INDUSTRIAL UNIVERSITY OF HOCHIMINH CITY

Chương 6:
HÀM BĂM – CHỮ KÝ SỐ

(Hash Function - DIGITAL SIGNATURES)

INDUSTRIAL UNIVERSITY OF HOCHIMINH CITY

HÀM BĂM
(Hash Function)

Nhu cầu toàn vẹn thông tin

- Các ứng dụng chú trọng mục tiêu Toàn vẹn
  - Tài liệu được sử dụng giống hệt tài liệu lưu trữ
  - Các thông điệp trao đổi trong một hệ thống an toàn không bị thay đổi/sửa chữa
- "Niêm phong" tài liệu/thông điệp
  - "Niêm phong" không bị sửa đổi/phá hủy đồng nghĩa với tài liệu/thông điệp toàn vẹn
  - "Niêm phong": băm (hash), tóm lược (message digest), đặt số kiểm tra (checksum)
  - Tạo ra "niêm phong": hàm băm

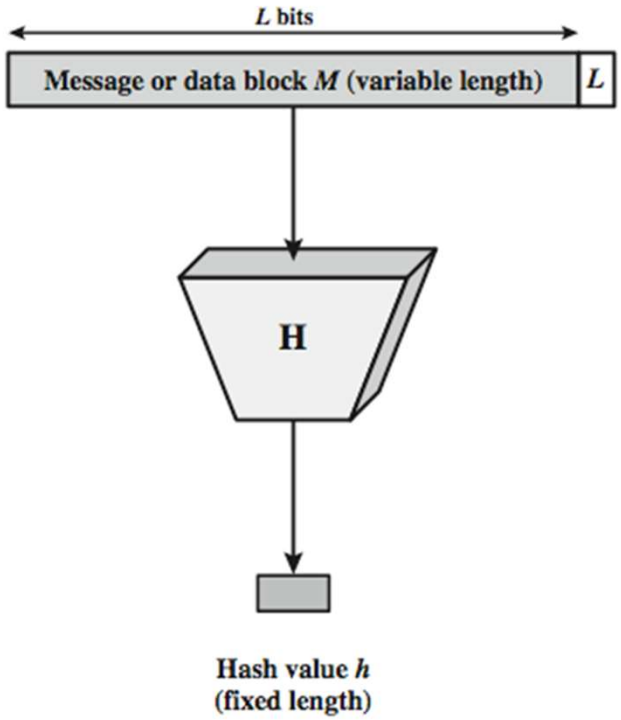
Hàm băm

- Mục tiêu an toàn
  - Toàn vẹn (Integrity)

1. Hàm băm (Hash Function)

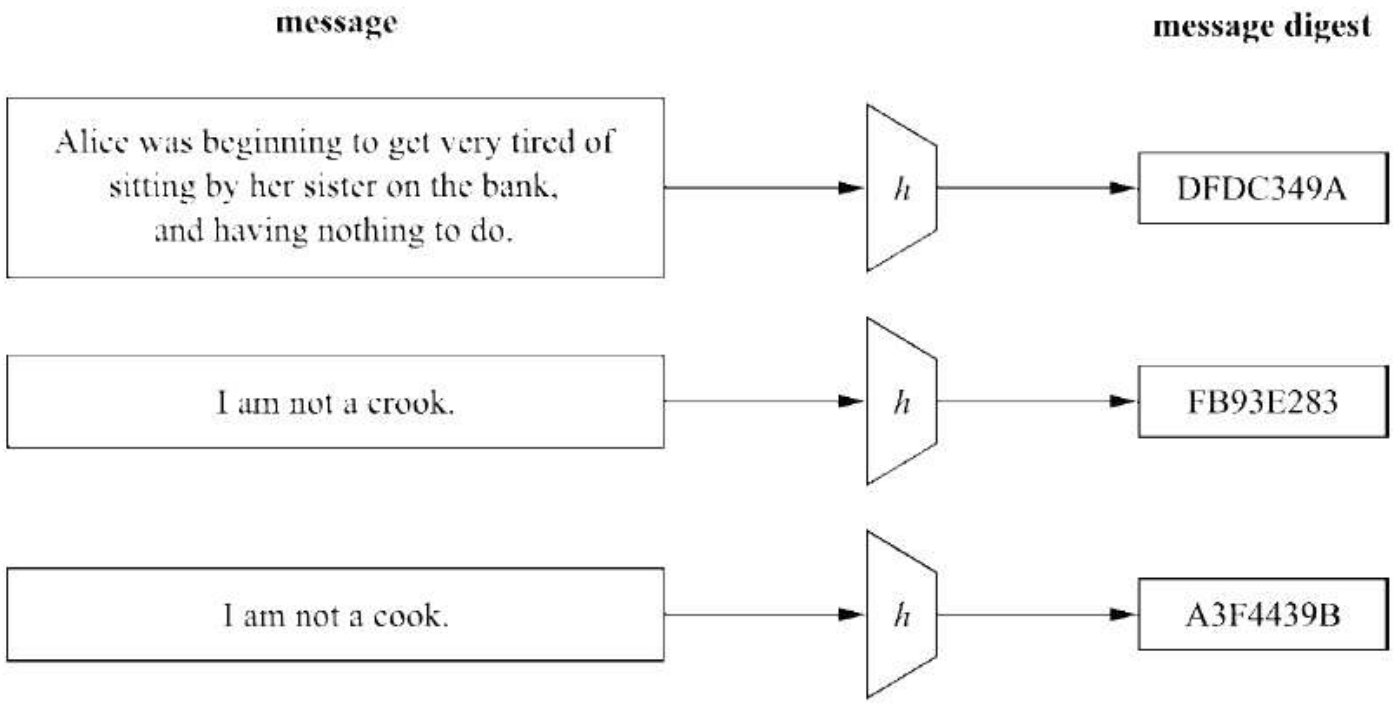
- **Hàm băm** là các thuật toán không sử dụng khóa để mã hóa, nó có nhiệm vụ băm thông điệp được đưa vào theo một thuật toán ***h*** **một chiều nào đó**, rồi đưa ra một **bản băm – văn bản đại diện – có kích thước cố định**. Do đó người nhận không biết được nội dung hay độ dài ban đầu của thông điệp đã được băm bằng hàm băm.
- Giá trị của hàm băm là **duy nhất**, và **không thể suy ngược** lại được nội dung thông điệp từ giá trị băm này.

