Crypto basics

Acknowledgement

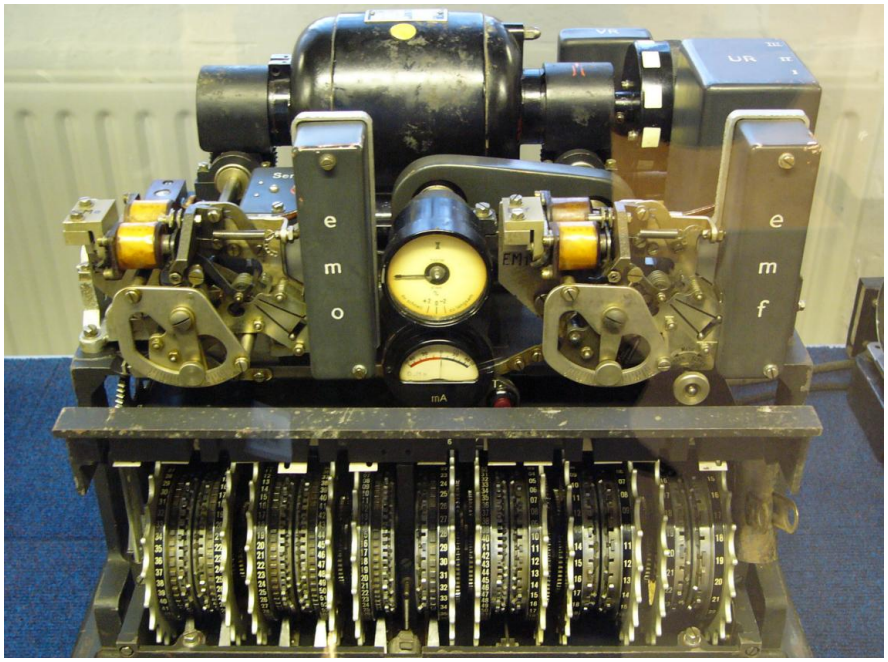
- Original slides made by
  - Patrik Okui <pokui@psg.com>
  - Sheryl Hermoso <sheryl@apnic.net>
  - Yoshinobu ‘maz’ Matsuzaki <maz@iij.ad.jp>
  - Randy Bush <randy@psg.com>

Overview

- What is Cryptography?
- Symmetric Key Cryptography
- Asymmetric Key Cryptography
- Block and Stream Cipher
- Digital Signature and Message Digest

Cryptography

- Cryptography is everywhere



German Lorenz cipher machine

Cryptography

- Cryptography deals with creating documents that can be shared secretly over public communication channels
- Other terms closely associated
  - Cryptanalysis = code breaking
  - Cryptology
    - Kryptos (hidden or secret) and Logos (description) = secret speech / communication
    - combination of cryptography and cryptanalysis
- Cryptography is a function of plaintext and a cryptographic key

$$C = F(P, k)$$

Notation:

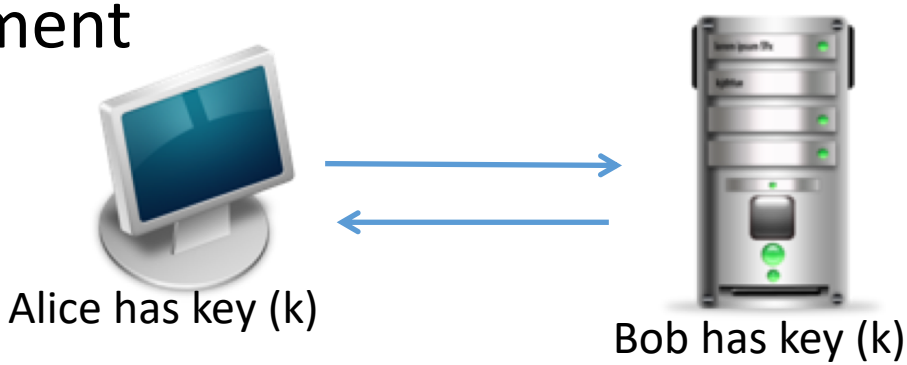
Plaintext (P)
Ciphertext (C)
Cryptographic Key (k)

Typical Scenario

- Alice wants to send a “secret” message to Bob
- What are the possible problems?
  - Data can be intercepted
- What are the ways to intercept this message?
- How to conceal the message?
  - Encryption

