北京大学 PEKING UNIVERSITY

瀚高 HIGH GO

数据加密理论及基础

PG开源开发实践-第三周

David & Cary

CONTENT
目 录

What Is Security?

What Is Security?

安全是什么？

- The word “**Security**” is a very broad concept and could refer to completely different procedures and methodology to achieve
- Safeguarding the company data is of utmost importance!
- Data compromises could lead to financial loss, reputation damage, consumer confidence disintegration, brand erosion, and non-compliance of government and industry regulation.
- Security on infrastructure software is even more important because any data compromises could have nation or city-wide impacts
- These damages are often very difficult to completely recover

What Is Security?

Common Database Compromises

User Compromise:

- Excessive privileges
- Privilege abuse
- Weak user authentication
- Weak password
- Default privilege too open

Data Compromise:

- Unmanaged and unprotected sensitive data
- Backup data exposure
- Stolen hard disks
- Unmanaged encryption keys

Network Compromise:

- Firewall rules
- Deep Packet Inspection (DPS)
- Vulnerability prevention
- Denial of Service (DOS) attack

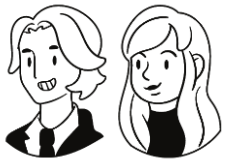
Vulnerability:

- Software bug
- Buffer overflow
- SQL injection
- Privileged escalation

What Is Security?

The Big Picture

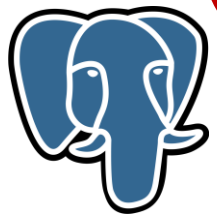
TDE happens here!



PostgreSQL
Users (Client)

Data Security
- TLS / SSH

Firewall



PostgreSQL
Server

Vulnerability
- Software bug

Data Security
- Data encryption

Hard disk

Authenticatio
Server

User Security
- password check
- User authentication
- Privilege checking

Bad Guy

Network Security
- Firewall Rules
- DOS attack

Privilege Settings

Where is security involved?

What Is Security?

Kerckhoffs's principle

- The system must be practically, if not mathematically, indecipherable;
- It should not require secrecy, and it should not be a problem if it falls into enemy hands;
- It must be possible to communicate and remember the key without using written notes, and correspondents must be able to change or modify it at will;
- It must be applicable to telegraph communications;
- It must be portable, and should not require several persons to handle or operate;
- Lastly, given the circumstances in which it is to be used, the system must be easy to use and should not be stressful to use or require its users to know and comply with a long list of rules.

What Is Security?

安全学什么？

- Cryptology
  - ✓ Cryptography
  - ✓ Cryptanalysis

Cryptography

Cryptography

Introduction

- Secret-key Cryptosystems
  - ✓ Plaintext （明文）
  - ✓ Ciphertext （密文）
  - ✓ Encryption （加密）
  - ✓ Decryption （解密）
  - ✓ Key （秘钥）
  - ✓ Alice and Bob, and Eve
- Public-key Cryptosystems
  - ✓ Public key
  - ✓ Private key
  - ✓ Rivest, Shamir and Adleman
- Block and Stream Ciphers
  - ✓ DES, 3DES, AES, RC4,
- Hybrid Cryptography

Cryptography

Introduction

- Message Integrity
  - ✓ Message Authentication Codes (MACs)
  - ✓ Signature schemes
  - ✓ Nonrepudiation
  - ✓ Certificates
  - ✓ Hash Functions

- Cryptographic Protocols
  - ✓ Identification scheme
  - ✓ Key distribution scheme
  - ✓ Secret sharing scheme

Cryptography

Introduction

- Three levels of security
  - ✓ computational security
  - ✓ provable security
  - ✓ unconditional security
- Four commonly considered attack models
  - ✓ known ciphertext attack
  - ✓ known plaintext attack
  - ✓ chosen plaintext attack
  - ✓ chosen ciphertext attack
- Various kinds of attacks against implementations of cryptography
  - ✓ padding oracle attack
  - ✓ side channel attacks
  - ✓ timing attacks
  - ✓ fault analysis attacks
  - ✓ power analysis attacks
  - ✓ cache attacks