1. Hàm băm (Hash Function)

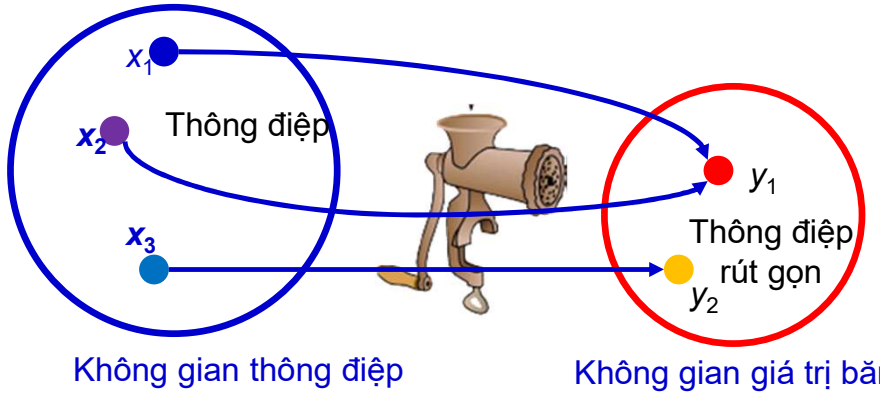


- Input: ***M có kích thước bất kỳ***
- Output – ***giá trị h có kích thước cố định, ngắn.***
- **H(x)** – ***hàm một chiều ("Khó để tính nghịch đảo")***

1. Hàm băm (Hash Function)



1. Hàm băm (Hash Function)



→Không gian giá trị Băm nhỏ hơn rất nhiều so với Không gian thông điệp về mặt kích thước →chắc chắn sẽ tồn tại đụng độ (trùng), nghĩa là **có hai tin *x* và *x"* mà giá trị Băm của chúng là giống nhau, tức là *h(x) = h(x")***

Tính chất hàm băm

1. ***Tính chống tiền ảnh*** (*Preimage resistant – one-way property*):

   Cho trước giá trị băm h việc tìm x sao cho H(x)=h là rất khó

2. ***Tính chống tiền ảnh thứ hai*** *(Second preimage resistant – weak collision resistant – Tính chống trùng yếu):*

   Cho thông điệp đầu vào x, việc tìm một thông điệp x' với (x'≠x) sao cho h(x')=h(x) là rất khó

3. ***Tính chống trùng mạnh*** *(Strong Collision resistant):*

   Không thể tính toán để tìm được hai thông điệp đầu vào

   $x_1 \neq x_2$ sao cho chúng có cùng giá trị băm

   (Nghịch lý ngày sinh – Birthday paradox)

Nghịch lý ngày sinh (birthday paradox)

Bài toán 1: *Giả sử trong phòng có M sinh viên. Vậy xác suất để có hai SV có cùng ngày sinh là bao nhiêu phần trăm? (1 năm 365 ngày khác nhau)*

- *Theo nguyên lý chuồng bồ câu Dirichlet: cần có 365+1 = 366 người để tìm thấy 2 người có cùng ngày sinh với xác suất 100%. Vì vậy với 30 người thì xác xuất này rất nhỏ. → Rất nhỏ, đúng không*
- *Tính theo xác suất thống kê toán học thì*

$$M(M-1) >= 2 \times 365 \times \log_e 2 \quad (*)$$

chỉ cần 23 người là đủ để xác suất hơn 50%. Vì vậy bài toán này gọi là ***nghịch lý ngày sinh***

Nghịch lý ngày sinh (birthday paradox)

Điều này muốn nói lên rằng, *trong nhiều trường hợp xác suất để hai mẩu tin có cùng bản Hash là không nhỏ như chúng ta nghĩ.*

→Tính chống trùng mạnh

Nghịch lý ngày sinh (birthday paradox)

Bài toán 2: Giả sử bạn đang ở trong một lớp học với **M** sinh viên. Hỏi M tối thiểu là bao nhiêu để tồn tại một bạn khác có cùng ngày sinh với bạn với xác suất (XS) lớn hơn 50%?

- XS để 1 người khác ngày sinh với bạn là 364/365.
- →XS để M người đều khác ngày sinh với bạn là $(364/365)^M$.
- →XS để tồn tại ít nhất một người có cùng ngày sinh với bạn là: $1-(364/365)^M$
- Để XS này >50% ➔ **M>=253 người**

→Tính chống trùng yếu

Nghịch lý ngày sinh (birthday paradox)

- Áp dụng cho hàm băm, ta thấy **tính chống trùng mạnh giống bài toán 1**; **tính chống trùng yếu giống bài toán 2**.
- Gọi **n** là số bit của giá trị băm h, có **$N=2^n$** giá trị băm khác nhau. Giả sử 2^n giá trị băm này là ngẫu nhiên, có khả năng xuất hiện như nhau.
- Thay 365 của bất phương trình (*) bằng 2^n

$$M(M-1)>=2 \times 2^n \times \log_n 2$$

Giải bất phương trình trên, ta có xấp xỉ

$$M > \sqrt[2]{2^n} = 2^{n/2}$$

Nghịch lý ngày sinh (birthday paradox)

- Để tìm ra **hai thông điệp có cùng giá trị băm** (vét cạn) thì phải thử bao nhiêu thông điệp khác nhau?

→Phải thử khoảng $2^{n/2}$ thông điệp khác nhau (xác suất > 50%)

- Ví dụ: Nếu **n=128** thì phải thử 2^{64} thông điệp (khá lớn), nghĩa là hàm hăm đạt được tính chống trùng mạnh (tương được tấn công vét cạn khóa của DES)

2. Ứng dụng hàm băm

- **Chứng thực thông điệp**

  (Message Authentication)
- **Chữ ký số**

  (Digital Signatures)
- **Các ứng dụng khác**

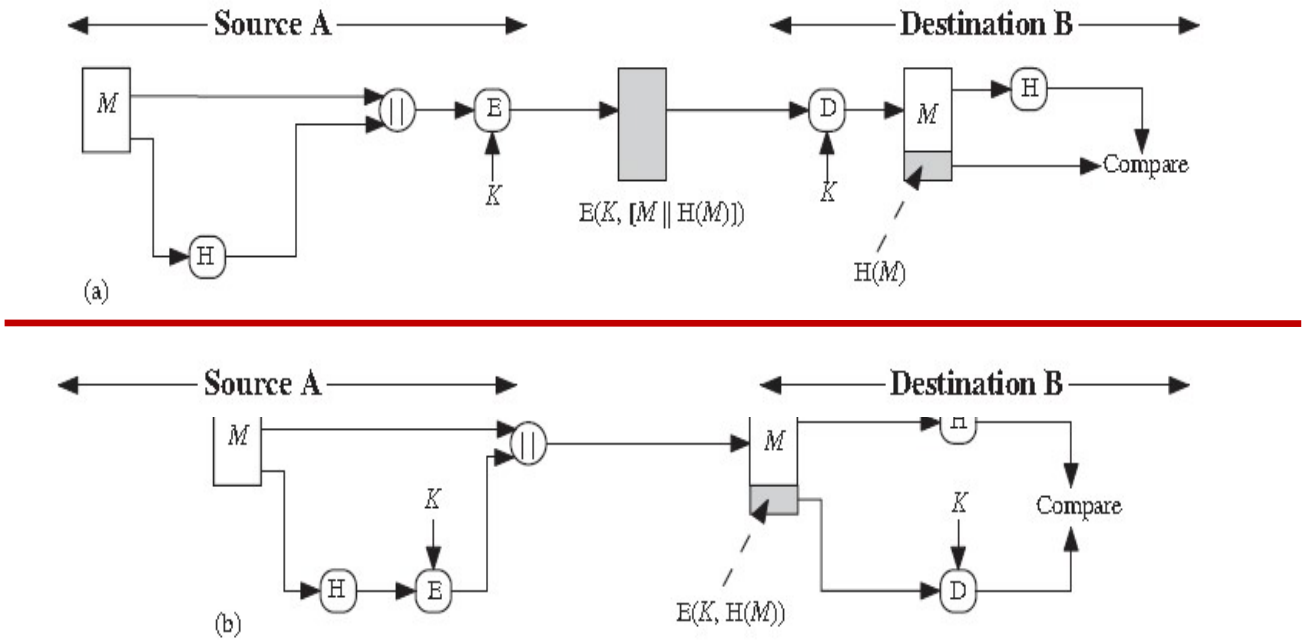
  (Other Applications)

