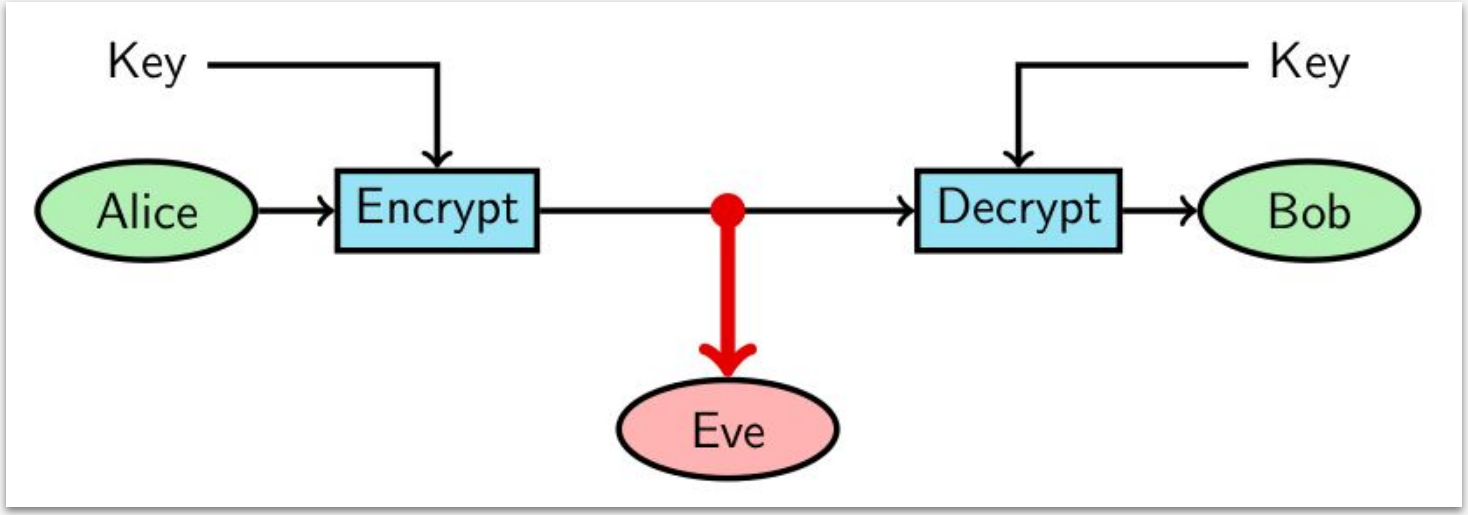
Cryptography

Mustakimur R. Khandaker

Origin of Cryptography

"Cryptography" is a Greek word that means "hidden writing".

Used to hide message from someone, and sometimes prevent them from creating a new message.



Cryptography

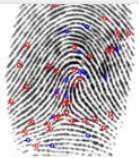
A security tool, not a general solution.
- Reliable if implemented properly.

Cryptography usually converts a communication security problem into a key management problem.
- Reliable unless used improperly.

Hide information securely, opens with identical keys (symmetric key cryptography)

Make something visible, but hinder changes (signatures)

Everyone can hide information, opens with private key (asymmetric key cryptography)



Enable simple checking of data (hash functions)

So now you must take care of the key security problem, which becomes a problem of computer security.

Terminology

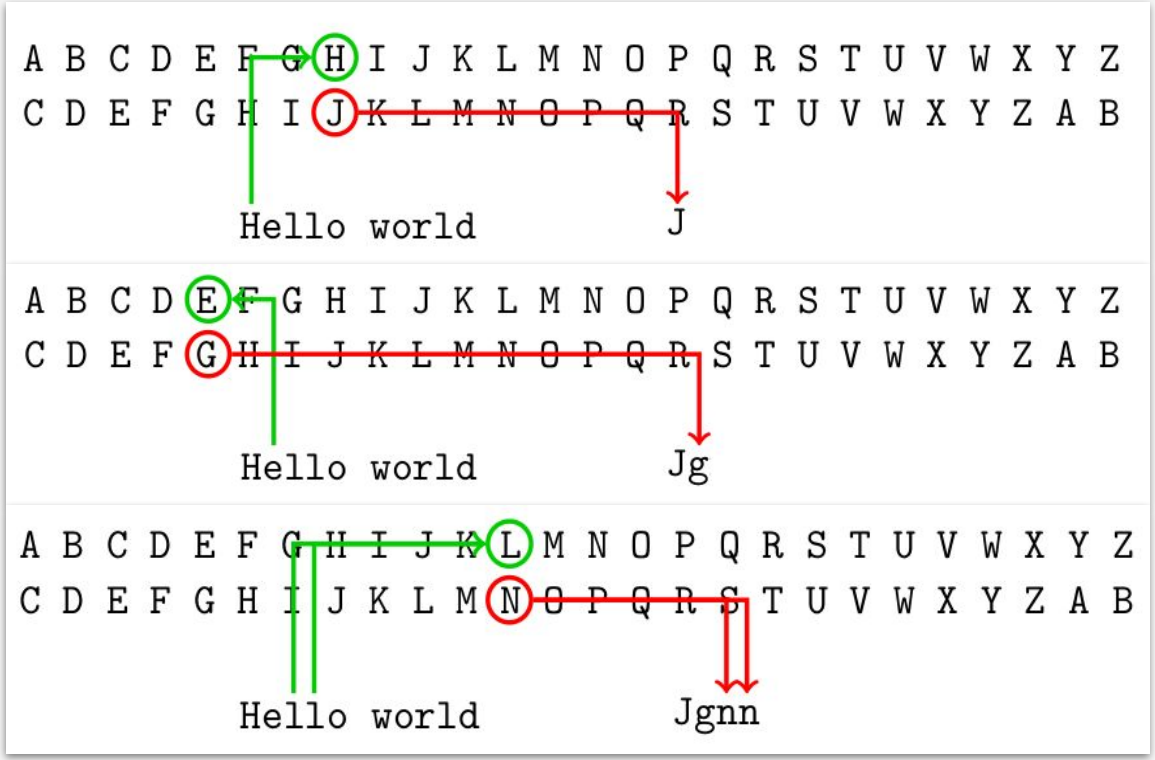
- The **plaintext** is the information in its normal form.
- The **ciphertext** is the transformed plaintext.
- The secret parameter for the encryption (known only to the sender and intended recipients) is called the **key**.