Cryptography

Classical Cryptography

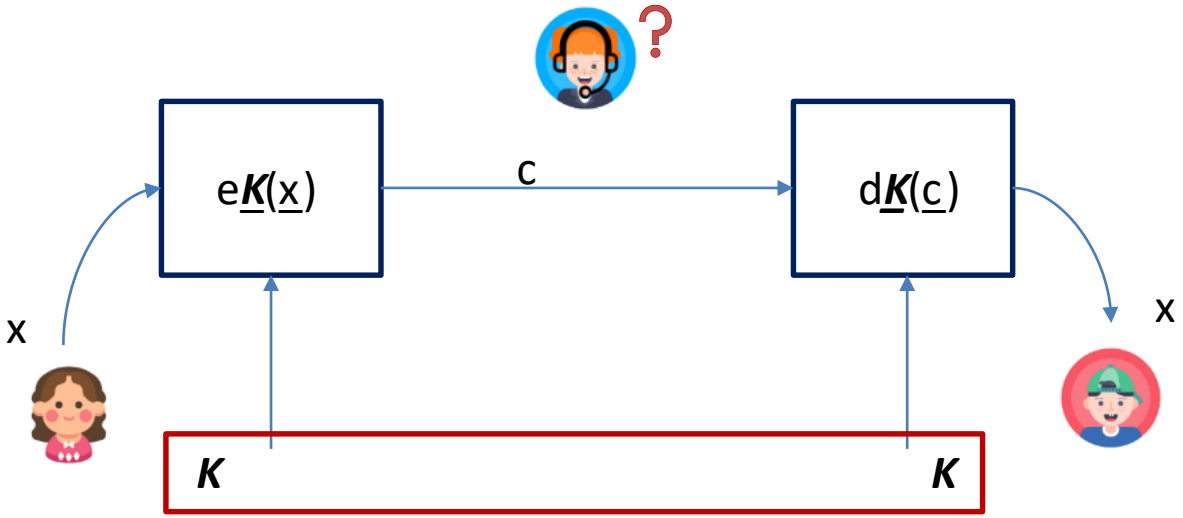
- a simple cryptosystem

A **cryptosystem** is a five-tuple (P, C, K, E, D), where the following conditions are satisfied:

1. P is a finite set of possible **plaintexts**;
2. C is a finite set of possible **ciphertexts**;
3. K, the **key space,** is a finite set of possible **keys**;
4. For each $K \in$ K, there is an **encryption rule** $eK \in$ E and a corresponding **decryption rule** $dK \in$D. Each eK: P→C and dK: C→P are functions such that $dK(eK(x)) = x$ for every plaintext element $x \in$ P.

Cryptography

Classical Cryptography



Cryptography

Classical Cryptography

- The Shift Cipher

the shift cipher, also known as Caesar's cipher, is one of the simplest and most widely known encryption techniques. It is a type of substitution cipher in which each letter in the plaintext is replaced by a letter some fixed number of positions down the alphabet.

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Suppose the key for a Shift Cipher is $K = 11$, and the plaintext is

wewillmeetatmidnight

22	4	22	8	11	11	12	4	4	19
0	19	12	8	3	13	8	6	7	19

7	15	7	19	22	22	23	15	15	4
11	4	23	19	14	24	19	17	18	4

Cryptography

Classical Cryptography

- The Substitution Cipher

a substitution cipher is a method of encrypting in which units of plaintext are replaced with the ciphertext, in a defined manner, with the help of a key; the "units" may be single letters (the most common), pairs of letters, triplets of letters, mixtures of the above, and so forth. The receiver deciphers the text by performing the inverse substitution process to extract the original message.

plaintext

a	b	c	d	e	f	g	h	i	j	k	l	m
X	N	Y	A	H	P	O	G	Z	Q	W	B	T

n	o	p	q	r	s	t	u	v	w	x	y	z
S	F	L	R	C	V	M	U	E	K	J	D	I

ciphertext

A	B	C	D	E	F	G	H	I	J	K	L	M
d	l	r	y	v	o	h	e	z	x	w	p	t

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
b	g	f	j	q	n	m	u	s	k	a	c	i

MGZVYZLGHCMHJMYXSSFMNHAHYCDLMHA ?