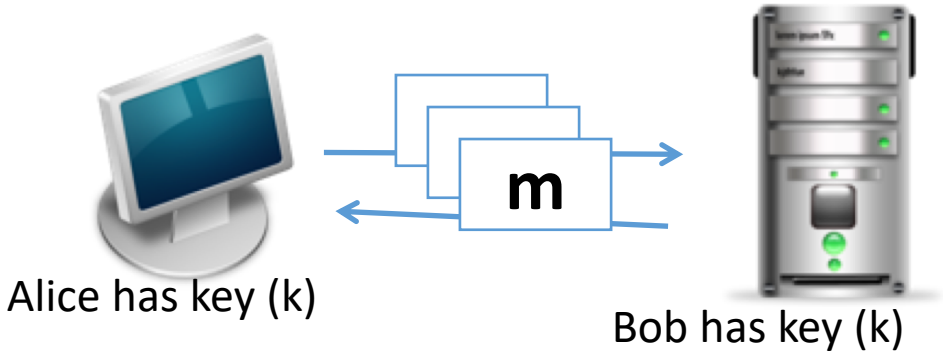
Crypto Core

- Secure key establishment



- Secure communication

Confidentiality and integrity

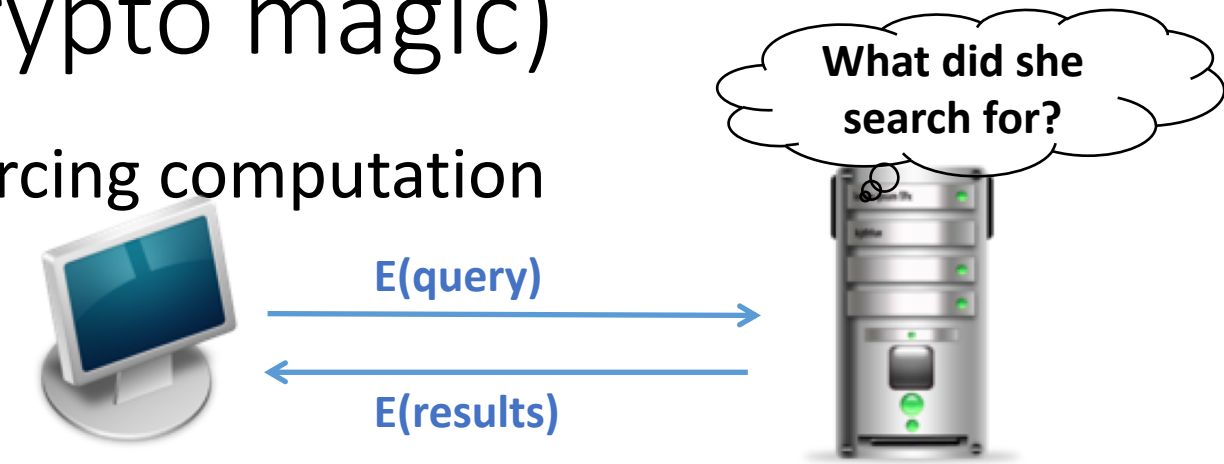


It can do much more

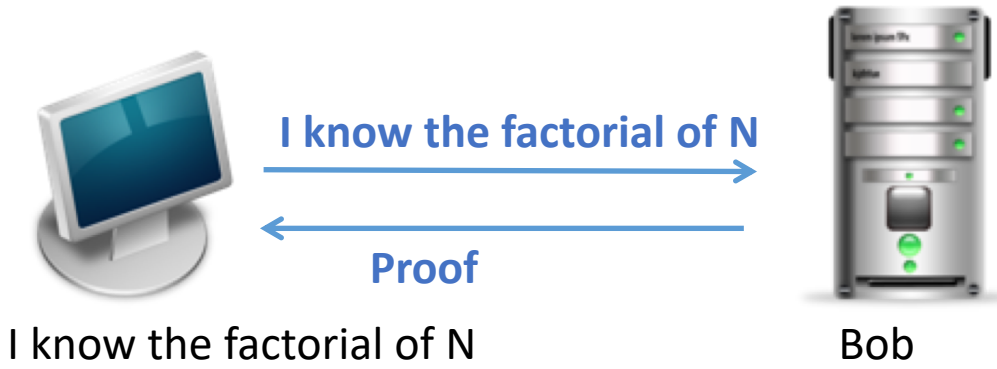
- Digital Signatures
- Anonymous communication
- Anonymous digital cash
  - Spending a digital coin without anyone knowing my identity
  - Buy online anonymously?
- Cryptocurrency
- Elections and private auctions
  - Finding the winner without actually knowing individual votes (privacy)

Other uses are also theoretically possible (Crypto magic)

- Privately outsourcing computation



- Zero knowledge (proof of knowledge)



Source: Dan Boneh, Stanford

History: Ciphers

- Substitution cipher
  - involves replacing an alphabet with another character of the same alphabet set
  - Can be mono-alphabetic (single set for substitution) or poly-alphabetic system (multiple alphabetic sets)
- Example:
  - Caesar cipher, a mono-alphabetic system in which each character is replaced by the third character in succession
  - Vigenere cipher, a poly-alphabetic cipher that uses a 26x26 table of characters

Caesar Cypher

ABCDEFGHIJKLMNOPQRSTUVWXYZ

DEFGHIJKLMNOPQRSTUVWXYZABC

Plain text HELLO

Encrypted text KHOOR

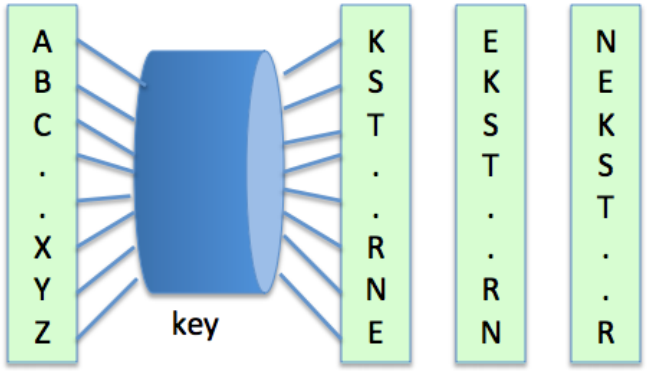
Transposition Cipher

- No letters are replaced, they are just rearranged.
- Rail Fence Cipher – another kind of transposition cipher in which the words are spelled out as if they were a rail fence.

