Cryptology and Information Security—Theory and Practice

D. J. Guan

Department of Computer Science
National Sun Yat-Sen University
guan@cse.nsysu.edu.tw

April 14, 2016

Abstract

In this talk, I will introduce Cryptology, which is the foundation of Information Security. I will emphasize the gap between the theory and implementation of cryptosystems. I will also talk about digital signature which is very important in the processing of official digital documents. Finally, I will introduce quantum cryptography, which is important if attackers have quantum computers.

Contents

Cryptology and Information Security

A sender S wants to send a message m to a receiver R by using a public channel.

$$S \Longrightarrow \stackrel{m}{\Longrightarrow} \Longrightarrow R$$

An eavesdropper may learn the secret m.

$$S \xrightarrow{m} \boxed{E_{k_e}(m)} \stackrel{c}{\Longrightarrow} \boxed{D_{k_d}(c)} \xrightarrow{m} R$$

$$D_{k_d}(E_{k_e}(m)) = m$$

Introduction to Cryptology and Information Security

Cryptography is the study of mathematical techniques related to aspects of information security such as:

1. Confidentiality, (Secrecy, or Privacy)
2. Data integrity
3. System Availability
4. Entity identification
5. Data authentication
6. Non-repudiation

The Goal of Information Security

Provide a system which can function properly, even if there are malicious users.

1. Can we design a secure system?
2. Can we prove that a system is secure?

The Gap between Theory and Implementation

The theory of modern cryptography is based on mathematics, algorithm and computational complexity.

In this talk, I will not emphasize on the theory of cryptography.

I will discuss more on the gap between the theory and implementation of cryptosystems.

Symmetric Key Cryptosystems

1. Traditional Cryptosystems
   shift cipher, substitution cipher, Vigenere Cipher, ...

2. Modern Cryptosystems
   (a) Block cipher: DES, AES, ...
   (b) Stream cipher: linear feedback shift register, ...

Symmetric Key Cryptosystems

Implementation: efficiency

Key selection:

1. Low entropy: passwords
2. High entropy: hash of passwords

Information Entropy

Entropy is a measure of uncertainty.

"Compress then encrypt" or "encrypt then compress" ?

Public Key Cryptosystems

$$A \stackrel{x}{\Longrightarrow} B$$

- Key generation
  1. B randomly chooses two large distinct primes p and q, (e. g. $p, q > 2^{1024}$).
  2. B computes $n = p \cdot q$ and $\phi(n) = (p-1)(q-1)$.
  3. B randomly chooses e, $\gcd(e, \phi(n)) = 1$.
  4. B computes $d \equiv e^{-1} \bmod \phi(n)$.
  5. B sends (n, e) to A.

Public Key Cryptosystems

- Encryption
  1. A computes $y = x^e \bmod n$.
  2. A sends y to B.
- Decryption
  1. B computes $x = y^d \bmod n$.

Security of RSA

1. If n can be factored efficiently, then RSA cryptosystems is not secure.

2. If d or e is too small, then RSA cryptosystems is not secure, even if n is very large.

3. Generate different set of keys (n, e_0, d_0) and (n, e_1, d_1) with the same modulus n is not secure.

Factoring Large Integers

1. If $\phi(n) = (p-1)(q-1)$ is known, then n can be factored.
2. If $|p-q|$ is small, e. g. $|p-q| < \sqrt[4]{n}$, then n can be factored.
3. If every prime power factor of $p-1$ is small, then n can be factored.
4. If every prime power factor of $p+1$ is small, then n can be factored.
5. If every prime power factor of $p+1 \pm 2\sqrt{p}$ is small, then n can be factored.

Factoring Large Integers

RSA Number	digits	bits	Factored on
RSA-100	100	330	1991/04/01
RSA-110	110	364	1992/04/14
RSA-120	120	397	1993/06/09
RSA-129	129	426	1994/04/26
RSA-130	130	430	1996/04/10
RSA-140	140	463	1999/02/02
RSA-150	150	496	2004/04/16
RSA-155	155	512	1999/08/22
RSA-160	160	530	2003/04/01
RSA-170	170	563	2009/12/29
RSA-576	174	576	2003/12/03
RSA-180	180	596	2010/05/08
RSA-640	193	640	2005/11/02
RSA-200	200	663	2005/05/09
RSA-768	232	768	2009/12/12

How to Select Primes in RSA

Randomly select large primes of the same size.

