Privacy and Security

- General concepts
  - Such as “privacy of an actor” and “home security”
  - Here, we focus on information processing
- Privacy
  - To seclude the information about an individual or group
    - Mainly by anonymization
- Security
  - To protect the stored and transferred information from leaks or corruptions
    - Mainly by cryptography

Anonymization

- To avoid to be identified a person (or a group) from given data
  - Identifier: passport number, student number, name, …
  - Linkable: tweet(time, GPS), blog(name, place) : name of tweeter
- In database
  - Anonymize by removing identifier (for OLAP)
  - Identify a person by combination of data
    - Example: Even though anonymize name, if you know the age of Alice and the database has only one tuple about a female whose age is 23…

name	age	gender	purchase
Alice	23	female	beef
Bob	23	male	beer
Carol	24	female	pork
David	22	male	milk

name	age	gender	purchase
*	23	female	beef
*	23	male	beer
*	24	female	pork
*	22	male	milk

k-anonymity [Latanya Sweeney 2002]

- k-anonymity: A release of data is said to have the *k*-anonymity property if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appear in the release.
- k-anonymization:

name	age	gender	purchase	
*	20< x <25	female	meat	k = 2
*	20< x <25	female	meat	
*	20< x <25	male	drinks	k = 2
*	20< x <25	male	drinks	

Cryptography (1/2)

- Long history
  - In ancient Egypt: secret hieroglyphics
  - In ancient Greece: the scytale of Sparta
  - In ancient Roma: Caesar cipher
    - Simply shifts the letters in the alphabet by a constant number of steps
    - "RETURN TO ROME" -> "UHWXUQ WR URPH"

