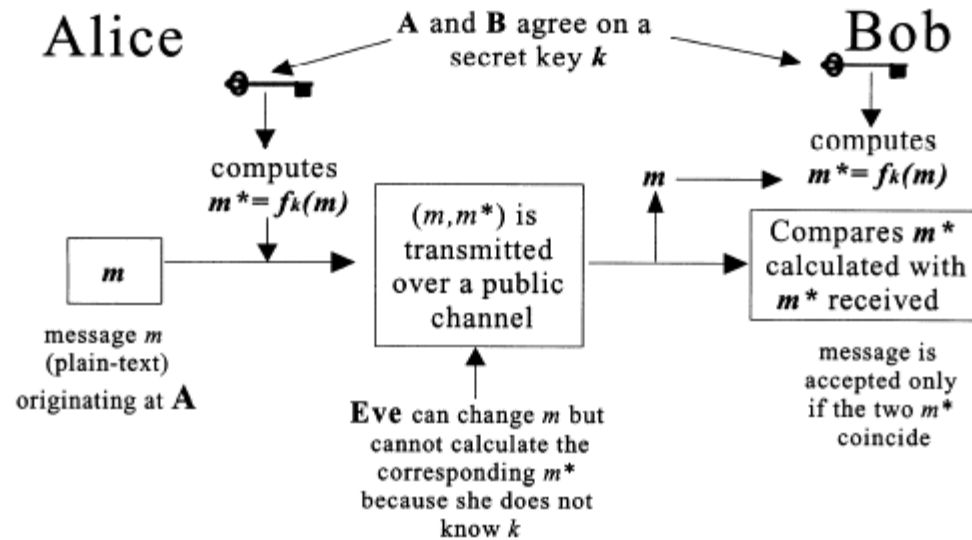


Figure 4.3: Authentication and message integrity check to ensure that nobody tampered with the message from A to B.

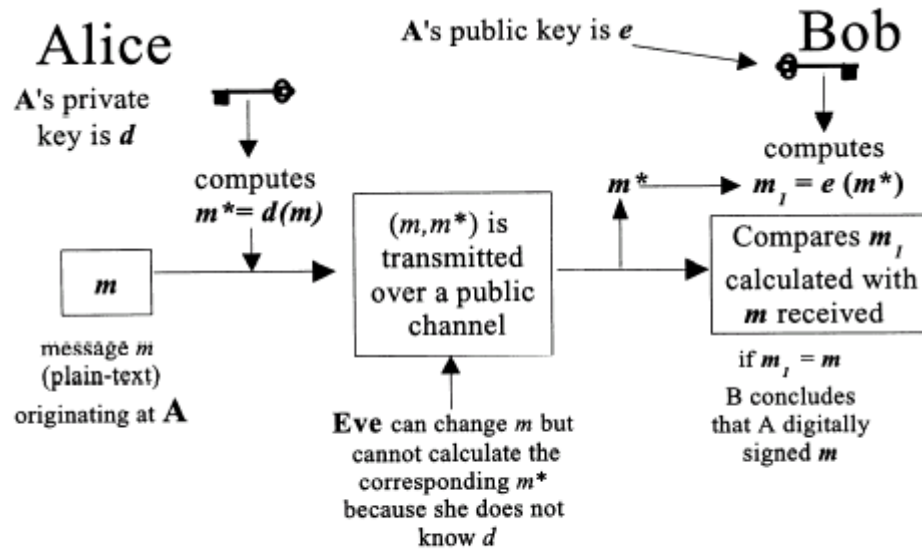


Figure 4.5: Digital Signature scheme using Public Key Cryptography.

Hash Functions

A hash function transform a string (message) M of variable length into a shorter binary string M_1 of a fixed length.

M_1 is a digest or “snapshot” of M and is called the hash of M .

M_1 can easily be calculated from M — but not the other way around!

A desirable feature of hash functions is that each bit of the output depends — in a complicated way — on all the bits in the input. (Shannon’s confusion and diffusion for hash functions).

Collisions necessarily occur (a collision being two different inputs that hash to the same output).

Secure Hash Functions

- (a) Given an output M_1 it is not computationally feasible to find an input M such that $f(M) = M_1$
- (b) It is computationally not feasible to find collisions, i.e., to find messages $x_1 \neq x_2$ such that $f(x_1) = f(x_2)$.

Examples:

MD4, MD5 which hashes to 128 bits

SHA1 which hashes to 160 bits

Practical Issues

- Technical

(hardware/software/protocol) must be designed for a wide variety of conditions beyond designers' control.

- Commercial

The business model used to commercialize new applications must make sense.

- Property Rights

The real value of a patent resides on having a good grasp of issues related to the prior art, the correct mathematics and proper engineering practices.

- Legal

The requirement to encrypt and protect certain data has now become the law in California and it is likely that other jurisdictions

Hot Issues

- Authentication
 - User identification/Biometrics
- E-Commerce
 - Digital money
- E-Government
 - How much is too much?/ Identification issues
- Key Lengths
 - RSA / ECC / symmetric encryption
- Digital Right Management
 - The next step
- Wireless Networks
 - 802.11
- Communication Protocols
 - (TCP/IP) etc.