Cryptography

Shannon’ s Theory

- 1949 - Shannon published “Communications Theory of Secrecy Systems”
- Observation that secrecy problem is analogous to “noisy-channel” problem
- Perfect secrecy
- Entropy

…

A necessary and sufficient condition for perfect secrecy can be found as follows: We have by Bayes’ theorem

$$P_E(M) = \frac{P(M)P_M(E)}{P(E)}$$

in which:

$P(M)$ = *a priori* probability of message M.
$P_M(E)$ = conditional probability of cryptogram E if message M is chosen i.e. the sum of the probabilities of all keys which produce cryptogram E from message M.
$P(E)$ = probability of obtaining cryptogram E from any cause.
$P_E(M)$ = *a posteriori* probability of message M if cryptogram E is intercepted.

For perfect secrecy $P_E(M)$ must equal $P(M)$ for all E and all M. Hence either $P(M) = 0$, a solution that must be excluded since we demand the equality independent of the values of $P(M)$, or

$$P_M(E) = P(E)$$

for every M and E. Conversely if $P_M(E) = P(E)$ then

$$P_E(M) = P(M)$$

and we have perfect secrecy. Thus we have the result:

Cryptography

Steam Cipher

- A stream cipher
  - ✓ is a **symmetric key** cipher where plaintext digits are combined with a pseudorandom cipher digit stream (keystream). In a stream cipher, each plaintext digit is encrypted one at a time with the corresponding digit of the keystream, to give a digit of the ciphertext stream. In practice, a digit is typically a bit and the combining operation is an exclusive-or (XOR).
  - ✓ For a stream cipher to be secure, its keystream must have a **large period**, and it must be impossible to recover the cipher's key or internal state from the keystream.
- Usages

  Stream ciphers are often used for their speed and simplicity of implementation in hardware, and in applications where plaintext comes in quantities of unknowable length like a secure wireless connection.
- Notable stream ciphers
  - ✓ ChaCha is becoming the most widely used stream cipher in software;
  - ✓ others include: RC4, A5/1, A5/2, Chameleon, FISH, Helix, ISAAC, MUGI, Panama, Phelix, Pike, Salsa20, SEAL, SOBER, SOBER-128, and WAKE.

https://en.wikipedia.org/wiki/Stream_cipher

Cryptography

Block Cipher

- A Block Cipher
  - ✓ the plaintext is divided into fixed-sized chunks called blocks. A block is specified to be a bitstring of some fixed length (e.g., 64 or 128 bits). A block cipher will encrypt (or decrypt) one block at a time.
  - ✓ Most modern-day block ciphers incorporate a sequence of permutation and substitution operations. A commonly used design is that of an iterated cipher. An iterated cipher requires the specification of a round function and a key schedule, and the encryption of a plaintext will proceed through N similar rounds.
- Usages
- Notable block ciphers:
  DES, IDEA, RC5, AES,

Cryptography

Steam Cipher vs. Block Cipher

- One byte of plaintext at a time
- uses 8 bits
- Complex
- uses only **confusion**
- reverse encrypted text is easy
- modes: CFB (Cipher Feedback) and OFB (Output Feedback)
- Fast

- One block of plaintext at a time
- 64 bits, 128 bits or more
- Simple
- Uses **confusion** as well as **diffusion**
- reverse encrypted text is hard
- modes: ECB (Electronic Code Book) and CBC (Cipher Block Chaining)
- Slow

Cryptography

Hash Functions

- a mathematical algorithm that maps data of arbitrary size (often called the "message") to a bit array of a fixed size (the "hash value", "hash", or "message digest"). It is a one-way function, that is, a function which is practically infeasible to invert.
- **SHA-0**: A retronym applied to the original version of the 160-bit hash function published in 1993 under the name "SHA". It was withdrawn shortly after publication due to an undisclosed "significant flaw" and replaced by the slightly revised version SHA-1.
- **SHA-1**: A 160-bit hash function which resembles the earlier MD5 algorithm. This was designed by the National Security Agency (NSA) to be part of the Digital Signature Algorithm. Cryptographic weaknesses were discovered in SHA-1, and the standard was no longer approved for most cryptographic uses after 2010.
- **SHA-2**: A family of two similar hash functions, with different block sizes, known as SHA-256 and SHA-512. They differ in the word size; SHA-256 uses 32-byte words where SHA-512 uses 64-byte words. There are also truncated versions of each standard, known as SHA-224, SHA-384, SHA-512/224 and SHA-512/256. These were also designed by the NSA.
- **SHA-3**: A hash function, chosen in 2012 after a public competition among non-NSA designers. It supports the same hash lengths as SHA-2, and its internal structure differs significantly from the rest of the SHA family.
- The corresponding standards are **FIPS PUB 180** (original SHA), **FIPS PUB 180-1** (SHA-1), **FIPS PUB 180-2** (SHA-1, SHA-256, SHA-384, and SHA-512). NIST has updated Draft FIPS Publication 202, SHA-3 Standard separate from the Secure Hash Standard (SHS).