Kerckhoff's principle

- *A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.*

Caesar (Shift) Cipher

Exchange every plaintext letter into the letter x positions further on in the alphabet.



Caesar Cipher Computation

Replace every plaintext letter with its (zero-offset) position in the alphabet (“A”=0, “B”=1, etc., up to the number of letters n).

Express the key as an integer **k** using the same system.

If the plaintext as an integer is **m**, the cryptogram as an integer

```
c = m + k modulo n
```

The cryptogram letter is then the letter corresponding to the number **c**.

The plaintext “H” gives **m** = 7, and **k** = 2 results in **c** = 7 + 2 (mod 26), so ciphertext is “J”.

Substitution Cipher

Create a table of plaintext characters and their corresponding crypto characters.

- Crypto characters can be just ordinary letters, but also anything else.
- Each crypto character must occur only once in the table to enable unique decryption.

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
G	Z	E	J	D	Y	I	T	Q	A	U	M	B	W	R	F	C	X	H	N	S	L	O	K	P	V

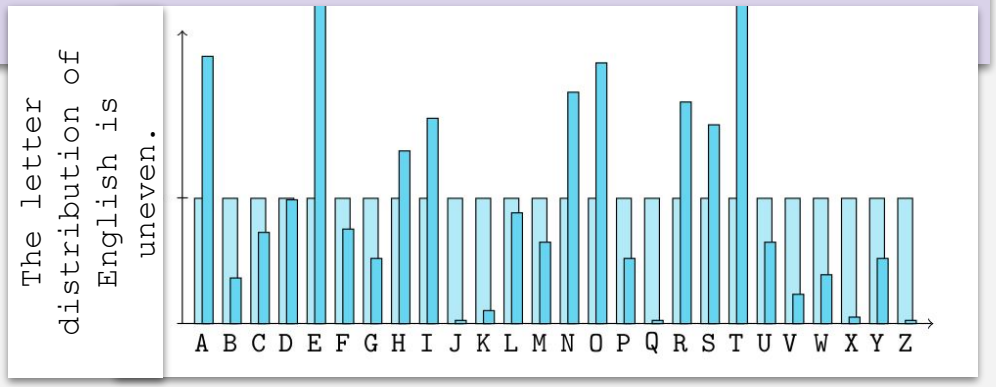
Breaking Simple Substitution

Every crypto letter will occur exactly as often as its plaintext counterpart occurs in plaintext.

Every combination of crypto letters (digrams, trigrams etc.) will occur as often as the corresponding plaintext combinations.

Count how often letters, bigrams and trigrams occur in the cryptogram, and try to identify the ones corresponding to common plaintext letters and common letter combinations.

Fill in so that remaining gaps form words.



Algorithm Strength

Algorithm strength is measured in the amount of work needed to break the cipher.

- The comparison is with brute force.
- For Caesar crypto, you need to check 26 keys, or just under 2^6 values.
- For simple substitution you need to check $26!$ keys ($\approx 4 \times 10^{26}$), or just over 2^{88} values.
- But with letter statistics this drops quickly.

There are 31536000 seconds in a year $\approx 2^{25}$.

- If you can try one key every microsecond, you can find a 45-bit key in around one year ($10^6 \approx 2^{20}$)
- If you have 1000 such processors, you can find a 55-bit key in around one year.

Modern Ciphers

Modern ciphers are immune to simple attacks.

- Algorithm weakness is almost never a problem with approved algorithms today.
- Badly chosen keys, bad key management, and bad implementations are the current problems.