2.1 Message Authentication

- Là một cơ **chế/dịch vụ** được dùng để **kiểm tra tính toàn vẹn** của một thông điệp.
- Đảm bảo rằng dữ liệu nhận được là chính xác như khi được gửi (không bị chỉnh sửa, chèn, hoặc thay thế)
- Trong nhiều trường hợp, có một yêu cầu là cơ chế chứng thực phải hỗ trợ nhận dạng người gửi (sender) là hợp pháp.
- Hàm băm dạng này, giá trị băm (h) được gọi là **tóm tắt thông điệp hoặc cốt thông điệp (*message digest*)**

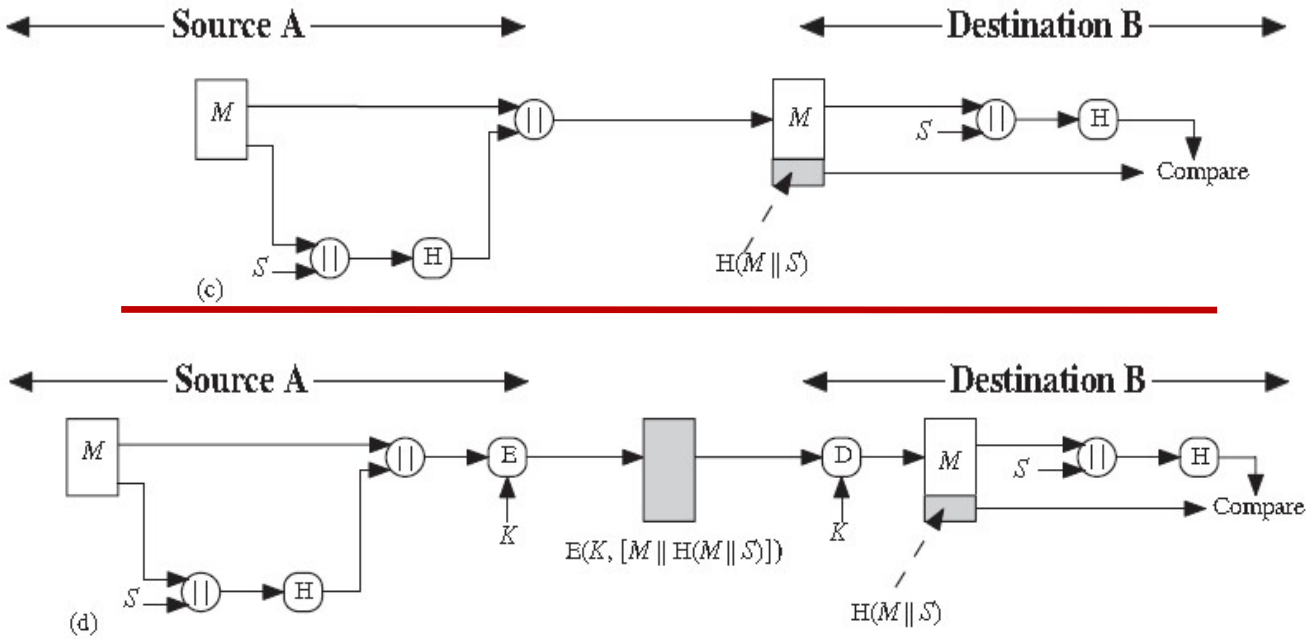
2.1 Message Authentication

- Ví dụ cơ chế chứng thực đơn giản



2.1 Message Authentication

- Ví dụ cơ chế chứng thực đơn giản (tt)



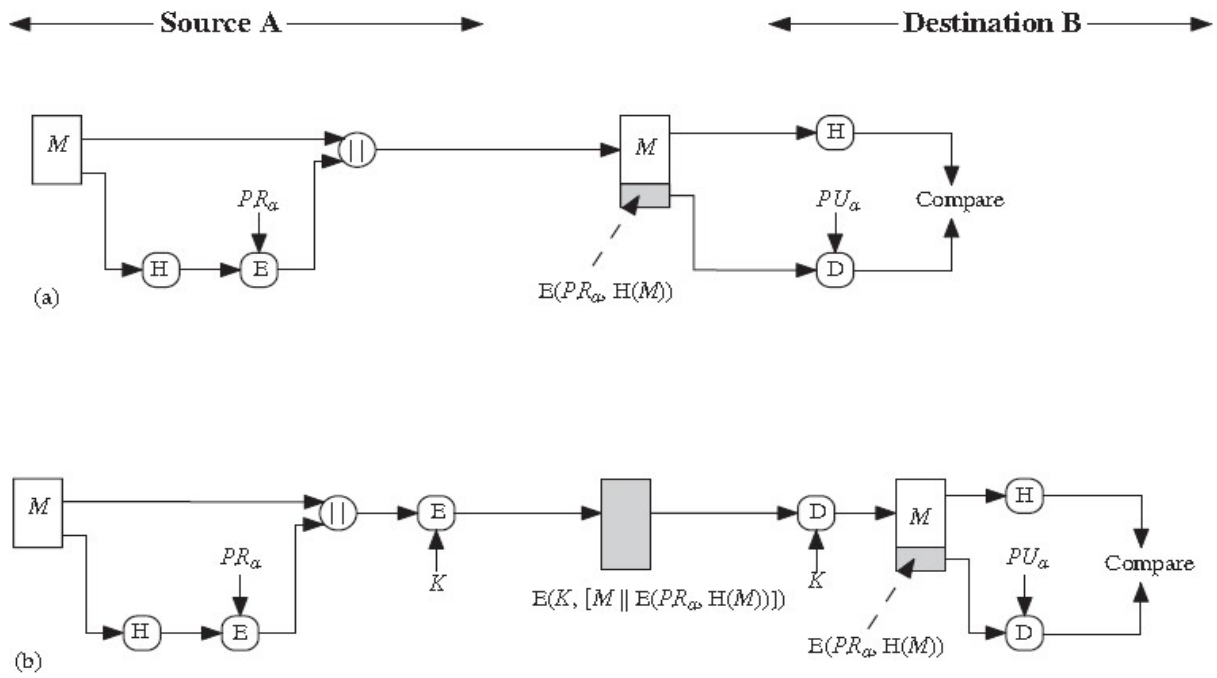
2.2 Digital Signatures

- Giá trị băm của thông điệp được mã hóa bằng **private key** của user, bất kỳ ai biết **public key** của user thì có thể thẩm tra thông điệp mà được gắn kết với chữ ký số.

- Kẻ tấn công muốn hiệu chỉnh thông điệp thì sẽ **cần phải biết private key** của user.