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
DEFGHIJKLMNOPQRSTUVWXYZABC
```

In World War 2

German Enigma encryption machine

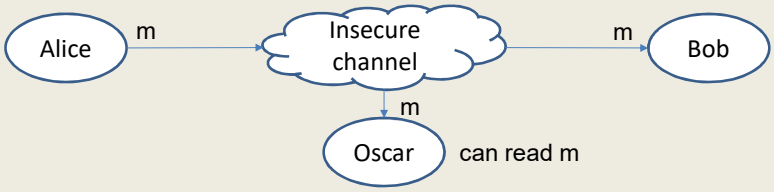
Alan Turning (The Imitation Game)

Cryptography (2/2)

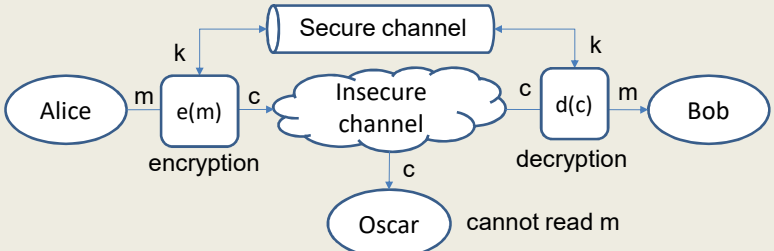
- Symmetric Algorithm
  - Traditional cryptography
- Asymmetric (or Public-Key) Algorithm
  - Introduced after 1976
- Cryptographic Protocols
  - Symmetric / Asymmetric algorithms are building blocks

Symmetric Cryptography

- Insecure channel

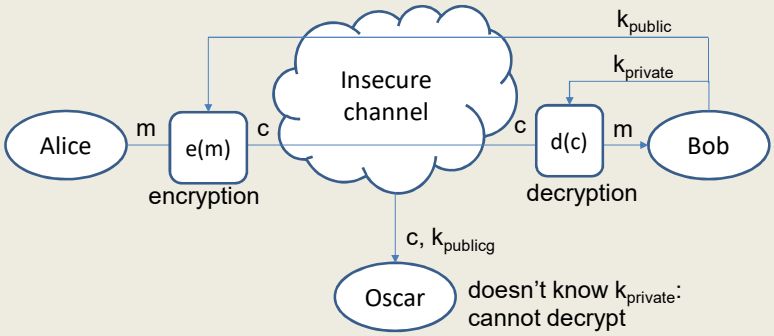


- Symmetric-key cryptosystem



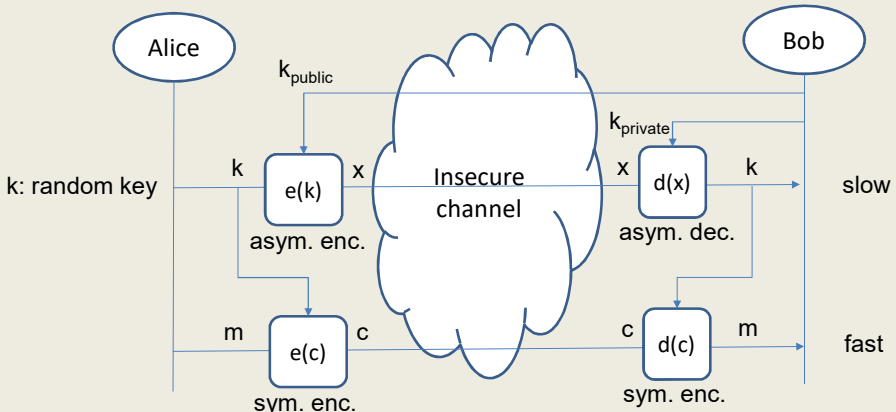
Asymmetric Cryptography

- Public-key cryptosystem (requires no secure channel)
  - Generate a pair of keys: public and private keys
  - Encrypt by public key, decrypted by private key
    - $c = e_{k_{public}}(m)$, $m = d_{k_{private}}(c)$



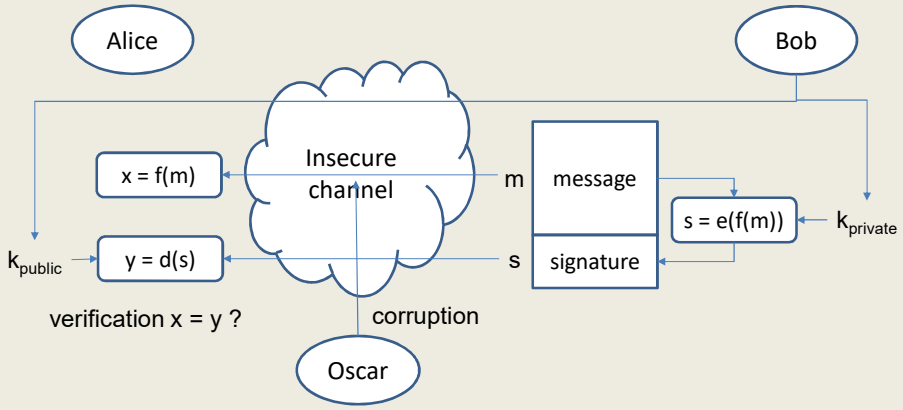
Key Transfer Protocol

- Symmetric key can be transferred using asymmetric cryptography



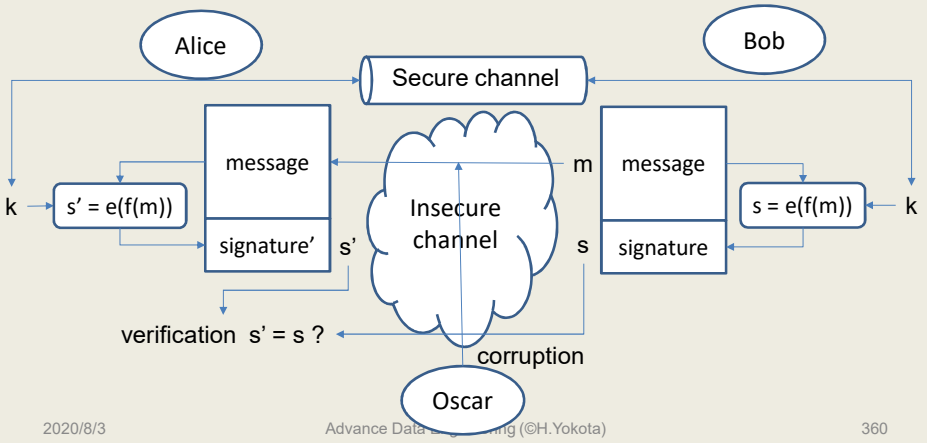
Digital Signature

- Only the person who creates a digital message should be capable of generating a valid signature.



Message Authentication Codes

- The difference between the message authentication codes (mac) and digital signature is that mac uses a symmetric key



Deterministic/Probablistic Encryption

- Deterministic Encryption

  $m_1 = m_2 \Rightarrow e(m_1) = e(m_2)$

  – Can be used for verification and search

  – Vulnerability

- Probablistic Encryption

  $m_1 = m_2 \Rightarrow e(m_1) \neq e(m_2)$

  – Cannot be used for verification or search

Cryptography Algorithms

- Symmetric
  - DES: Data Encryption Standard
  - AES: Advanced Encryption Standard
- Asymmetric
  - RSA: Ronald Rivest, Adi Shamir, Leonard Adleman
    - Based on large integer number factorization problem
  - Discrete Logarithm Problem based
    - Diffie-Hellman Key Exchange
  - Elliptic Curve Crhptosystem

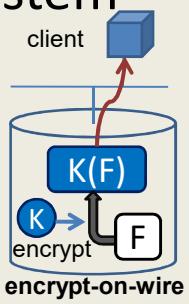
Security Levels

- If the best known attack requires 2^n steps
  - Security level of *n* bit

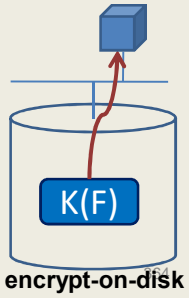
Algorithm Family	Cryptosystems	Security Level (bit)			
		80	128	192	256
Integer factorization	RSA	1024 bit	3072 bit	7680 bit	15360 bit
Discrete logarithm	DH, DSA, Elgamal	1024 bit	3072 bit	7680 bit	15360 bit
Elliptic curves	ECDH, ECDSA	160 bit	256 bit	384 bit	512 bit
Symmetric-key	AES, 3DES	80 bit	128 bit	192 bit	256 bit

Encryption on a storage system

- Encryption schemes for the confidentiality on a network storage [Riedel et al., 2002]
  - Encrypt-on-wire scheme