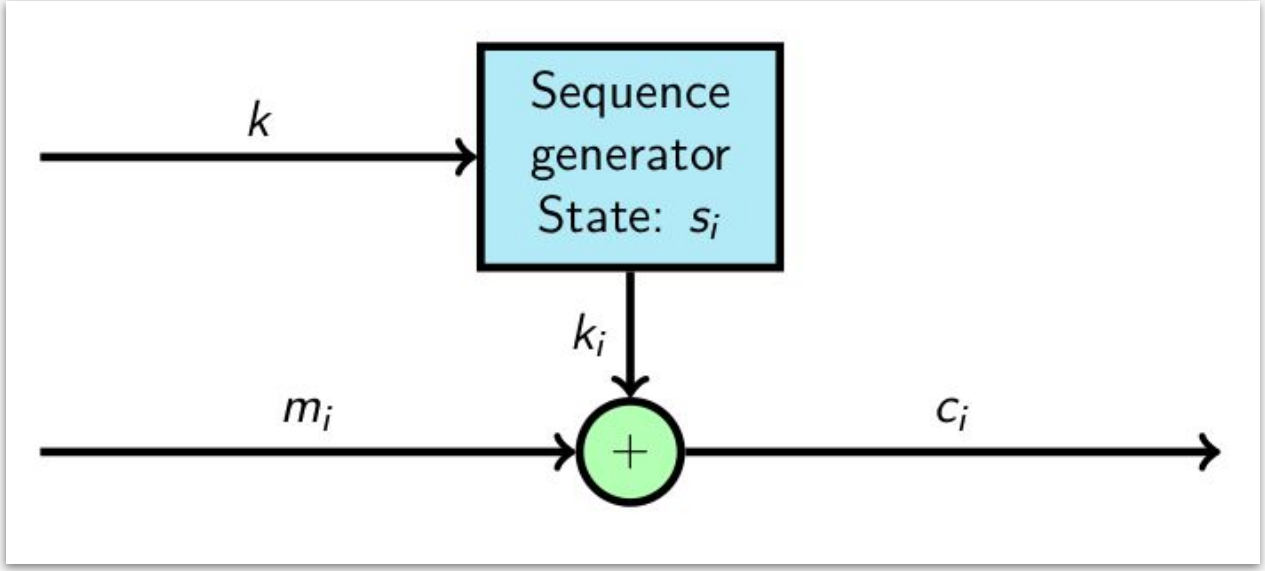
There are three basic algorithm types:

- **Stream ciphers**, very fast and suited for confidentiality, not data integrity.
- **Block ciphers**, can be *symmetric key* which are fast, and *asymmetric* that are slower, both suited for confidentiality and integrity (with some precautions).
- **One way functions**, fast, only for data integrity.

Stream Cipher: The One-time-pad

Useful for sending secret data without preserved data integrity.

- Use a long *pseudo*-random bit sequence as key, generated from a short *seed* (key) **k** and initial state s_0.
- Add this to the plaintext to form the ciphertext.



Evaluation of one-time pad

Advantages

- Easy to compute encrypt, decrypt from key, text.
- As hard to break as possible.
  - This is an information-theoretically secure cipher.
  - Given ciphertext, all possible plaintexts are equally likely, assuming that key is chosen randomly.

Disadvantages

- Key is as long as the plaintext.
  - How does sender get key to receiver securely?

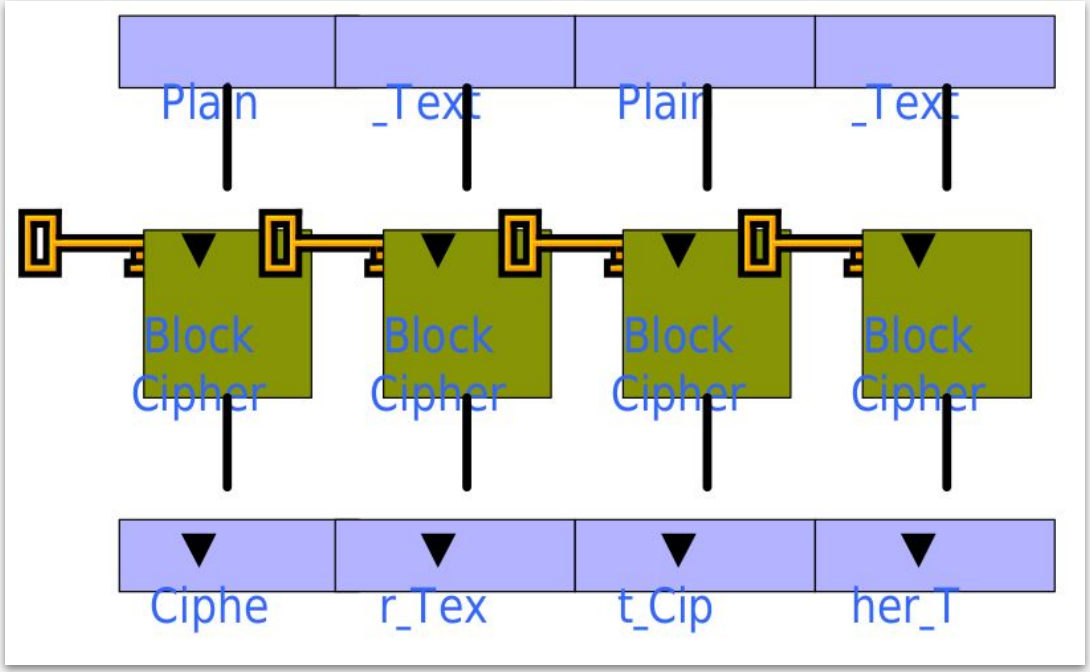
Block Cipher: Electronic Code Book

In short, also known as ECB.