Cryptography

Hash Functions

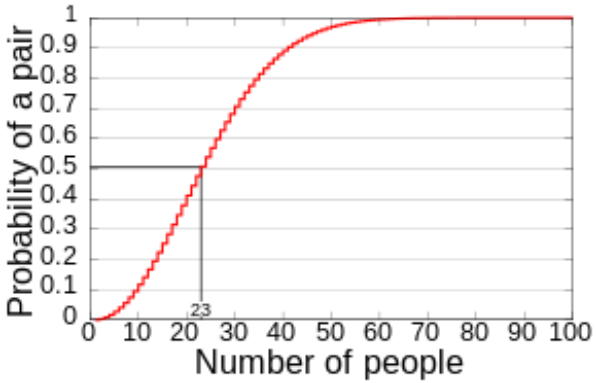
- Birthday Paradox
  The probability of having a "collision" when items are selected among m possible outcomes is:

$$P(m,n) = 1 - \frac{m^{(n)}}{m^n}$$

where

$$m^{(n)} = m(m-1)(m-2)....(m-n+1)$$

- the number of people you need in a room to find two people with the same birthday - it only requires 23 people to get better than 50% (0.507) probability of a match
- P(365, 30) ~ 71%
- P(365, 70) ~ 99.9%



https://en.wikipedia.org/wiki/Birthday_problem

Cryptography

Message Authentication

- a message authentication code (MAC), sometimes known as a tag, is a short piece of information used to authenticate a message, in other words, to confirm that the message came from the stated sender (its authenticity) and has not been changed.
- The MAC value protects a message's data integrity, as well as its authenticity, by allowing verifiers (who also possess the secret key) to detect any changes to the message content.
- The term message integrity code (MIC) is frequently substituted for the term MAC, especially in communications to distinguish it from the use of the latter as media access control address (MAC address).
  - ✓ Various standards FIPS PUB 113 Computer Data Authentication, withdrawn in 2002, defines an algorithm based on DES.
  - ✓ FIPS PUB 198-1 The Keyed-Hash Message Authentication Code (HMAC)
  - ✓ ISO/IEC 9797-1 Mechanisms using a block cipher
  - ✓ ISO/IEC 9797-2 Mechanisms using a dedicated hash-function
  - ✓ ISO/IEC 9797-3 Mechanisms using a universal hash-function
  - ✓ ISO/IEC 29192-6 Lightweight cryptography - Message authentication codes

Cryptography

Reference Material

- https://crypto.stanford.edu/~dabo/cryptobook/BonehShoup_0_5.pdf
- http://web.engr.oregonstate.edu/~rosulekm/crypto/
- http://netlab.cs.ucla.edu/wiki/files/shannon1949.pdf
- Introduction to Modern Cryptography (3rd edition)
  http://www.cs.umd.edu/~jkatz/imc.html
- a free online course “Cryptography I” by Stanford University
  https://www.coursera.org/learn/crypto/home/info

Data Security Principles

Data Security Principles

数据安全原则

The basic principles and methodologies involved in data security include but not limited to:

- Symmetrical Encryption
- Asymmetrical Encryption (a.k.a Public Key Cryptography)
- Block Cipher Mode of Operation (a.k.a Stream Cipher)
- Key Exchange Algorithms
- Data Integrity Check / Data Authentication

We will go over each of these principles next…

Basic Terminology

基本词汇

- Plaintext: the original message
- Ciphertext: the coded message
- Cipher: algorithm for transforming plaintext to ciphertext
- Key: secret info used in cipher known only to sender/receiver
- encipher (encrypt): converting plaintext to ciphertext
- decipher (decrypt): recovering ciphertext from plaintext