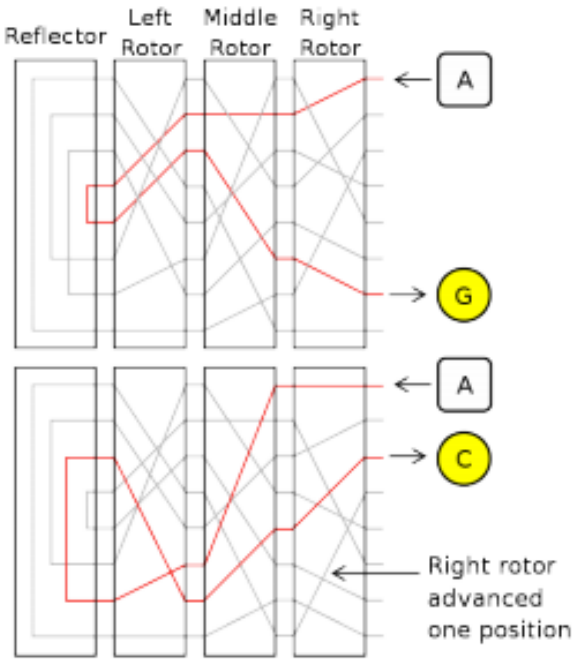
```
T...U...B...N...J...E...E...E...Y..
.H.Q.I.K.R.W.F.X.U.P.D.V.R.H.L.Z.D.G.
..E...C...O...O...M...O...T...A...O
```

History: Rotor Machines (1870-1943)

- Hebern machine – single rotor



- Enigma - 3-5 rotors



Source: Wikipedia (image)

Modern Crypto Algorithms

- Specifies the mathematical transformation that is performed on data to encrypt/decrypt
- Crypto algorithm is NOT proprietary
- Analyzed by public community to show that there are no serious weaknesses
- Explicitly designed for encryption

Kerckhoff's Law (1883)

- The system must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience.

- In other words, the security of the system must rest entirely on the secrecy of the key.

Properties of a Good Cryptosystem

- There should be no way short of enumerating all possible keys to find the key from any amount of ciphertext and plaintext, nor any way to produce plaintext from ciphertext without the key.
- Enumerating all possible keys must be infeasible.
- The ciphertext must be indistinguishable from true random values.

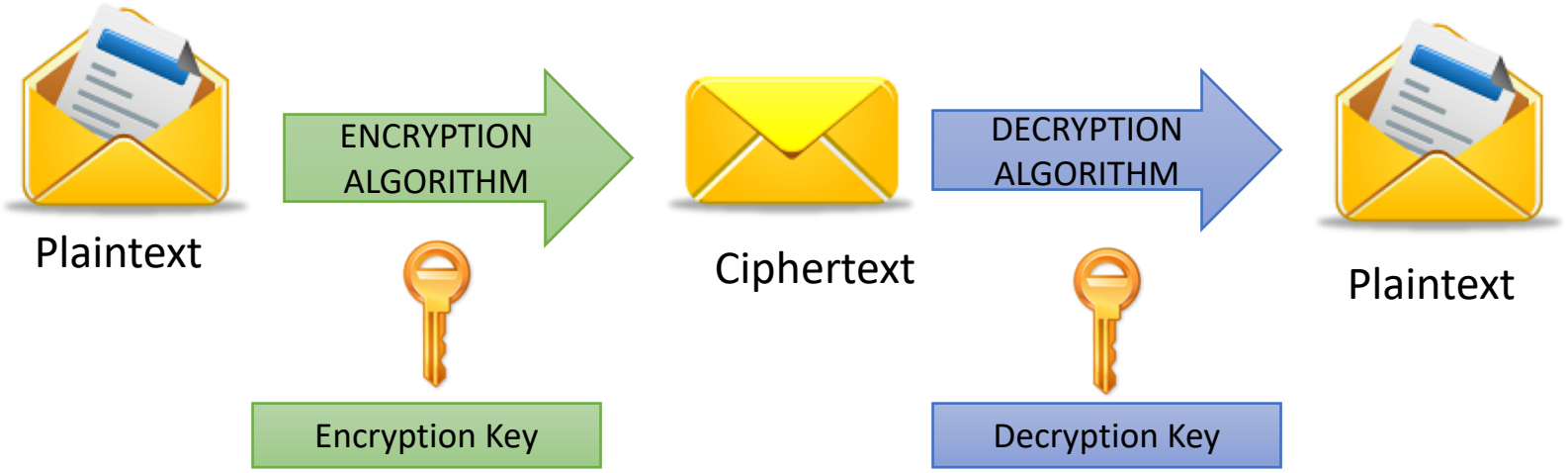
Encryption

- Process of transforming plaintext to ciphertext using a cryptographic key
- Used all around us
  - In Application Layer – used in secure email, database sessions, and messaging
  - In session layer – using Secure Socket Layer (SSL) or Transport Layer Security (TLS)
  - In the Network Layer – using protocols such as IPsec

Encryption

- Benefits of good encryption algorithm:
  - Resistant to cryptographic attack
  - They support variable and long key lengths and scalability
  - They create an avalanche effect
  - No export or import restrictions