2.2 Digital Signatures

- Ví dụ:



2.3 Other Applications

Dùng lưu trữ mật khẩu (băm password):

- Hàm băm được dùng để tạo **one-way password file**, trong cơ chế này giá trị băm của password được lưu, điều này tốt hơn là lưu chính bản rõ password. → password không bị truy xuất bởi kẻ tấn công nơi chứa password.
- Khi user nhập vào một password, thì giá trị băm của password được so với giá trị băm được lưu để kiểm tra.

2.3 Other Applications

Dùng nhận diện xâm hại (*intrusion detection*) và nhận diện virus (*virus detection*).

- Tính, lưu và bảo mật giá trị băm H(F) của các tập tin trong hệ thống (thể lưu trên CD-R)
- Kẻ xâm hại cần phải hiệu chỉnh F mà không thay đổi H(F)

2.3 Other Applications

- **Dùng:**

  Xây dựng hàm ngẫu nhiên giả

  (*pseudorandom function - PRF*)

  hoặc

  Phát sinh số ngẫu nhiên giả (*pseudorandom number generator - PRNG*)

SHA-1

- SHA = Secure Hash Algorithm
  - Được đề xuất và bảo trợ bởi NIST
  - Dùng trong hệ DSS (Digital Signature Standard) của NIST
  - Được sử dụng rộng rãi
  - SSL, PGP, SSH, S/MIME, IPSec
- Đầu vào bội số của 512 bit
- Giá trị băm 160 bit
- 80 vòng lặp tính toán

MD5

- MD = Message Digest
- MD5 được đề xuất bởi Rivest vào năm 1991
- Được sử dụng rộng rãi
  - Truyền tập tin
  - Lưu trữ mật khẩu

INDUSTRIAL UNIVERSITY OF HOCHIMINH CITY

CHỮ KÝ SỐ

(DIGITAL SIGNATURES)

GV: Nguyễn Thị Hạnh

Nội dung chính

1. Khái niệm về chữ ký số (Digital Signature)
2. Các dịch vụ bảo mật cung cấp bởi chữ ký số
3. Một vài chữ ký số thông dụng
4. Mô tả vài áp dụng thực tế của chữ ký số

(*Cryptography & Network Security*. McGraw-Hill, Inc., 2007., Chapter 13)

1. Khái niệm về chữ ký số

- 1976, Diffie & Hellman lần đầu tiên đề cập đến khái niệm Chữ ký số
- là một trong ứng dụng quan trọng nhất của mã hóa khóa công khai.
- 1989, phiên bản thương mại Chữ ký số đầu tiên trong Lotus Notes, dựa trên RSA
- Ứng dụng:
  - Hợp đồng số
  - Bầu cử điện tử
  - Giao dịch ngân hàng
  - ….

1. Khái niệm về chữ ký số

Mục tiêu:

- Xác thực (Authentication)

- Chống phủ nhận (Non-repudiation)

→ Chữ ký số không những giúp xác thực thông điệp mà còn bảo vệ mỗi bên khỏi bên kia.

1. Khái niệm về chữ ký số

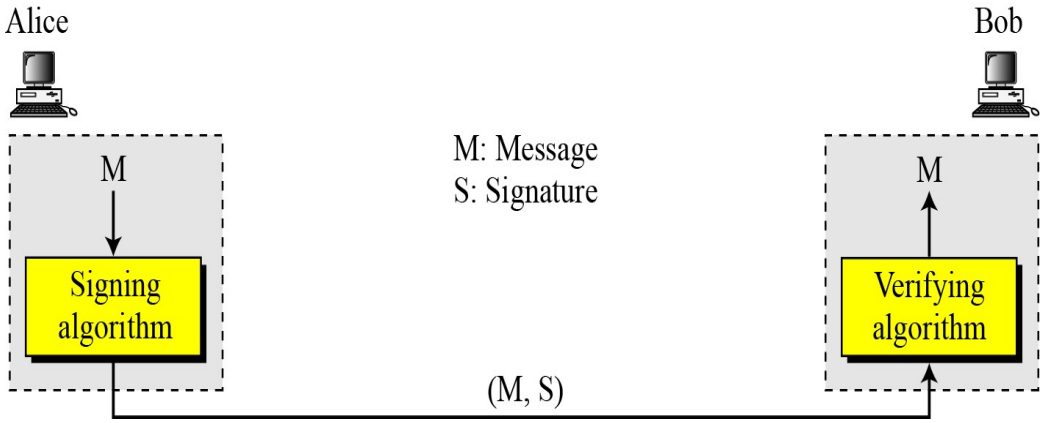
- **Chữ ký số** được xây dựng dựa trên mật mã khóa công khai.

- Nhằm bảo đảm vấn đề **toàn vẹn dữ liệu**, và **chống sự chối bỏ trách nhiệm** khi đã ký.
- Chữ ký điện tử và chữ ký tay đều có chung đặc điểm là rất khó có thể tìm được hai người có cùng một chữ ký. Chữ ký điện tử được người ký tạo ra bằng Khoá riêng và phần có đặc tính duy nhất của văn bản được ký."

Quy trình sử dụng chữ ký số

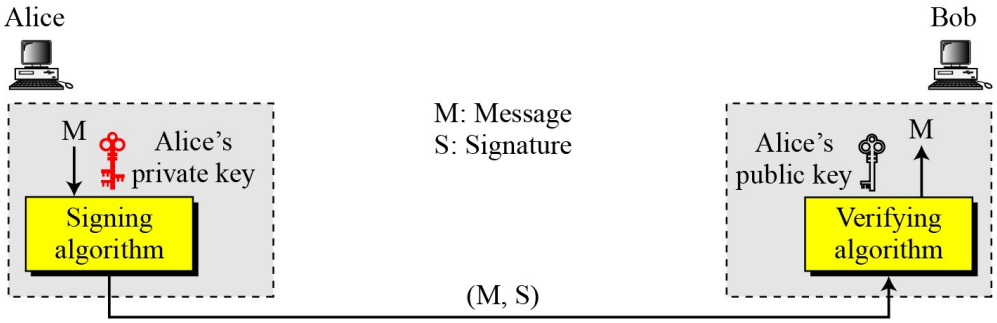
- Người gửi dùng một thuật toán tạo chữ ký (signing algorithm) để ký thông điệp. Thông điệp và chữ ký được gửi cho người nhận.
- Người nhận dùng thuận toán thẩm tra chữ ký (Verifying algorithm) để thẩm tra.
- Nếu đúng, thông điệp được chấp nhận, ngược lại sẽ từ chối thông điệp