Symmetrical Encryption

对称加密

- Symmetrical Encryption is a type of encryption where only one secret key is used to encrypt and decrypt a message.
- The connecting client will use this secret key to encrypt the message and send the encrypted messages to server
- The server will use the same secret key to decrypt the encrypted messages and vice versa
- Most frequently used encryption method to secure your data

Common Algorithms:

- AES-128, AES-192, AES-256
- Blowfish
- DES

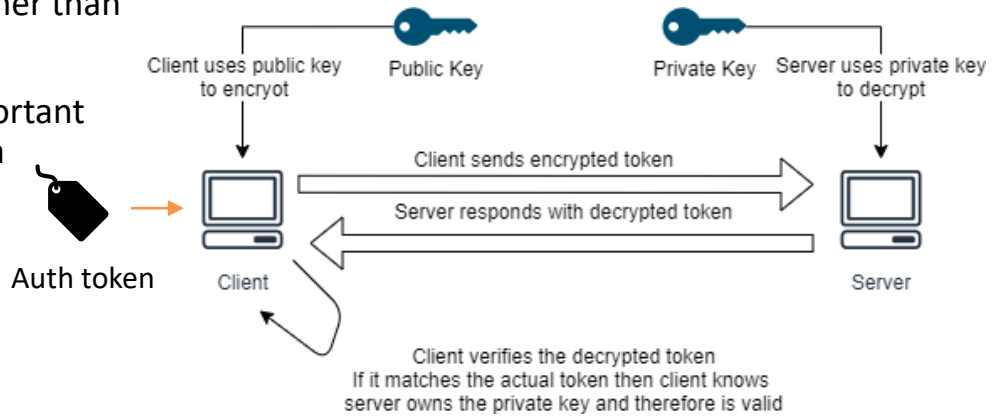
Asymmetrical Encryption

非对称加密

- Asymmetrical encryption uses **two distinct keys** called **public** and **private** keys; Public key is used for encryption and private key is used for decryption.
- Both keys are different but related by math and it is much **slower** than symmetrical encryptions.
- Public key can be distributed publicly while private key is to be kept private.
- This essentially forms a secured **one-way communication** and is normally used **for mutual authentication** rather than data protection.
- In addition to data encryption, trust is another important aspect in security. Transport Layer Security (TLS) is a protocol that employs both.

Common Algorithms:

- RSA (1024, 2048, 4096bit)
- Elliptic Curve
- DSA



Block Cipher Mode Of Operation

分组密码模式

- Block Cipher Mode of Operation is normally used with **symmetrical encryption** to encrypt or decrypt a stream of data block by block.
- There are several available modes of block cipher operations that have different strengths and weaknesses.
- Most modes require a complete block of 16 bytes to be able to encrypt. In the case where the input stream is not in multiple of 16, **padding** of 0s are normally appended to fill the blocks

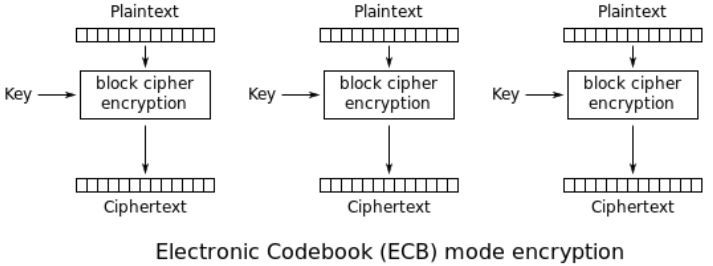
Common Algorithms:

- Cipher Block Chaining (CBC)
- Counter (CTR)
- Electronic Codebook (ECB)
- … and many more

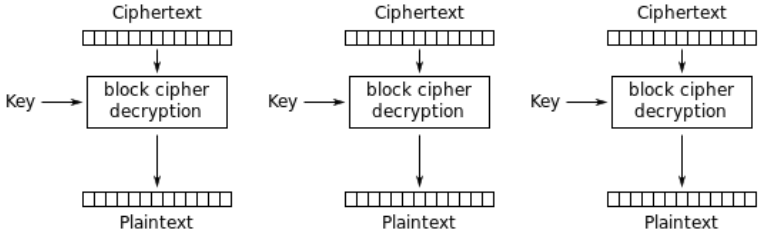
Block Cipher Mode Of Operation - ECB

ECB 分组密码模式

- ECB stands for **Electronic Code Book**
- One of the simplest and least secured block cipher modes
- Lacks randomness and therefore the encrypted results (especially image data) could still see patterns of original data.
- Not recommended by many software applications today