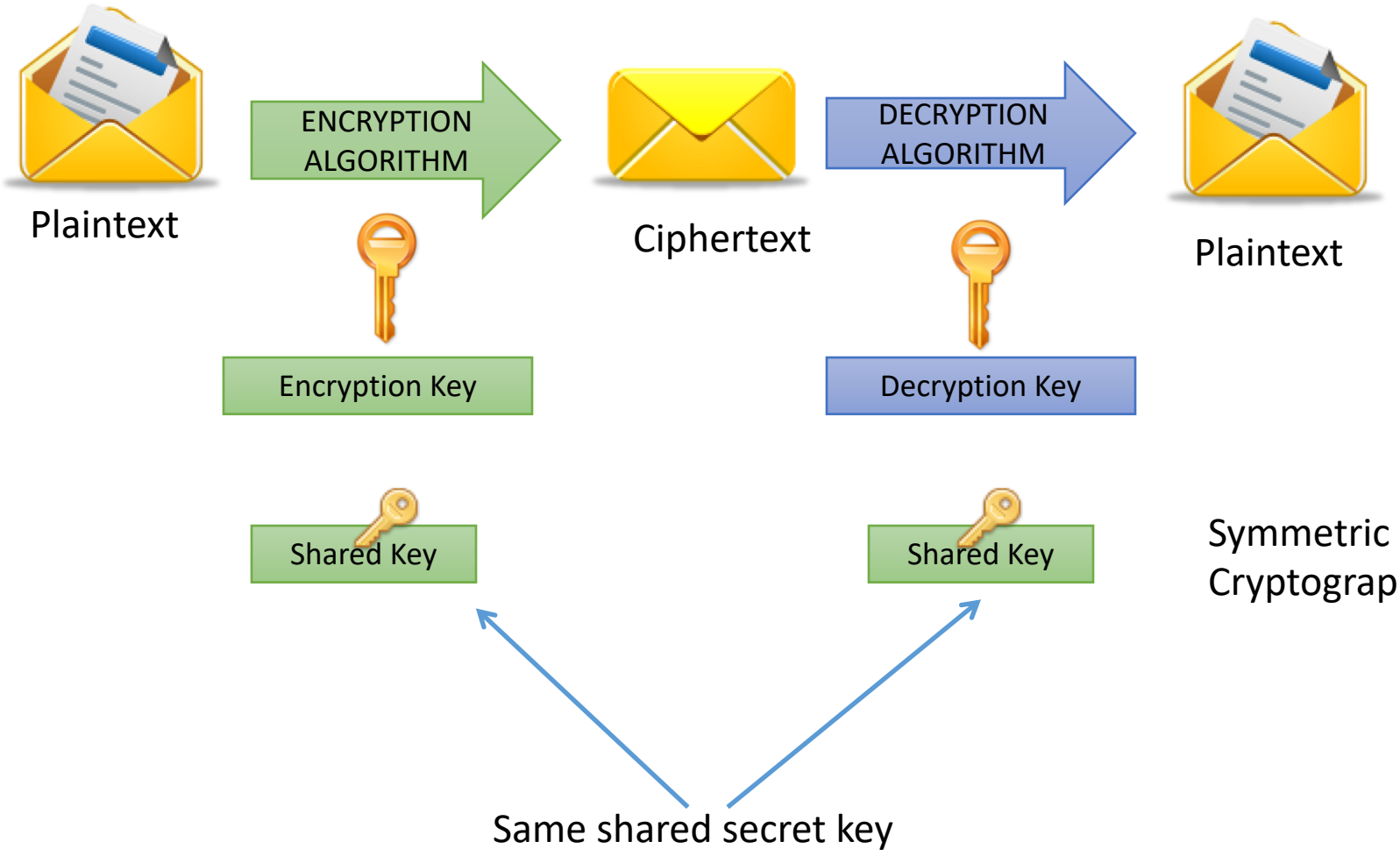
Encryption and Decryption



Symmetric Key Algorithm

- Uses a single key to both encrypt and decrypt information
- Also known as a secret-key algorithm
  - The key must be kept a “secret” to maintain security
  - This key is also known as a private key
- Follows the more traditional form of cryptography with key lengths ranging from 40 to 256 bits.
- Examples:
  - DES, 3DES, AES, RC4, RC6, Blowfish

Symmetric Encryption



Block Cipher

- Transforms a fixed-length block of plain text into a block of ciphertext
  - operate on a pre-determined block of bits (one byte, one word, 512 bytes, so forth), mixing key data in with the message data in a variety of different ways
- Common block ciphers:
  - DES and 3DES (in ECB and CBC mode)
  - Skipjack
  - Blowfish
  - RSA
  - AES
  - IDEA
  - SAFER

Stream Cipher

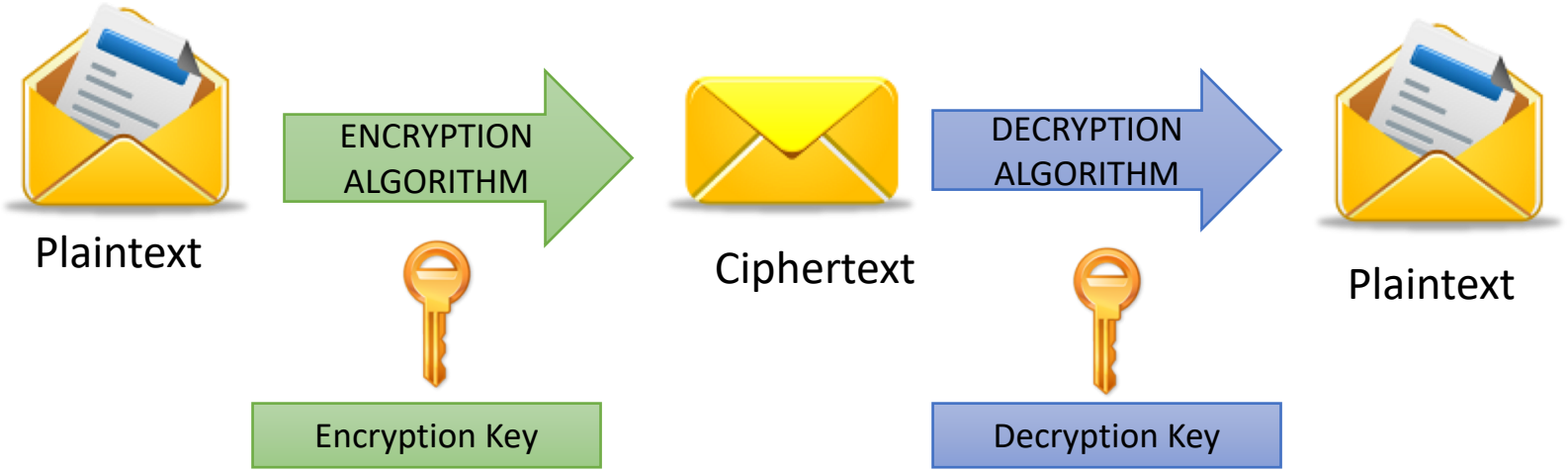
- Use smaller units of plaintext than what are used with block ciphers.
  - encrypts bits of the message at a time
  - typically bit-wise
  - They perform some operation (typically an exclusive OR) with one of these key bits and one of the message bits
- They either have a very long key (that eventually repeats) or a reusable key that generates a repeatable but seemingly random string of bits.
- Common stream ciphers:
  - RC4
  - DES and 3DES (running OFB or CFB mode)
  - SEAL