Quy trình sử dụng chữ ký số



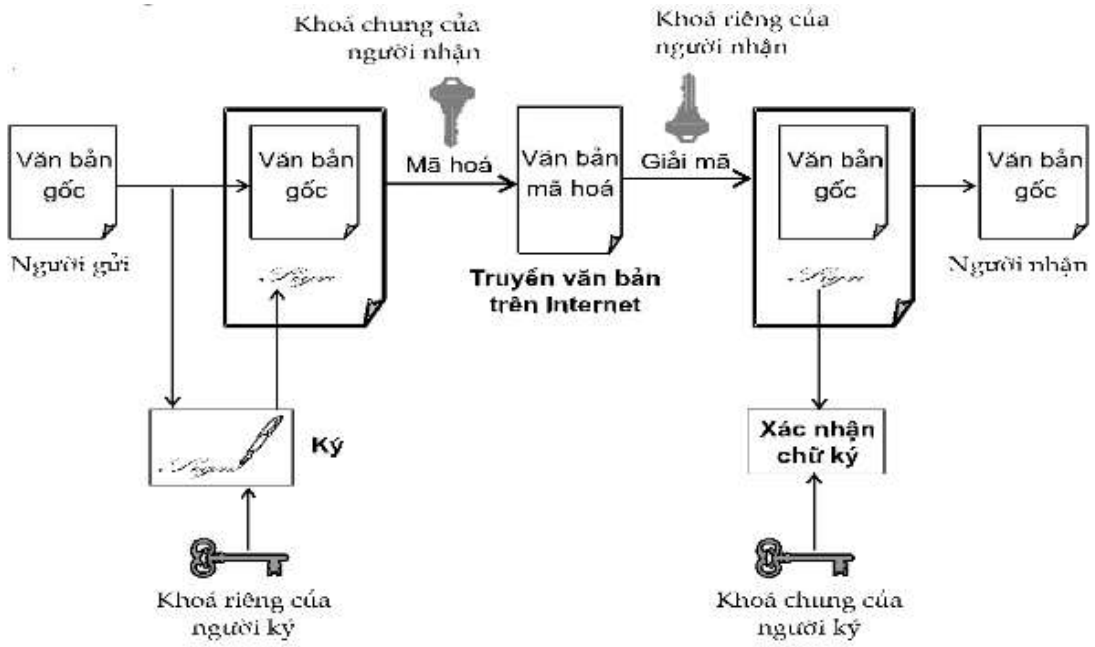
Chữ ký số dựa trên mã hóa khóa công khai

- Một digtal signature cần hệ thống public-key. Người ký (signer) **ký bằng private key** của chính mình; người kiểm tra (verifier) **kiểm tra bằng public key của signer**



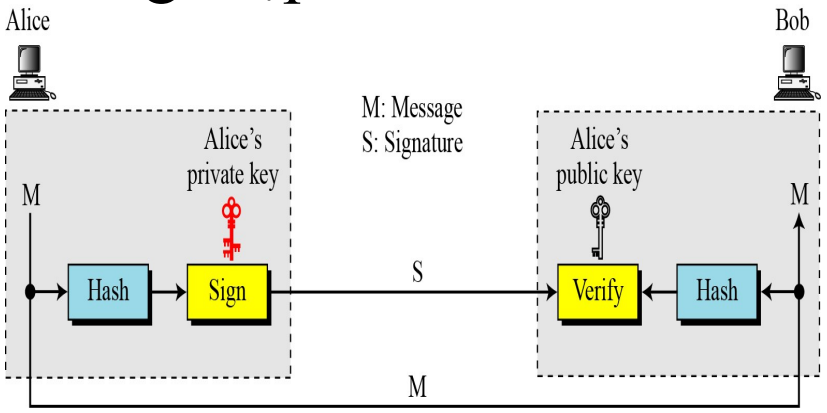
- Dùng khóa bí mật secret key, ký và thẩm tra chữ ký được không?

Quy trình sử dụng chữ ký số

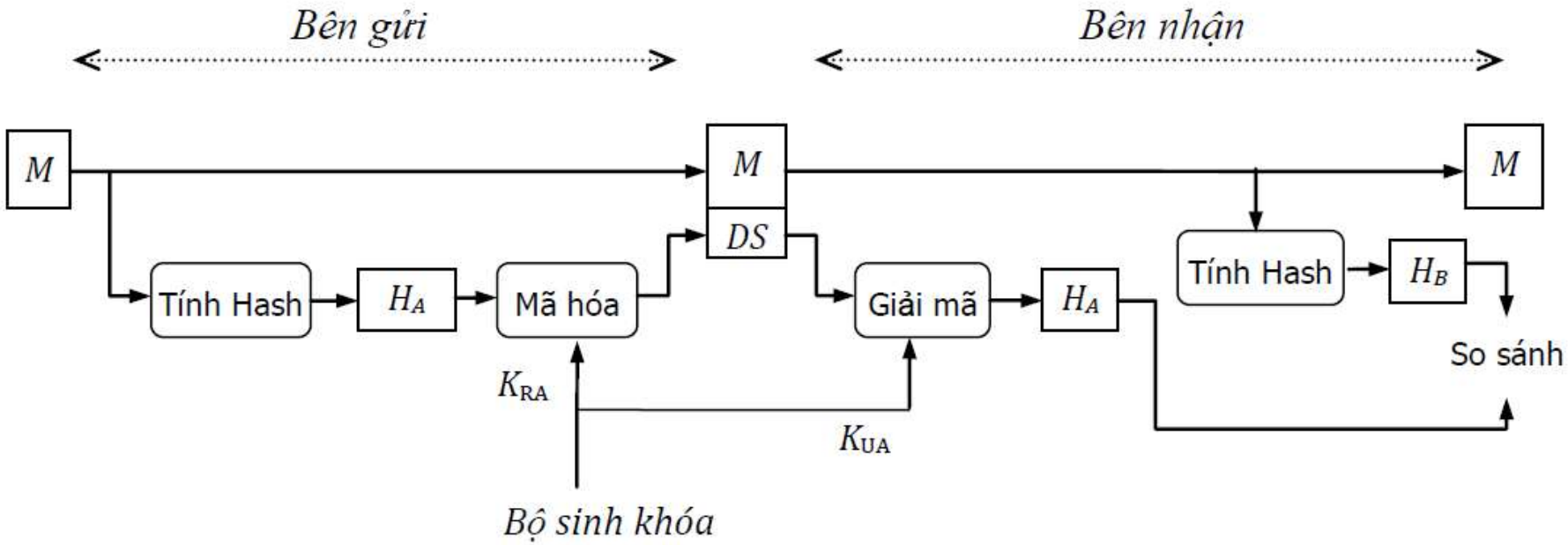


Chữ ký trên giá trị băm của thông điệp

- **Ký trên giá trị băm của thông điệp thì chữ ký sẽ ngắn hơn ký trên thông điệp**
- Người gửi có thể ký trên cốt thông điệp và người nhận có thể kiểm tra trên cốt thông điệp



Chữ ký trên giá trị băm của thông điệp



DS: Data signature – chữ ký điện tử

Nhận xét

- Chữ ký không phải là nét vẽ ngoằn ngoèo khó bắt chước mà là một dãy số trích từ đặc trưng văn bản đã được mã hóa.
- So với chữ ký thông thường, chữ ký số có ưu thế vượt trội hơn chữ ký tay.
  - Chính xác tuyệt đối
  - Kiểm định dễ dàng và chính xác