Electronic Codebook (ECB) mode encryption



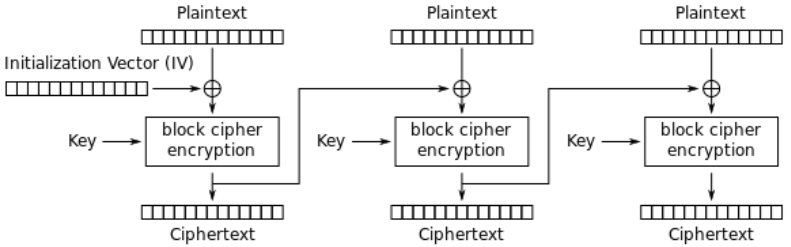
Electronic Codebook (ECB) mode decryption

Images source: https://en.wikipedia.org/wiki/Block_cipher_mode_of_operation

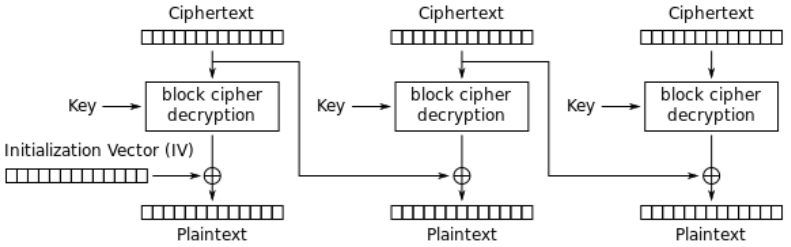
Block Cipher Mode Of Operation – CBC

ECB 分组密码模式

- CBC stands for **Cipher Block Chaining**
- Simple and more secured than EBC
- Provides considerable randomness to the encrypted results guaranteed by Initialization Vector (IV). The encrypted current block is used as the IV for the next until the end
- Require the same IV for decryption
- Management of IV is important
- Still a popular block cipher mode today



Cipher Block Chaining (CBC) mode encryption



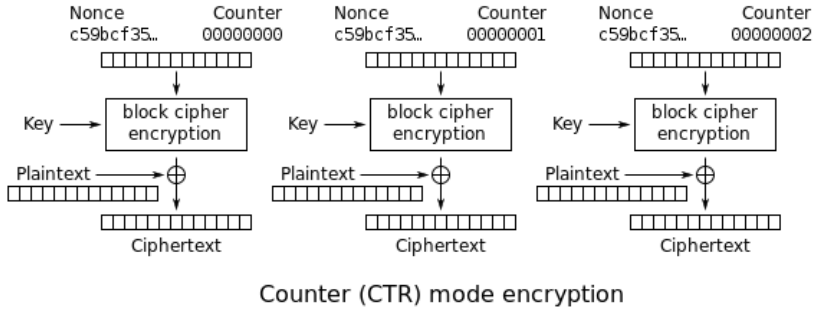
Cipher Block Chaining (CBC) mode decryption

Images source: https://en.wikipedia.org/wiki/Block_cipher_mode_of_operation

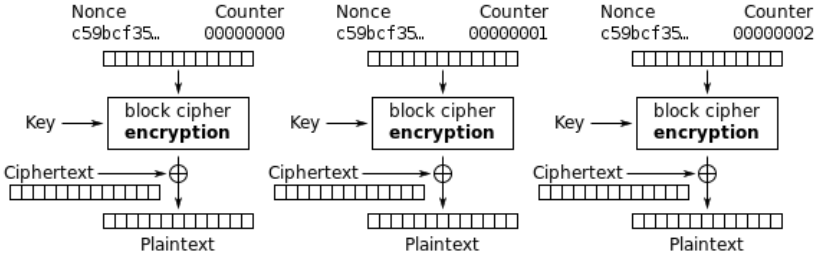
Block Cipher Mode Of Operation - CTR

CTR 分组密码模式

- CTR stands for **Counter**
- More secured than CBC
- Provides high degree of randomless guaranteed by the Counter values fed into each block
- Each encrypted block should be fed with different counter value
- Ideal for encrypting a stream of data having IDs or sequence numbers such as WAL records, buffer entries in PostgreSQL because these can generally be used as counter value
- Popular choice of block cipher today.