Data Encryption Standard (DES)

- Developed by IBM for the US government in 1973-1974, and approved in Nov 1976.
- Based on Horst Feistel's Lucifer cipher
- block cipher using shared key encryption, 56-bit key length
- A typical block cipher with block size of 64 bits
- Now considered as insecure

DES: Illustration

64-bit blocks of input text



56-bit keys +
8 bits parity

DES is broken

- Developed in the early 1970s
- The 56-bit key length is too short for modern computation
- Bronek in 22 hours in 1999, but known to be vulnerable earlier
- So it lasted 25 years, OK for early days
- But messages encrypted with DES in 1990 can be read now trivially

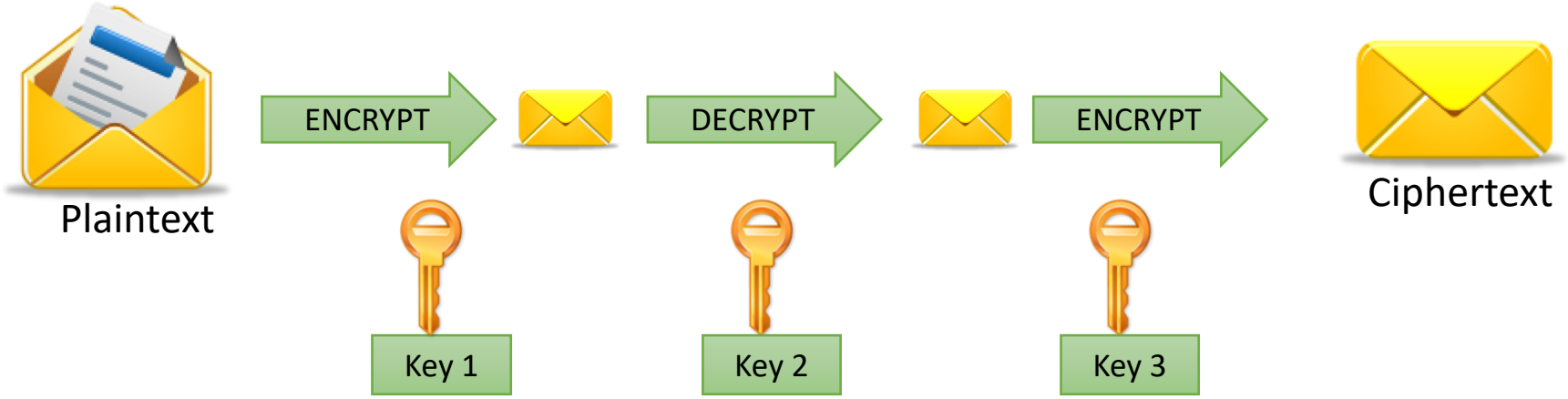
Triple DES

- 3DES (Triple DES) – a block cipher that applies DES three times to each data block
- Uses a key bundle comprising of three DES keys (K1, K2, K3), each with 56 bits excluding parity.
- DES encrypts with K1, decrypts with K2, then encrypts with K3

$$C_i = E_{K3}(D_{K2}(E_{K1}(P_i)))$$

- Disadvantage: very slow

3DES: Illustration



- Note:
  - If Key1 = Key2 = Key3, this is similar to DES
  - Usually, Key1 = Key3

Advanced Encryption Standard (AES)

- Published in November 2001
- Symmetric block cipher
- Has a fixed block size of 128 bits
- Has a key size of 128, 192, or 256 bits
- Based on Rijndael cipher which was developed by Joan Daemen and Vincent Rijmen
- Better suited for high-throughput, low latency environments

Rivest Cipher

- Chosen for speed and variable-key length capabilities
- Designed mostly by Ronald Rivest
- Each of the algorithms have different uses

RC Algorithm	Description
RC2	Variable key-sized cipher used as a drop in replacement for DES
RC4	Variable key sized stream cipher; Often used in file encryption and secure communications (SSL)
RC5	Variable block size and variable key length; uses 64-bit block size; Fast, replacement for DES
RC6	Block cipher based on RC5, meets AES requirement