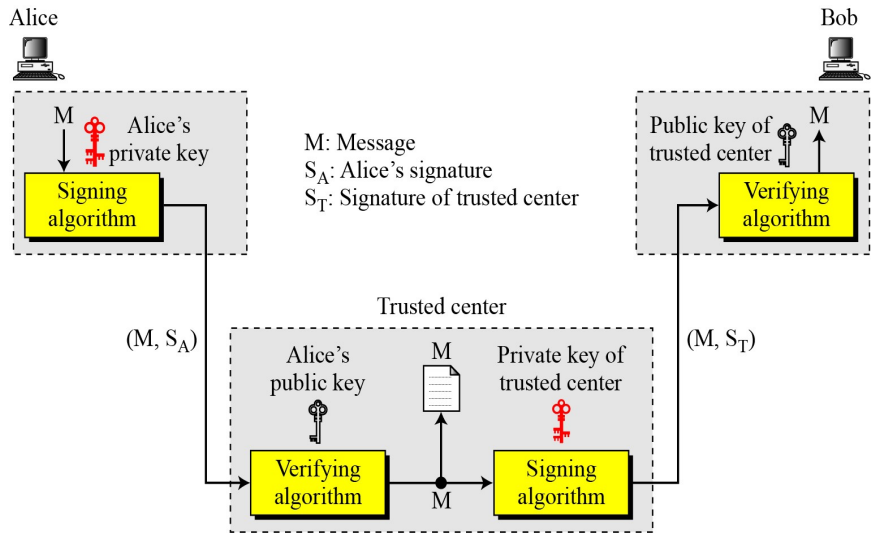
"Chữ ký điện tử mở đường cho các dịch vụ có độ tin cậy cao"

2. Các dịch vụ bảo mật cung cấp bởi chữ ký số

- Xác thực thông điệp (Message authentication)
  - Một cơ chế chữ ký số giống như một chữ ký truyền thống là có thể cung cấp sự xác thực thông điệp
- Toàn vẹn thông điệp (Message Interity)
  - Chúng ta không thể có được cùng chữ ký nếu thông điệp đã được hiệu chỉnh
- Chống từ chối (Nonrepudiaton)
  - Trust center
- Bảo mật (Confidentiality)

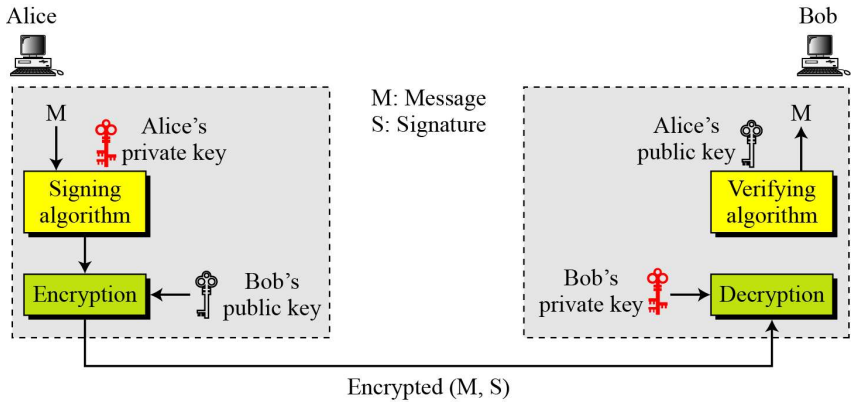
Nonrepudiation

- Nonrepudiation có thể được cung cấp bằng cách dùng một trusted Center



Confidentiality

- Thêm confidentiality vào cơ chế digital signature



Một Digital Signature không cung cấp tính bí mật. Nếu cần bảo mật, một tầng khác của encryption/decryption được áp dụng

Tấn công trên Digital Signature

- Các loại tấn công trên Digital Signature (Attack Types)

- Giả mạo chữ ký (Forgery)

Attack Types

- **Key-Only Attack**: Eve chỉ có thể truy suất đến các thông tin công khai của Alice, Để giả mạo thông điệp, Eve cần tạo chữ ký của Alice để thuyết phục Bob rằng thông điệp đến từ Alice

- **Known-Message Attack**: Eve có thể truy suất đến một hoặc nhiều cặp thông điệp – chữ ký, Eve cố gắng tạo một thông điệp khác và giả chữ ký của Alice trên thông điệp đó

Attack Types

- **Chosen-Message Attack**: Eve bằng cách này hay cách khác sắp đặt Alice ký một hoặc nhiều thông điệp cho cô ta. Eve có một cặp choose-message/signature. Sau đó Eve tạo ra một thông điệp khác, với nội dung cô ta muốn, và giả mạo chữ ký của Alice lên đó

Forgery Types

- **Existential Forgery**: có thể tìm và chỉ ra được thông điệp (có thể vô nghĩa), đối phương có thể tạo chữ ký hợp lệ trên thông điệp này

- **Selective Forgery**: cho trước một thông điệp, đối phương tạo chữ ký hợp lệ trên thông điệp này (rất khó)

3. Vài chữ ký số thông dụng

- RSA Digital Signature Scheme
- ElGamal Digital Signature Scheme
- Schnorr Digital Signature Scheme
- Digital Signature Standard (DSS)
- Elliptic Curve Digital Signature Scheme

Những biến thể chữ ký

- **Time Stamped Signatures**
  - Một tài liệu được ký cần được gắn nhãn thời gian (Timestamped) để ngăn chặn tài liệu bị phát lại (replay) bởi đối phương
  - Ví dụ: Alice ký một yêu cầu đối với nhà băng của cô ta, Bob chuyển tiền cho Eve, Tài liệu yêu cầu này có thể bị chặn và phát lại bởi Eve nếu không có nhãn thời gian gắn trên tài liệu

Những biến thể chữ ký

Blind Signatures

- Giả sử có một tài liệu mà chúng ta muốn có chữ ký mà không muốn tiết lộ nội dung của tài liệu đối với người ký.
- Ví dụ: Nhà khoa học phát minh ra một lý thuyết rất quan trọng mà cần được ký bởi công chứng viên, công chứng viên sẽ ký nhưng sẽ không biết gì về nội dung của phát minh.

Những biến thể chữ ký

Blind Signatures

- Các bước thực hiện:
  - Bob tạo một thông điệp, ẩn (Blind) nó, và gửi thông điệp ẩn này cho Alice
  - Alice ký thông điệp ẩn và trả về chữ ký trên thông điệp ẩn.
  - Bob bỏ ẩn chữ ký để thu về chữ ký trên thông điệp gốc