Counter (CTR) mode encryption



Counter (CTR) mode decryption

Images source: https://en.wikipedia.org/wiki/Block_cipher_mode_of_operation

Key Exchange Algorithm

密钥交换算法

- Key exchange algorithm is a math algorithm designed to make both client and server agree on a secret key without sending the key to each other.
- This is done by pure math equations and require several steps of intermediate token exchange.
- In the end, both client and server will compute to the same value, which can be used as the secret key for **symmetrical encryption algorithms**
- Both **SSH** and **TLS** uses a key exchange algorithm during the **handshake** stage to determine a secret key for the data encryption for that session.

Common key exchange algorithms

- Diffie-Hellman (DH)
- Elliptic Curve Diffie-Hellman (ECDH)
- Ephemeral Diffie-Hellman (DHE)

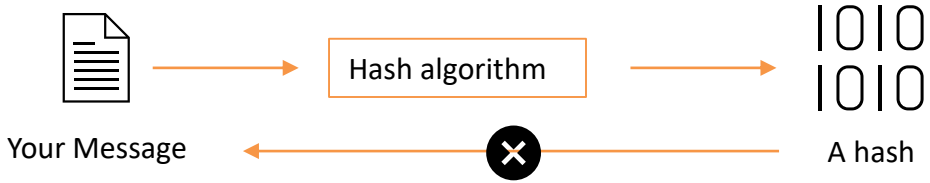
Data Integrity Check

数据完整性检查

- Data integrity check refers to the methods to ensure that the data stream has been received without being tempered with during transmission. Think of it as a data checksum or an authentication token
- This process is normally done by Message Authentication Code (MAC) or Hash-based MAC (HMAC) algorithms
- Ensuring data integrity is very important security measure to avoid man-in-the-middle attack.
- HMAC is a more secured than MAC for computing the authentication token as it requires an additional cryptographic key to perform the task.
- Both HMAC and MAC involve **hashing**, which turns a message of various lengths into a fixed size output (a.k.a Message Digest (MD))but not the other way around and the same input always produces the same output.

Common HMAC algorithms

- HMAC-SHA256
- HMAC-SHA512
- HMAC-SHA1
- HMAC-MD5



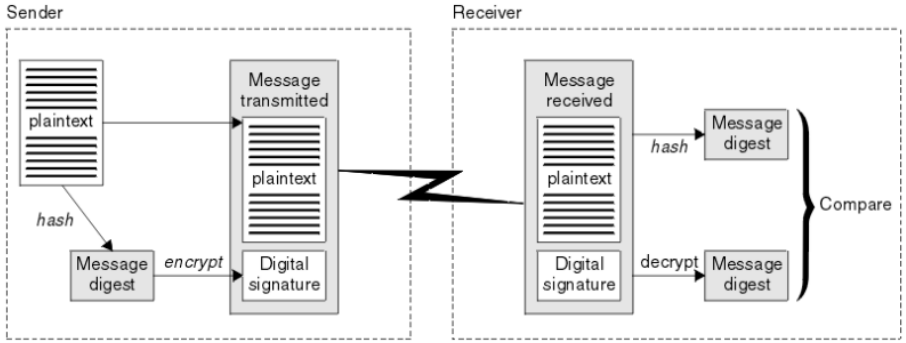
Digital Signature

电子签名

- A digital signature is formed by encrypting a representation of a message and for efficiency, usually operates on a message digest rather than the message itself.
- So, computing a digital signature involves both hashing and asymmetrical encryption
- Has 3 guarantees:
  - Authentication: ensure the message is received by the trusted entity
  - Integrity: ensure the message is not altered
  - Non-repudiation: ensure that the entity signing the message cannot deny having signed it later

Common digital signature algorithms

- RSA with SHA
- ECDSA with SHA



Transport Layer Security (TLS)