Symmetric Key Algorithm

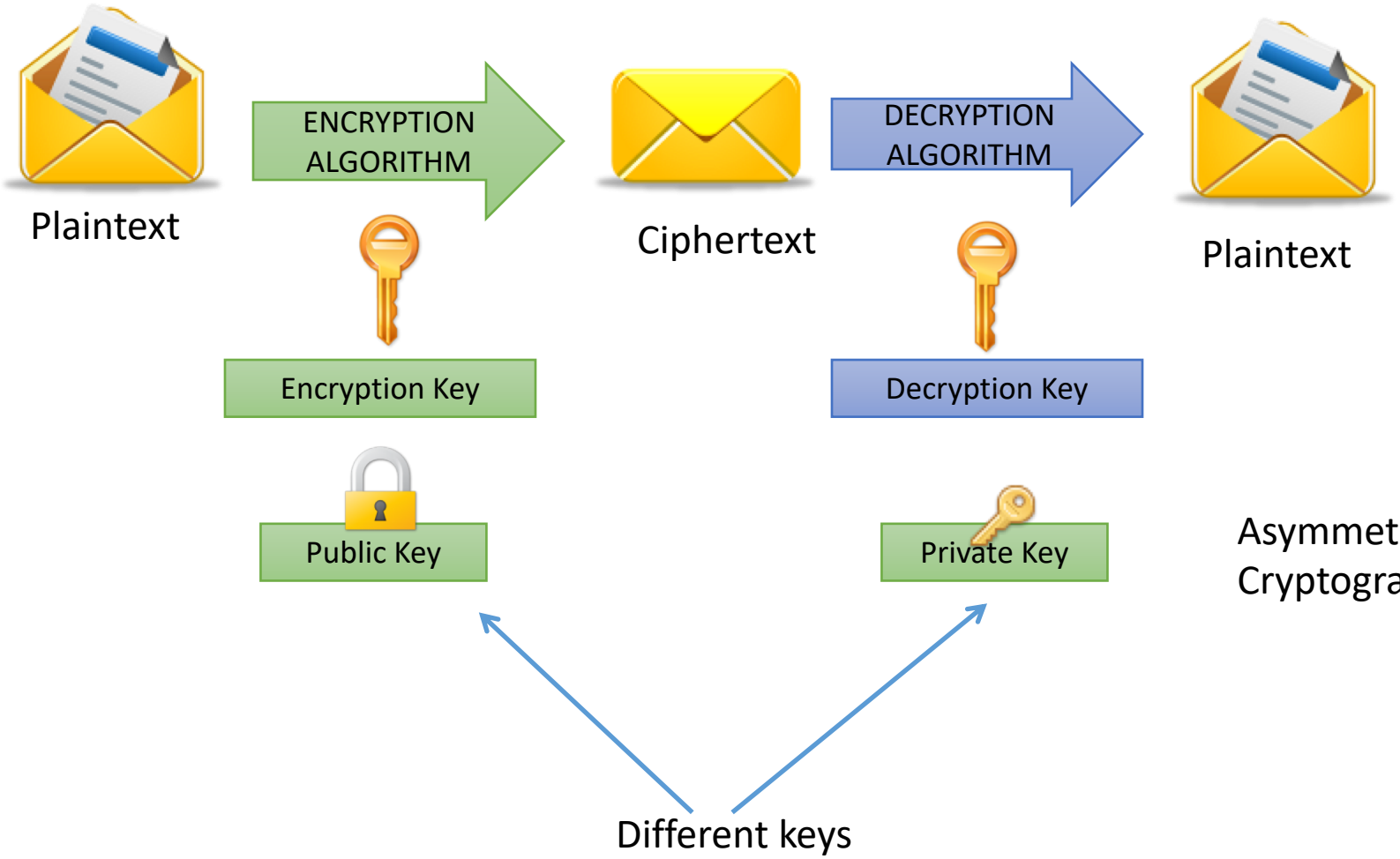
Symmetric Algorithm	Key Size
DES	56-bit keys
Triple DES (3DES)	112-bit and 168-bit keys
AES	128, 192, and 256-bit keys
IDEA	128-bit keys
RC2	40 and 64-bit keys
RC4	1 to 256-bit keys
RC5	0 to 2040-bit keys
RC6	128, 192, and 256-bit keys
Blowfish	32 to 448-bit keys

Note:
Longer keys are more difficult to crack, but more computationally expensive.

Asymmetric Key Algorithm

- Also called public-key cryptography
  - Keep private key to yourself and protected
  - Send/share the public key to anyone
- Separate keys for encryption and decryption (public and private key pairs)
- Examples:
  - RSA, DSA, Diffie-Hellman, ElGamal, PKCS

Asymmetric Encryption



Fun of Asymmetric Key Algorithm

- A data encrypted by a public key is able to decrypt by the corresponding private key
- A data encrypted by a private key is able to decrypt by the corresponding public key

Asymmetric Key Algorithms

- RSA – the first and still most common implementation
- DSA – specified in NIST's Digital Signature Standard (DSS), provides digital signature capability for authentication of messages
- Diffie-Hellman – used for secret key exchange only, and not for authentication or digital signature
- ElGamal – similar to Diffie-Hellman and used for key exchange
- PKCS – set of interoperable standards and guidelines

Symmetric vs. Asymmetric Key

Symmetric	Asymmetric
Generally fast Same key for both encryption and decryption	Can be 1000 times slower Uses two different keys (public and private) Decryption key cannot be calculated from the encryption key Key lengths: 512 to 4096 bits Used in low-volume

Mostly used in combination

Checksum and Hash

Checksum

- Data integrity refers to maintaining and assuring the accuracy and consistency of data over the research lifecycle
- Checksums provide a way to monitor the integrity of your data

Credit Card (check digit example)

- Double every other, and sum every digits
- If the sum is a multiple of 10, the number is valid

- 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 2
- 2 2 6 4 10 6 14 8 18 0 2 2 6 4 10 2
- 2+2+6+4+1+0+6+1+4+8+1+8+0+2+2+6+4+1+0+2
- =60

Data integrity

- What is a hash?
  - like a fingerprint of a file
  - produces a condensed representation of a message (hashing)
  - used to verify whether two copies of a file are identical
- The fixed-length output is called the checksum, hash or message digest
- Each time you run a checksum, a string of digits is created for each file. If just 1 byte of data has been altered, the same process will generate a different string.