Chứng thư số

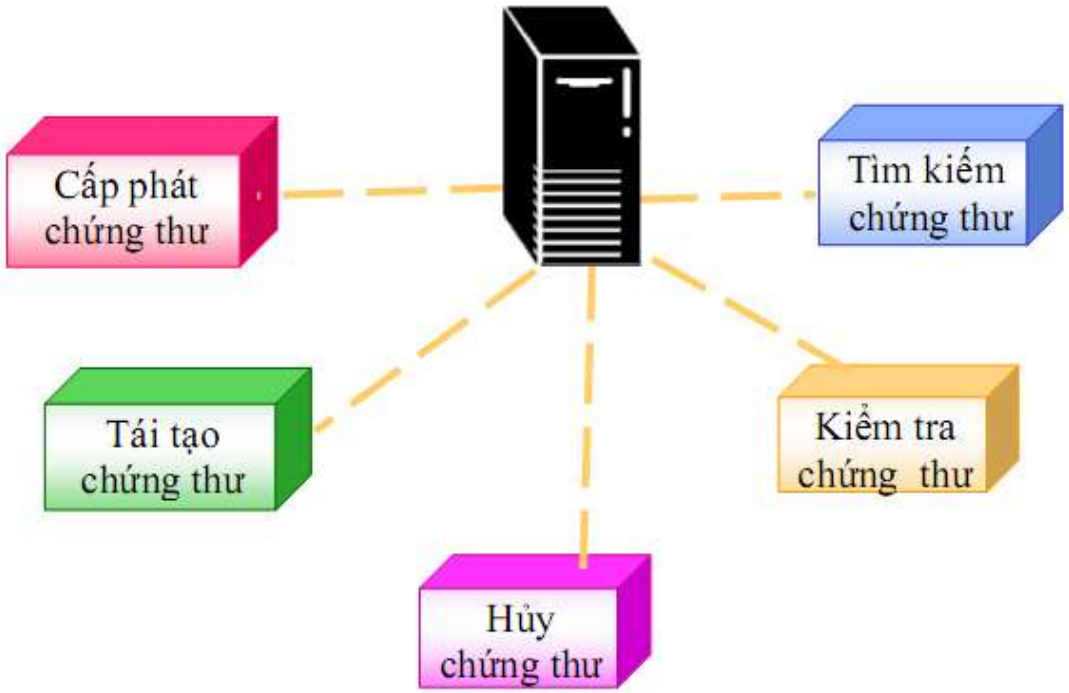
- Chứng thư số (digital certificate):
  - là một **văn bản cung cấp khóa công khai**
  - được cung cấp và quản lý bởi một tổ chức gọi là **nhà cung cấp chứng chỉ** (certificate authority, hay viết tắt là CA).
  - Chứng thư số hoạt động nhờ vào nguyên lý **bên thứ ba tin cậy** (trusted third party – TTP), ở đây bên thứ ba chính là CA. Thông thường, 2 bên giao tiếp với nhau không tin nhau, tuy nhiên, nếu họ cùng tin vào một bên thứ ba (TTP), và bên thứ ba đã xác thực 2 bên này, thì 2 bên này sẽ tin tưởng lẫn nhau.

Chứng thư số

- Là chứng thực để gắn một chìa khóa công khai với một thực thể (cá nhân, máy chủ, cty,…). Hay nói cách khác, CTS giúp xác định chìa khóa công khai thuộc về thực thể nào.
- Một CTS thường gồm chìa khóa công khai và một số thông tin khác về thực thể sở hữu chìa khóa đó.
- Chứng thư số thuộc sở hữu của nhà cc chứng thư số, viết tắt CA (certificate authority).

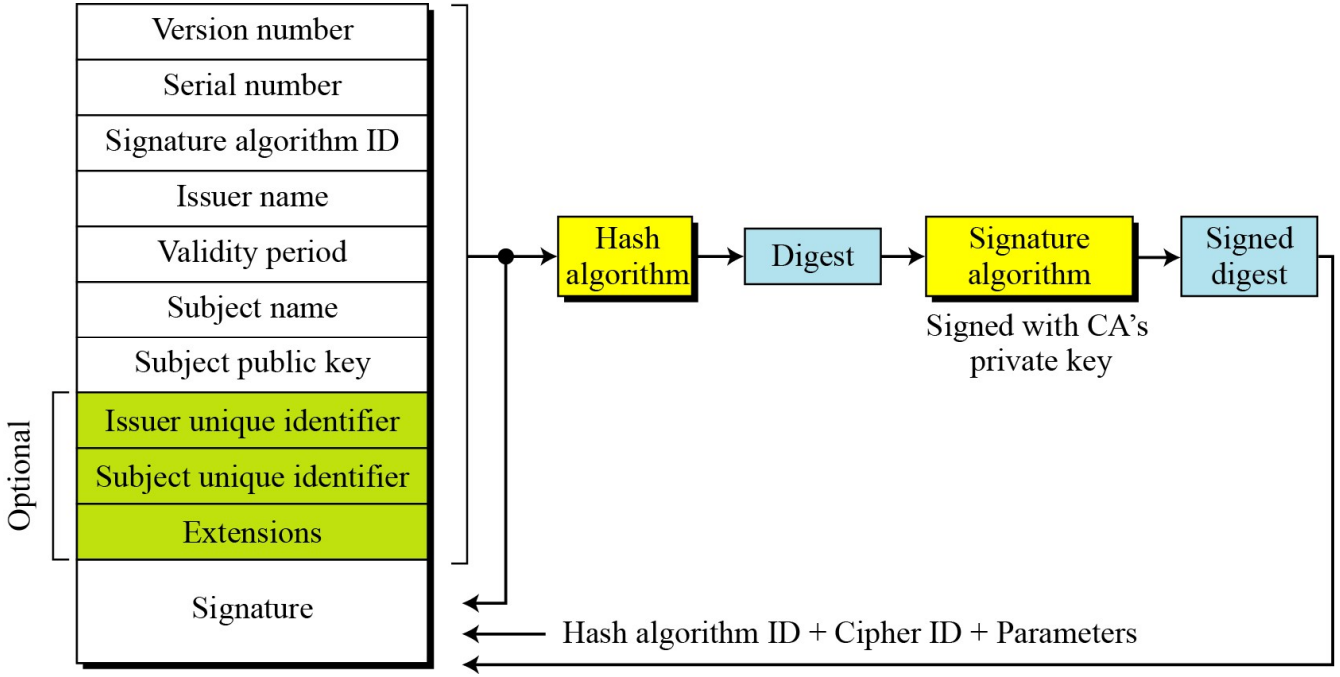
Cơ quan thẩm quyền phát hành chứng thư số (CA)

Các chức năng cơ bản



Chứng thư số với X.509

X.509 Certificate Format



Chứng thư số với X.509

- Version: Chỉ định phiên bản của chứng nhận X. 509
- Serial Number: Số loạt phát hành được gán bởi CA. Mlo64i CA nên gán một mã số loại duy nhất cho mỗi giấn chứng nhận mà nó phát hành
- Signature Algorithm ID: Chỉ rõ thuật toán tạo chữ ký mà được CA sử dụng để ký giấy chứng nhận
- Issuer Name: Tên tổ chức CA phát hành chứng thư số. Hai CA khác nhau không được dùng cùng một tên phát hành
- Validity Period: Gồm hai giá trị chỉ định khoảng thời gian mà giấy chứng nhận có hiệu lực