- Divide plaintext into blocks.
- Encrypt each block independently, with same key.

Problem:

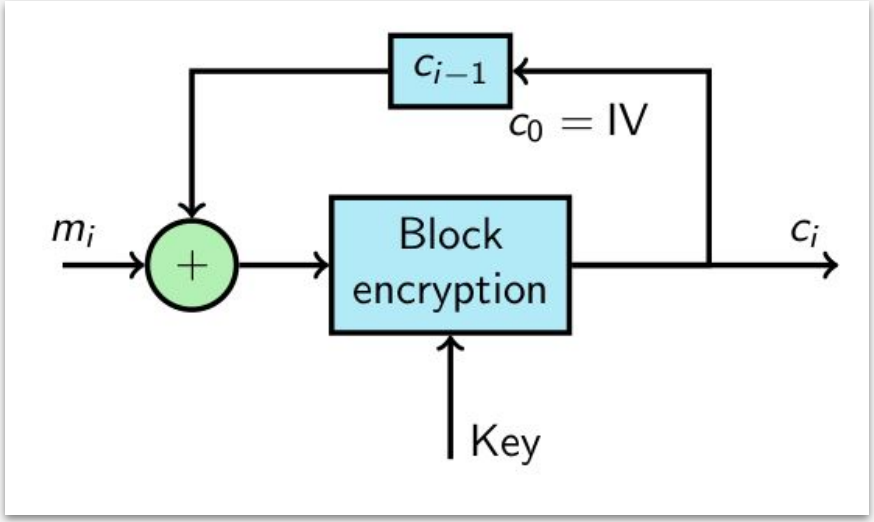
- Identical blocks encrypted identically.



Block cipher: Cipher Block Chaining

Most common mode to preserve data integrity.

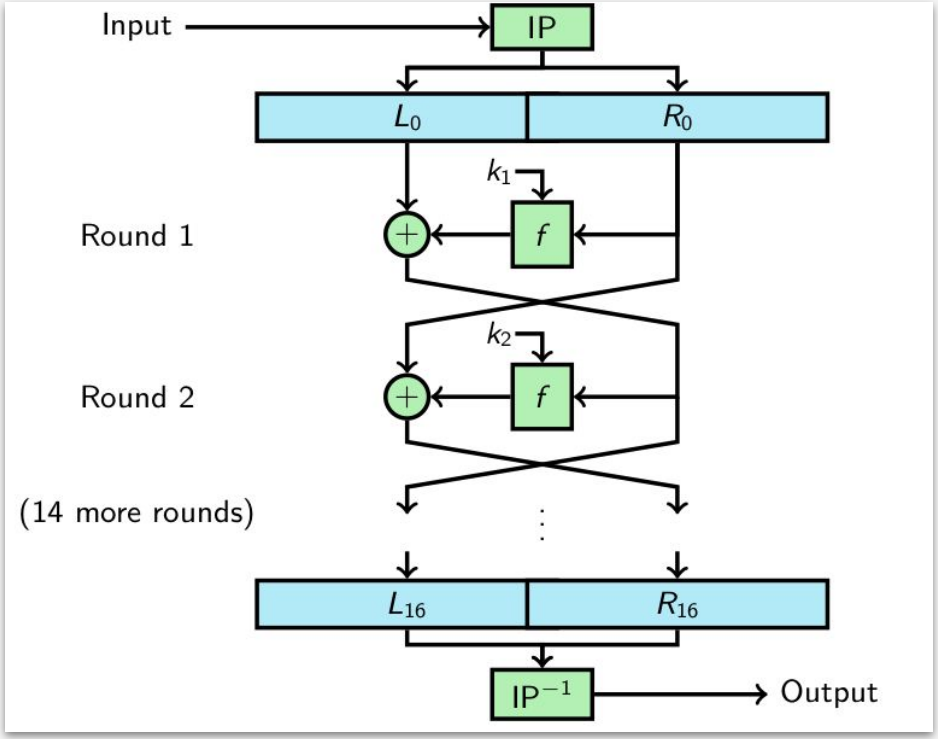
- One whole block at a time.
- Feedback of previous cipher block.
- Propagates transmission errors (one bit transmission error destroys one whole block plus one more bit).



Block Cipher: DES

Messages are treated in blocks of characters with fixed block size.
- The key remains fixed for at least one session.