Hash Functions

- A hash function takes an input message of arbitrary length and outputs fixed-length code.
  - Given x, we can compute the value f(x).
  - Given f(x), it is hard to get the value of x.
- A form of signature that uniquely represents the data
  - Collision-free
- Uses:
  - Verifying file integrity - if the hash changes, it means the data is either compromised or altered in transit.
  - Digitally signing documents
  - Hashing passwords

Hash Functions

- Message Digest (MD) Algorithm
  - Outputs a 128-bit fingerprint of an arbitrary-length input
  - MD5 is getting obsolete
- Secure Hash Algorithm (SHA)
  - SHA-1 is obsolete
  - Widely-used on security applications (TLS, SSL, PGP, SSH, S/MIME, IPsec)
  - SHA-2 families - SHA-256, SHA-384, SHA-512 are commonly used, which can produce hash values that are 256, 384, and 512-bits respectively
- RIPEMD
  - Derived from MD4, but performs like SHA
  - RIPEMD-160 is the most popular version

Digital Signature

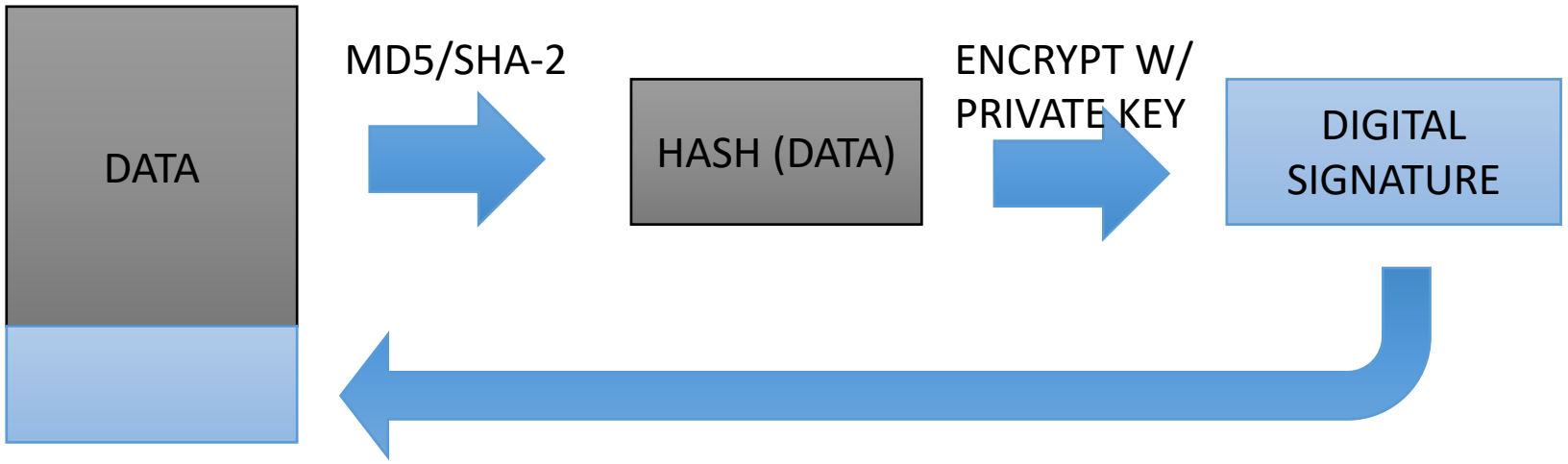
- A digital signature is a message appended to a packet
- The sender encrypts message with own private key instead of encrypting with intended receiver's public key
- The receiver of the packet uses the sender's public key to verify the signature.
- Used to prove the identity of the sender and the integrity of the packet

Digital Signature

- Two common public-key digital signature techniques:
  - RSA (Rivest, Shamir, Adelman)
  - DSS (Digital Signature Standard)
- Used in a lot of things:
  - Email, software distribution, electronic funds transfer, etc
- A common way to implement is to use a hashing algorithm to get the message digest of the data, then use an algorithm to sign the message