    - Data is stored in clear, and encrypted when transmitted (e.g., SSL: Secure Socket Layer)
  - Encrypt-on-disk scheme
    - Data is stored in cipher, and transmitted without any encryption process
- Encrypt-on-disk scheme is more efficient than encrypt-on-wire scheme for the performance and confidentiality.
  - Storage server does not require as much encryption work with data transfer.
  - Encrypt-on-disk scheme protects data in storage while encrypt-on-wire scheme cannot.



encrypt-on-wire



encrypt-on-disk

Revocation on encrypt-on-disk (1/2)

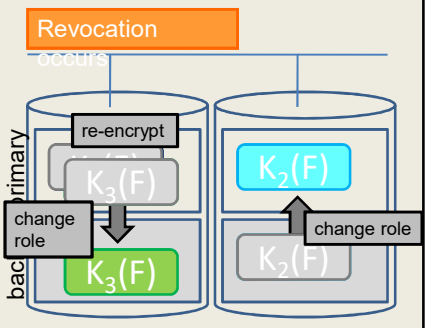
- With encrypt-on-disk, shared files must be re-encrypted when revocations occur.
  - There are possibilities of information leakage , if the revoked user holds the cryptographic key and intercepts the files.
- Re-encryption methods [Fu, 1999]
  - Active Revocation:
    - Files are re-encrypted immediately after the revocation.
    - It is enough secure
      - Revoked users are immediately unable to decrypt data
    - It has a problem of performance
      - Even authorized users cannot access them until re-encryptions are completed.

Revocation on encrypt-on-disk (2/2)

- Lazy Revocation:
  - The re-encryptions is delayed until the files are next updated
  - It is more efficient in respect of performance
    - Encryption involved in update process can be combined with the re-encryption required for revocations
    - The re-encryption work for multiple revocations are performed together if the file is not frequently updated
  - It is vulnerable
    - Data stored before update are still encrypted with the old key, which can be accessed by the revoked users.
- There is a trade-off problem between performance and security.

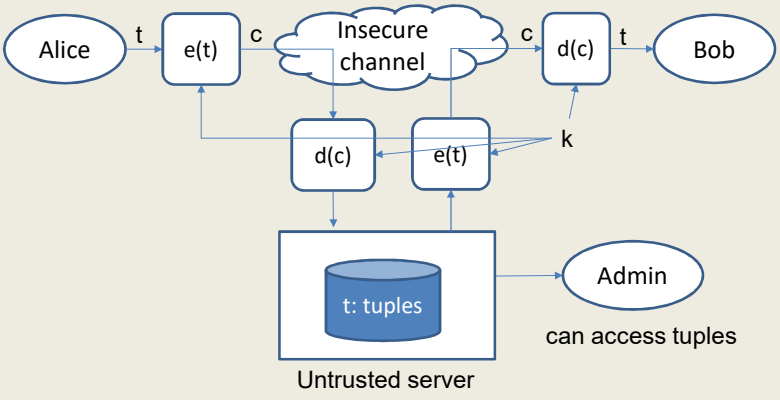
BA-Rev (Backup Assisted Revocation)

- We have proposed BA-Rev to attack the trade-off problem. [Takayama et al., 2007]
  - BA-Rev utilizes the primary-backup configuration.
- Outline
  1. Stores backup data with encrypted by key (K_2) different from that in primary (K_1)
  2. When a revocation occurs, their roles is changed
  3. Old primary data is re-encrypted by another key (K_3) and stored as backup
- Do not need wait for re-encryption
  - Authorized users can access the file immediately after the revocation.
  - Re-encryption processes are performed in background.



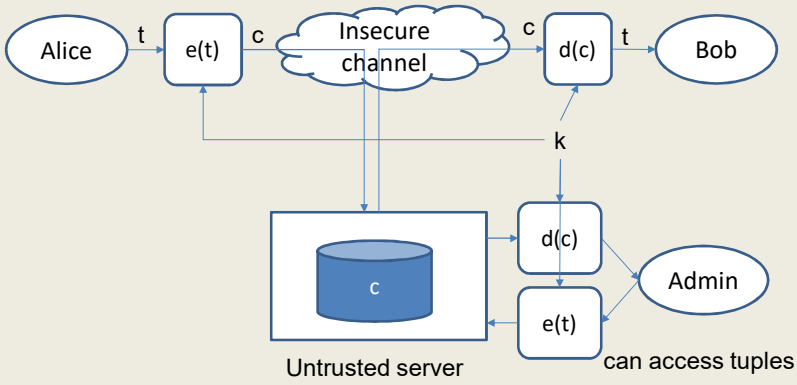
Untrusted Server

- Encrypted on wire approach
  - Malicious administrator



Untrusted Server

- Even encrypted on disk approach
  - If malicious administrator knows the key



Proxy Re-encryption

- Proxy re-encrypts cypher data