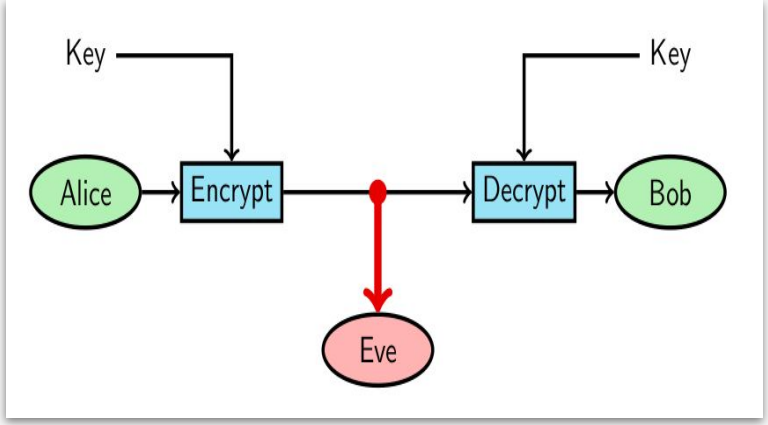
Data Encryption Standard (DES)
- Developed at IBM.
  - widely used.
- Permute input bits.
  - Repeat application of a S-box function.
- Apply inverse permutation to produce output.
- Appears to work well in practice.
- Improvements:
  - Triple DES, AES.

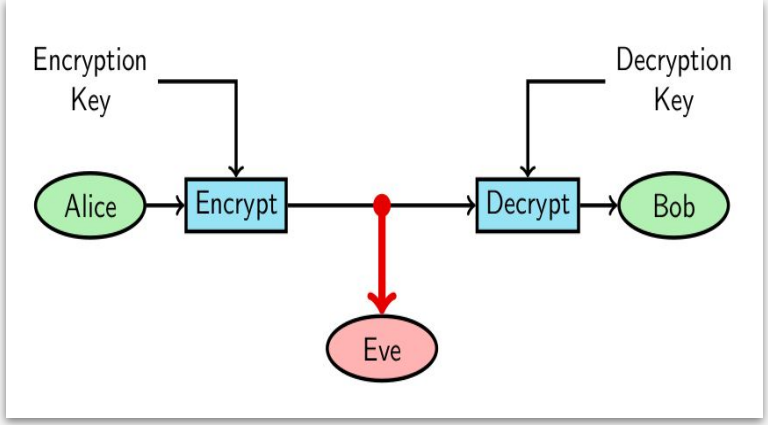


Symmetric vs Asymmetric Key Cryptography

In symmetric key systems, the encryption key is the same as the decryption key.

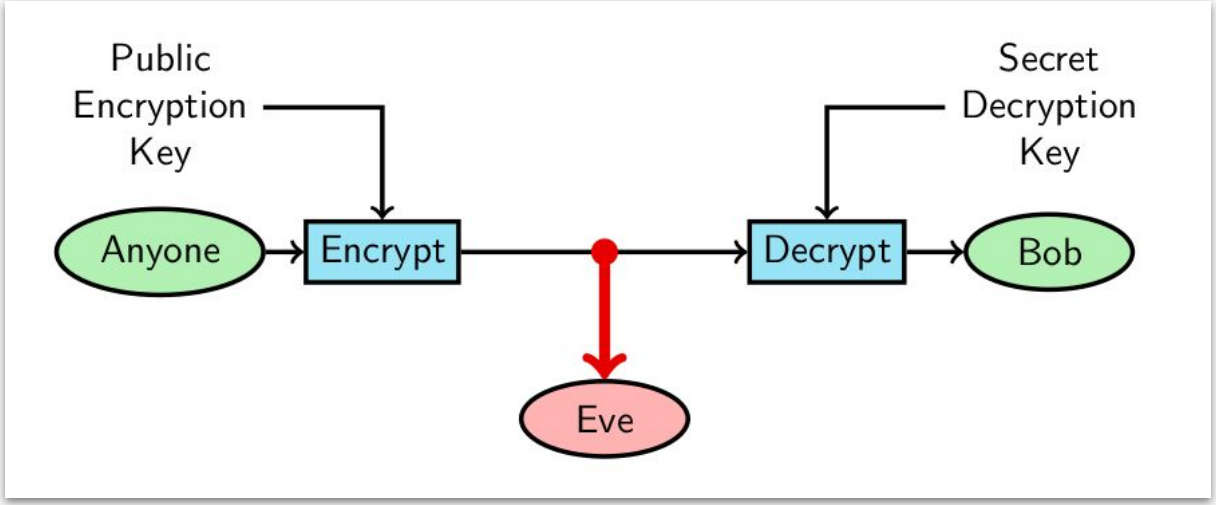
In asymmetric key systems, the encryption key and the decryption key are different.





Public Key Cryptography

The encryption key can be public, so that anyone can send secret messages to Bob.