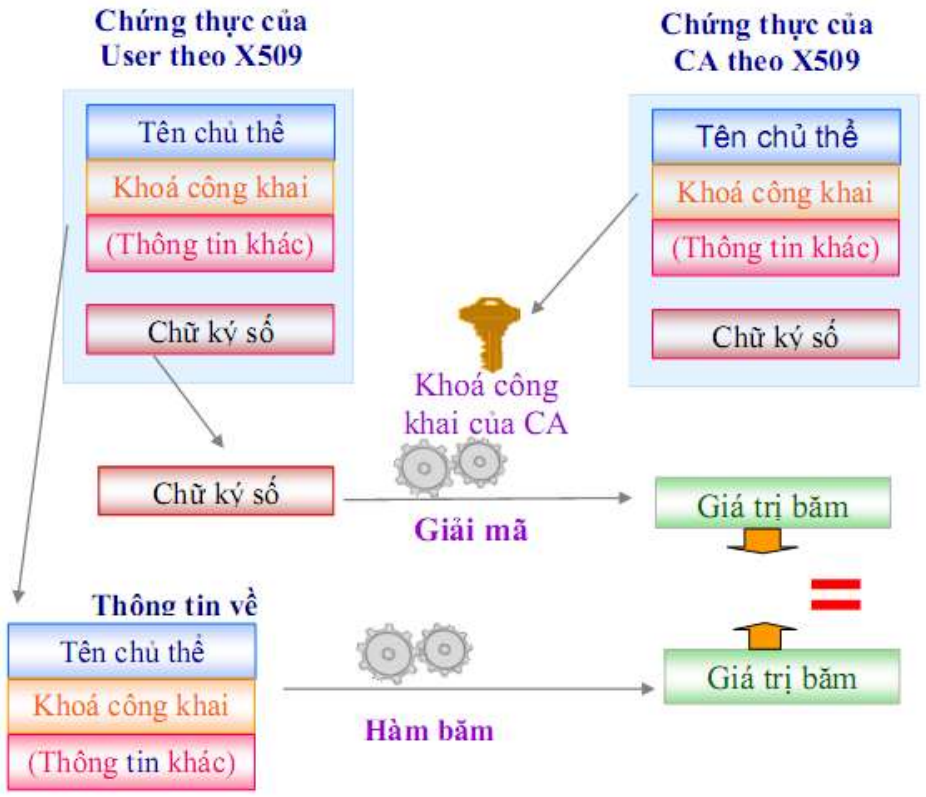
Chứng thư số với X.509

- Subject Name: Tên chủ thể được cấp chứng thực
- Public key: Khóa công khai của chủ thể được chứng thực
- Signature: Chữ ký số được tổ chức CA áp dụng

Chứng thư số với X.509

- Thay mới (Certificate Renewal)
  - Certificate hết hạn dùng, nếu không có vấn đề thì CA cấp lại một certificate mới trức khi hết hạn
- Thu hồi (Delta Revocation)
  - Private key ứng với Public Key có thể bị thỏa hiệp
  - CA không muốn chứng nận người dùng nữa
  - Private Key của CA mà có thể dùng thẩm tra Certificate, có thể đã bị thỏa hiệp
  - → CA phát hành danh sách các thu hồi (Certificate Revocation List - CRL)

Sơ đồ kiểm tra chứng thư số

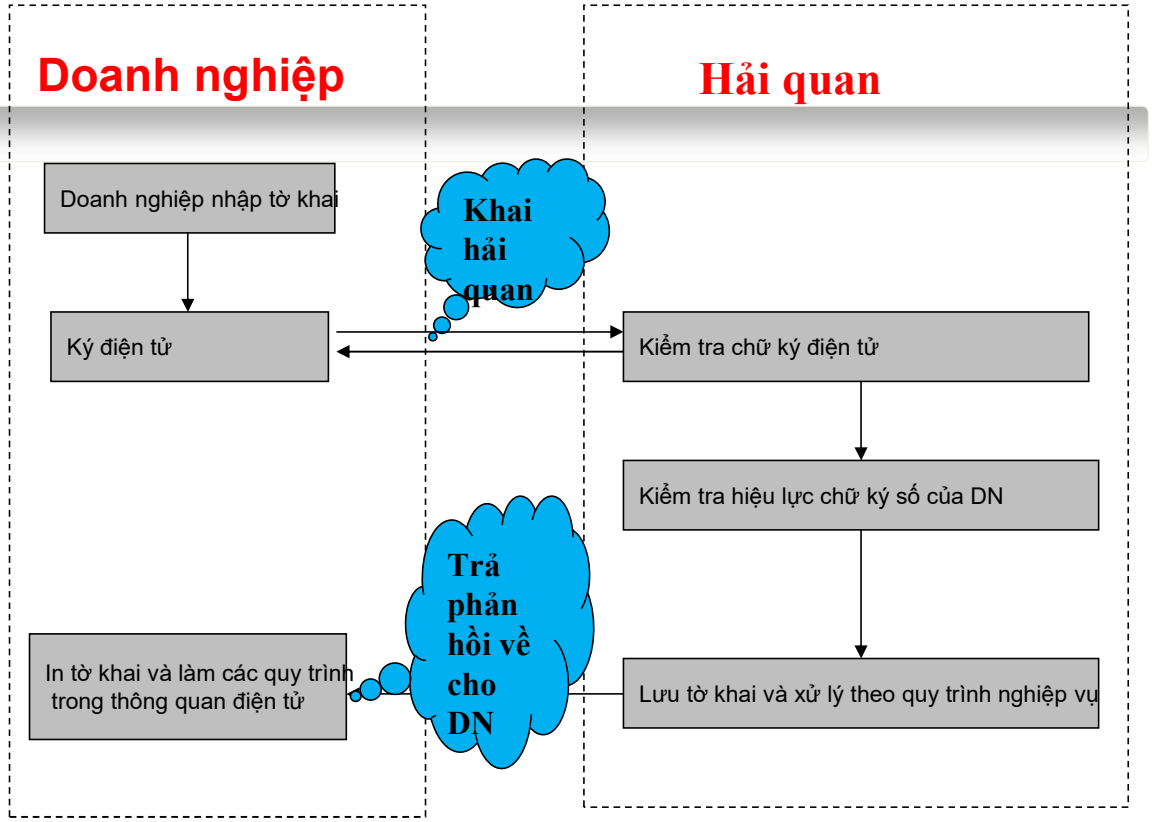


Tình hình sử dụng chữ ký điện tử trong DN

Theo "Báo cáo tình hình phát triển và ứng dụng chữ ký số tại Việt Nam năm 2018", tính đến thời điểm 31/3/2019 có:

- 703.753 DN sử dụng chữ ký số trong lĩnh vực thuế trên tổng số 711.748 DN đang hoạt động, đạt tỷ lệ 98,87%, tăng 55.623 tổ chức, doanh nghiệp so với năm 2018;
- 232.431 DN sử dụng chữ ký số trong hoạt động trong lĩnh vực hải quan, tăng 90.338 doanh nghiệp so với năm 2018
- 441.096 doanh nghiệp sử dụng chữ ký số trong kê khai bảo hiểm xã hội, tăng 231.678 doanh nghiệp so với năm 2016.
- 2.824 DN dùng chữ ký số trong lĩnh vực chứng khoán, năm 2018 là 2.815

VÍ DỤ CHỮ KÝ SỐ TRONG THỦ TỤC HQĐT



Các nhà cung cấp chữ ký điện tử ở VN

09 doanh nghiệp được Bộ TT & TT cấp phép triển khai dịch vụ chữ ký số công cộng bao gồm:

1. VINA-CA
2. VIETTEL-CA
3. BKAV-CA
4. VNPT-CA
5. NEWTEL-CA
6. NACENCOMMSCT-CA
7. FPT-CA
8. CK-CA
9. SAFECERT-CA