TLS介绍

- Used to be called Secured Socket Layer (SSL)
- Transport Layer Security (TLS) is one of the most widely used communication protocol today
- It is designed based on all the security principles discussed so far.
- Uses X.509 certificate plus digital signature to ensure authenticity, integrity and trust between client and server
- Uses key exchange algorithms to exchange a session encryption key
- Uses symmetrical encryption to encrypt the data.
- **https** is one of the most popular protocols using TLS.
- Since TLS uses all of the principles to communicate, it is important to ensure both client and server can support the algorithms required
- This is where cipher suites come into play

Example of TLS cipher suite:

TLS_DHE_RSA_WITH_AES_256_CBC_SHA256

- **DHE**:

use Ephemeral Diffie-Hellman key exchange algorithm

- **RSA**:

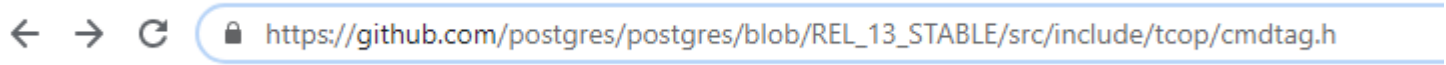
use RSA asymmetrical keys for authentication

- **AES_256_CBC** :

use AES-256 symmetrical encryption with CBC block cipher mode

- **SHA256** :

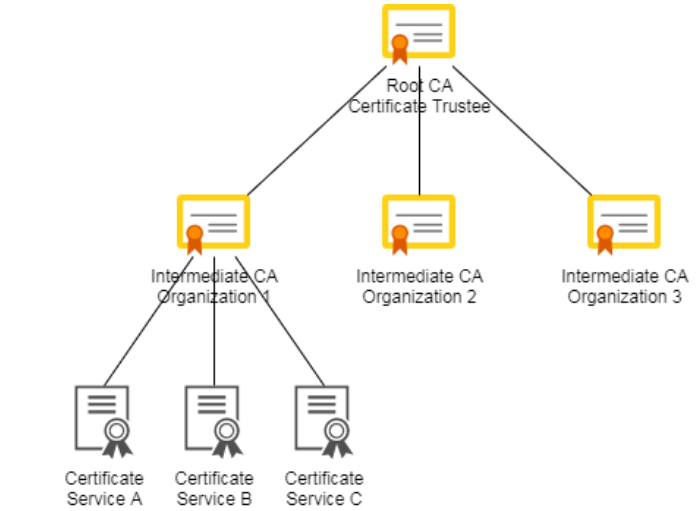
use SHA-256 as message authentication algorithm to compute digital signature

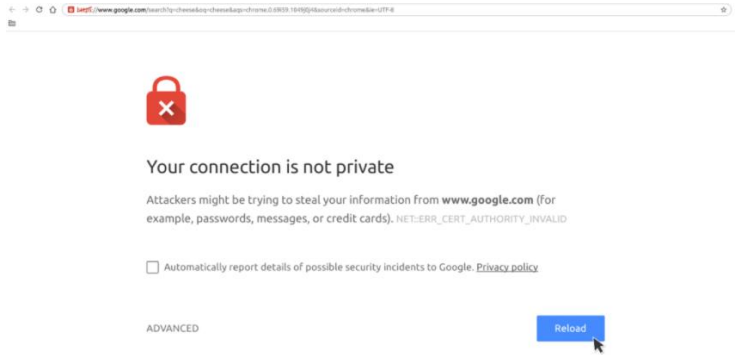


X509 Certificate Chain of Trust

TLS介绍

- The certificate created at the top hierarchy is called a “**Root CA**” (Root **Certificate Authority**) and is normally created by a trusted organization.
- This “root CA” is able to sign additional “Intermediate CA” that can be distributed to other organizations.
- The intermediate CA can then be used to create and sign individual entity certificates to be used by services like HTTPS, FTPS...etc.
- Client and server can **optionally** perform a **certificate validation** against its CA Certificate to see if the given certificate can establish the chain of trust back to the Root CA. This is not always required as it depends on your security policy settings.
- It is **possible** for a client to communicate with a server having an expired or invalid certificate.
- In the case with HTTPS, the browser will give user a popular warning “**Your connection is not private**” and it is up to you if you would like to proceed or not.





HIGH GO

融知与行 瀚且高远

THANKS