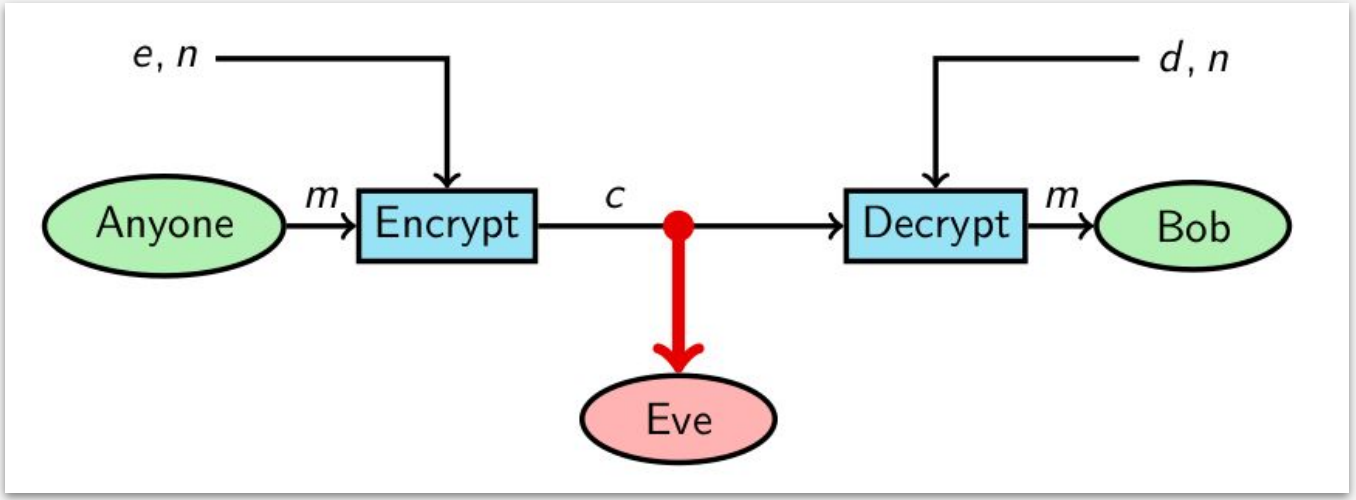
Each user has a pair of keys, one public encryption key ***e***, one secret decryption key ***d***.

- It is computationally hard to calculate ***d*** from ***e***.
- This is achieved by basing the algorithm on a difficult problem from number theory.

Example: RSA

Choose two very large primes ***p*** and ***q***, and publish ***n = pq***

- Choose a public encryption key ***e;*** $(\gcd(e, (p-1)(q-1)) = 1)$
- Calculate the secret decryption key ***d;*** $ed \equiv 1 \bmod (p-1)(q-1)$
- Encrypt with $c = m^e \bmod n$
- Decrypt with $m = c^d \bmod n$



Show Calculation

p = 11
q = 5
n = 55
$\phi(n) = 40$
e = 7
d = 23

How Well Does RSA Work?

Can generate modulus, keys fairly efficiently.

- Efficient rand algorithms for generating primes ***p,q***.
  - May fail, but with low probability.
- Given primes ***p,q*** easy to compute `n=p*q` and `φ(n) = (p - 1) * (q - 1)`

Public key ***n***, ***e*** does not reveal ***d***.

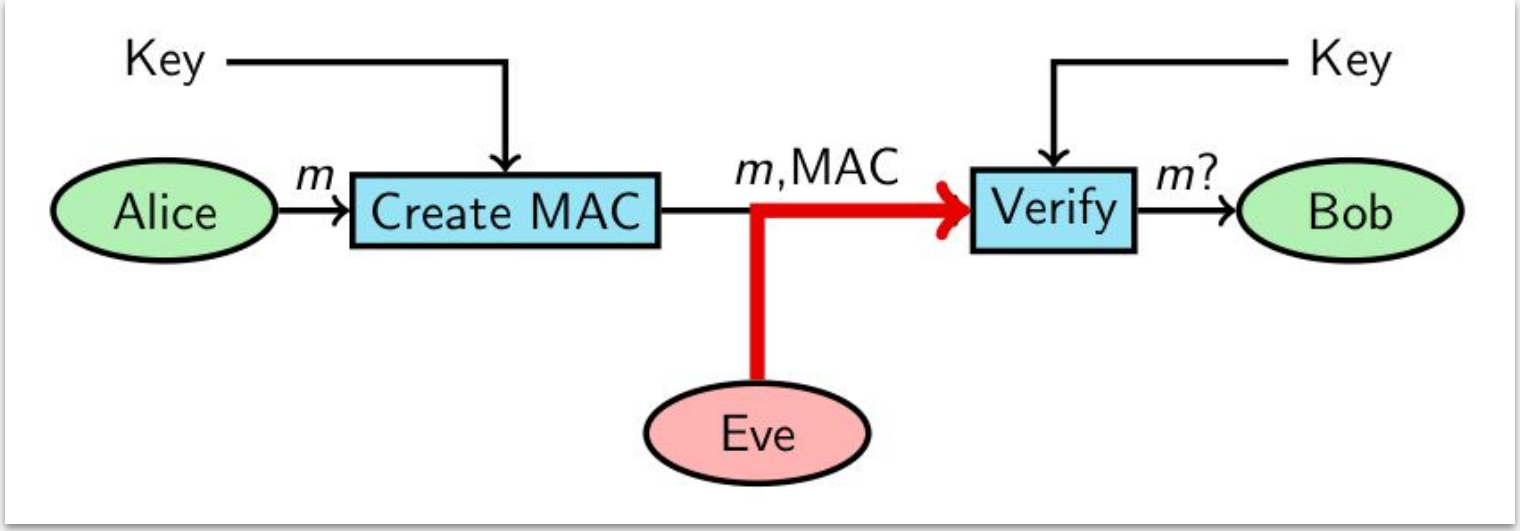
- This is not proven, but believed.