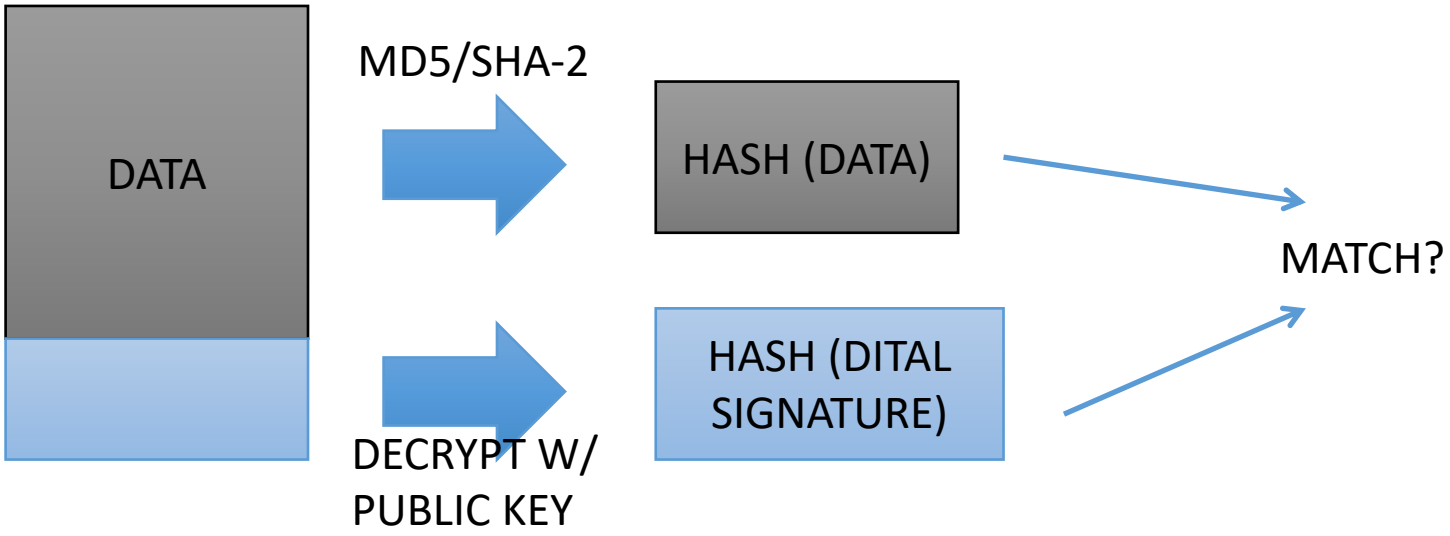
Digital Signature Process

1. Hash the data using one of the supported hashing algorithms (MD5, SHA-256)
2. Encrypt the hashed data using the sender's private key
3. Append the signature (and a copy of the sender's public key) to the end of the data that was signed)



Signature Verification Process

1. Hash the original data using the same hashing algorithm
2. Decrypt the digital signature using the sender's public key. All digital signatures contain a copy of the signer's public key
3. Compare the results of the hashing and the decryption. If the values match then the signature is verified. If the values do not match, then the data or signature was probably modified.



Hash collision

- hash(m1) and hash(m2)
  - hash is a checksum operation
  - m1/m2 are different messages
- hash(original_file) = hash(fake_file)
  - If there is a way to create a file that has the same checksum values as the target file, that's…problem
- Use a modern checksum algorithms to check data integrity
  - like SHA256, SHA512

sha utilities

- on UNIX
  - SHA256: $ sha256sum <file>
  - SHA512: $ sha512sum <file>
- on Mac
  - SHA256: $ shasum -a 256 <file>
  - SHA512: $ shasum -a 512 <file>
- on Windows
  - SHA256: > certutil -hashfile <file> SHA256
  - SHA512: > certutil -hashfile <file> SHA512

Need GUI? (Windows)

- Create a new bat file named ‘sha256.bat’ which contains the following 3 lines

```
@echo off
certutil -hashfile  “%~1” SHA256
pause
```

- Open explorer, type sendto on the address bar
- Put the bat file there
- How to use it
  - Select a file, right click, select sendto then sha256, you will get a SHA256 checksum of the file

md5 utilities (old)

- on UNIX
  - $ md5sum <file>
- on Mac
  - $ md5 <file>
- on Windows (7 or later)
  - > certutil -hashfile <file> MD5

sha1 utilities (old)

- on UNIX
  - $ sha1sum <file>
- on Mac
  - $ shasum <file>
- on Windows
  - > certutil -hashfile <file> SHA1

hands on

Exercise 1: text file

- Create a new text file with the following contents

```
hello!<enter>
```

- Get checksums of the file
  - sha256, sha512

Result should be:

- Mac&Linux:
  - sha256: 69b63d7c7f5e81dcc55ca4be4f45add1757f3b8992c8ac34dd3e80e2f8d814e7
  - sha512: ac3c78e58cda47f8df0bbe4bb1002689d82d8379b1d0636df1cf3dbcca63b73d59835d4c485c055ff94a9022f4cc0d3bea3967b4fc4edd00fa6121bce541455a
- Windows
  - sha256:
  - 1c0b8448b72c55f2f614640252d14ec132fa031d7c5c35d93b6bce780cfe0e92
  - sha512:
  - d91190cce04a17036bb57086c69bd8c377193880d4404d81c940624d6a3251d4623d2c53fc45934de9bef85f073b0acca8bd796fe4bec7a6c6eecfdee2334888

If you are interested in

- made on Windows
  - hello-win.dat
- made on Mac
  - hello-unix.dat

Why? It's "newline" problem

- On Mac&Linux
  - 68 65 6c 6c 6f 21 0a |hello!.| -- <LF>
- On Windows:
  - 68 65 6c 6c 6f 21 0d 0a |hello!..| -- <CR><LF>

- They look like identical, but actually different
- That's why ftp clients have 'text' mode to translate newlines accordingly
  - binary mode just transfer the data as is

Exercise 2: rename the file

- Change the filename to something else
- Get checksums of the files
  - sha256, sha512
- Compare the checksums with the result you get on the Exercise 1

Exercise 3: modify the file

- Delete ’o’ from the file, now the contents should be

```
hell!<enter>
```

- Get checksums of the files
  - sha256, sha512
- Compare the checksums with the result you get on the Exercise 1

Exercise 4: re-modify the file

- Add ’o’ to the file, now the contents should be

```
hello!<enter>
```

- Get checksums of the files
  - sha256, sha512
- Compare the checksums with the result you get on the Exercise 1

Exercise 5: downloaded file

- Get checksums of this pdf file
  - sha256, sha512
- Compare the checksums with your neighbors