But if n can be factored, all is lost ...

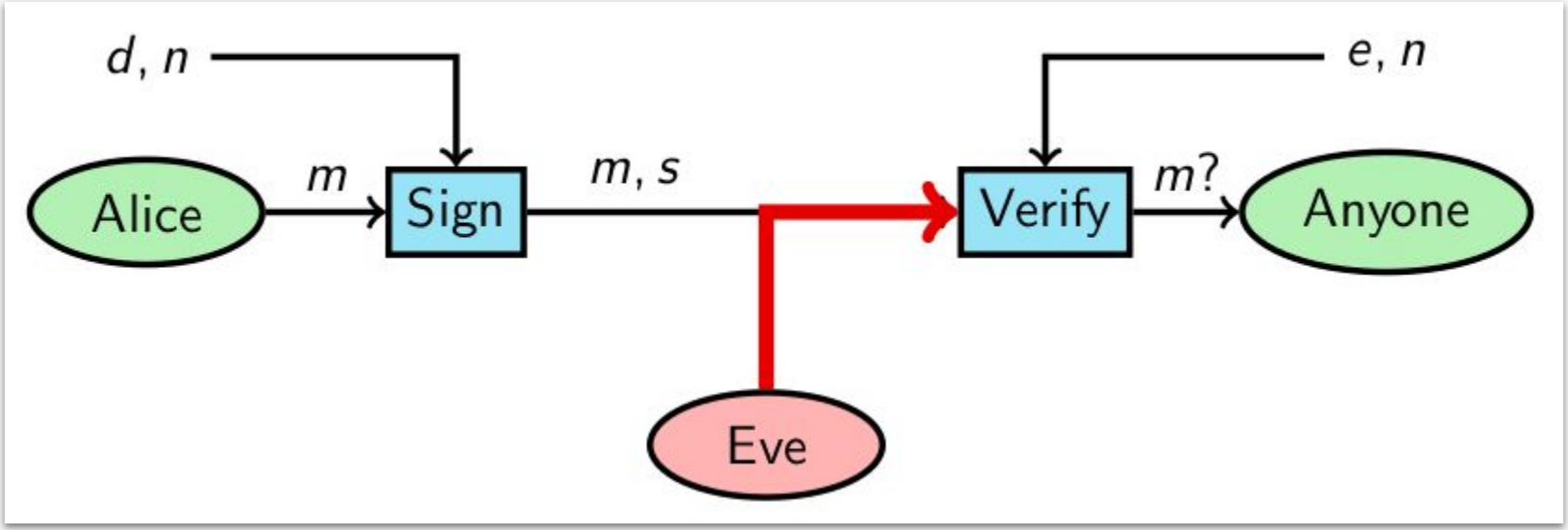
MAC

A Message Authentication Code (MAC) is a “seal of authenticity” created in a symmetric key system.

- If the same MAC can be created at the verifier, then the message hasn’t changed in transit.
- Everyone that can check the MAC can create the MAC.



Digital Signatures



In RSA, the secret *"decryption"* key now is the **secret signing key**, while the public *"encryption"* key now is the **public verification key**.

- The same pair ***e, d*** should not be used both for confidentiality and integrity.

Attacks

Confidentiality considers indistinguishability under…

- Chosen Plaintext Attack (CPA) An attacker can obtain the ciphertext for any provided plaintext (but does not have the key).
- Chosen Ciphertext Attack (CCA) An attacker can obtain the plaintext for any provided ciphertext (but does not have the key).

Integrity

- PTXT - Integrity of Plaintext - computationally infeasible to produce a ciphertext decrypting to a message that the sender had never encrypted.
- CTXT - Integrity of Ciphertext - computationally infeasible to produce a ciphertext not previously produced by the sender.

One-way Hash Functions

Large messages take time to MAC or sign.

A hash function creates an output that is much smaller than the message (or file).

- For any hash function, it should be easy to calculate $h(x)$ from ***x***.
- Then, it is easy to create MACs or digital signatures for $h(x)$.

A one-way function is one where it is easy to calculate $y = f(x)$, but computationally hard to calculate $x = f^{-1}(y)$.

- But a hash function has output much shorter than the input.

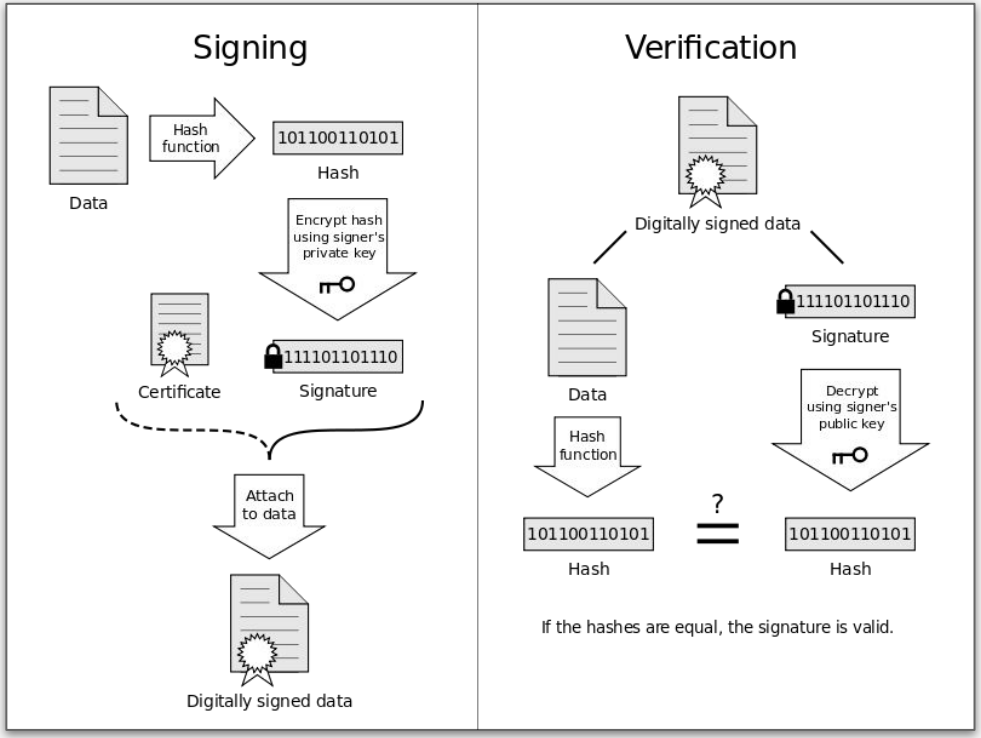
There must be messages for which the output collides (so f^{-1} doesn't exist).

- ***"Collision resistant"*** does not mean that no two messages get the same hash value, it means it is hard to find two such messages.

Example: RSA Digital Signatures

Set up for RSA and choose a collision-resistant hash function.

- Sign using $s = (h(m))^d \bmod n$
- Verify by checking that *h(m)* equals $s^e \bmod n$



Types of Hash Functions

MD5
- 128-bit output.
- Designed by Ron Rivest, used very widely.
- Collision-resistance broken (summer of 2004 and it keeps getting worse).

RIPEMD-160
- 160-bit variant of MD5.

SHA-1 (Secure Hash Algorithm)
- 160-bit output.
- US government (NIST) standard as of 1993-95.
- Also the hash algorithm for Digital Signature Standard (DSS).

How Strong is SHA-1?

Every bit of output depends on every bit of input.
- Very important for collision resistance.

Brute-force inversion requires 2^{160} ops, birthday attack on collision resistance requires $2^{80.}$

Recent weaknesses (2005)
- Collisions can be found in 2^{63} ops - maybe 2^{60} People are losing confidence.

Password Security Review

Summarize system:

- ❑ Identify assets: What do you wish to protect.
- ❑ Identify adversaries and threats.
- ❑ Identify vulnerabilities.
- ❑ Calculate the risks.
- ❑ Evaluate controls/mitigation strategies.
- ❑ Iterate.

Vulnerabilities:

- ➔ Online guessing/dictionary attack.
- ➔ Offline guessing/dictionary attack.
- ➔ Shared passwords.
- ➔ Password fallback schemes.

Mitigation Strategies

For traditional password management:
- Salts.
- Encrypted Storage.
- Challenge/Response.

Alternative to passwords:
- Graphical passwords, phrases.
- Tokens/dongles.
- Biometrics.

Multi-factor authentication:

"Something you forget, something you lose, and something you used to be."

- Who has used 2 (or more) factor auth?

Public-Key Infrastructure (PKI)

Anyone can send Bob a secret message.
- Provided they know Bob's public key.

How do we know a key belongs to Bob?
- If imposter substitutes another key, read Bob's mail.

One solution: PKI
- Trusted root authority (VeriSign, IBM, United Nations).
  - Everyone must know the verification key of root authority.
- Root authority can sign certificates.
- Certificates identify others, including other authorities.
- Leads to certificate chains.

De facto Standard Algorithms

NIST (National Institute for Standards and Technology) registers standards (algorithms) for federal use in the US.

- Stream ciphers: The A family in GSM and E family in Bluetooth.
- Symmetric block ciphers: DES (56 bits), AES (128-256 bits), ...
- Asymmetric ciphers: RSA (any key size) and Elliptic Curve Cryptography (ECC, any key size), ...
- Cryptographic hashes: MD5 (128 bit hashes, weak), SHA-1 (160 bit hashes, deprecated), SHA-2 (224-512 bit hashes), ...

Limitation of Cryptography

Most security problems are not crypto problems.

- This is good.
  - Cryptographically works!
- This is bad.
  - People make other mistakes; crypto does not solve them.

Example:

- Deployment and management problems.
- Ineffective use of cryptography.

Security failures not publicized.

< Cryptography />