Random?

1. pseudo-random number generators: random()
2. /dev/urandom files
3. quantum devices

More on RSA Cryptosystem and Factoring

Theorem 1 *If the secret key* (d) *can be computed from the public key* (e *and* n) *efficiently, then* n *can be factored efficiently.*

Is breaking RSA cryptosystem equivalent to factor n?

Other Public-key Cryptosystems

1. Based on Discrete Logarithm Problem
   ElGamal Cryptosystem
2. Use groups defined by elliptic curves
3. Based on solving shortest non-zero vector in a lattice
4. Based on error correction code
5. Based on composition of multivariate functions
6. Based on quantum information

Elliptic Curve Cryptography

1. There is no known adaptation of the index calculus method to the discrete logarithm problem on elliptic curves.

2. It is believed that a cyclic subgroup of an elliptic curve of size 160 bits will provide the same security strength as a cryptosystem based on $\mathbf{Z}_n$ with 512-bit n.

The hardest ECC discrete logarithm problem broken to date had a 112-bit key for the prime field case and a 109-bit key for the binary field case.

Note that some elliptic curves do have index-calculus-like method for solving the discrete logarithm problem.

Bilinear Mapping

Bilinear functions can be constructed by the using additive groups based on elliptic curves.

$$e(\alpha x + \beta y) = e(x + y)^{\alpha\beta}$$

Digital Signature

RSA digital signature scheme: A signs a message m.

1. Key generation
   (a) A randomly chooses two large distinct primes p and q, (e. g. $p, q > 2^{1024}$).
   (b) A computes $n = p \cdot q$ and $\phi(n) = (p-1)(q-1)$.
   (c) A randomly chooses e, $\gcd(e, \phi(n)) = 1$.
   (d) A computes $d \equiv e^{-1} \bmod \phi(n)$.
   (e) A announces (n, e).

Digital Signature

1. Compute Signature
   (a) B computes the signature of m: $y = x^d \bmod n$.
2. Verify
   (a) Given (x, y), everyone can verify the signature by testing if $x \equiv y^e \pmod{n}$ or not.

Hash Function

A cryptographic hash function h is a function from domain A to range B which is easy to compute and hard to invert.

$$h : A \rightarrow B$$

The domain A is usually much larger than the range B.

1. Given x, it is easy to compute $h(x)$.
2. Given y, it is hard to find x, $h(x) = y$.
3. Given x_1, it is hard to find x_2, $x_2 \neq x_1$ but $h(x_1) = h(x_2)$.
4. It is hard to find x_1 and x_2, $x_1 \neq x_2$, but $h(x_1) = h(x_2)$.

Hash Function

To encrypt a large file, it is required to divide the file into small blocks, and encrypt each block.

To sign a large document, we first hash the document, and then sign the hash of the document.

Hash functions: MD5, SHA1, SHA2, SHA3, ...

Security of Hash Functions

1. Birthday attack
2. Wang et al. found collisions for some hash functions.

Quantum Information and Post Quantum Cryptography

In 1982 Richard Feynman observed that certain quantum mechanical effects cannot be simulated efficiently on a traditional computer.

It is speculated that computations may be done more efficiently by using these quantum effects, including superposition and entanglement.

Quantum computing models

1. In 1980 Benioff introduced a quantum Turing machine model.

2. In 1989 Deutch proposed the quantum circuit model.

3. In 1993 Yao showed that the *uniform* quantum circuit model of computation is equivalent to the quantum Turing machine model.

Quantum Computers

Quantum computers make direct use of quantum-mechanical phenomena, such as superposition and entanglement, to perform operations on data.

In 2001, researchers demonstrated Shor's algorithm to factor 15 using a 7-qubit NMR computer.

In 2011, researchers at the University of Bristol created an all-bulk optics system that ran a version of Shor's algorithm to successfully factor 21.

Classical bits and Quantum Bits

classical bits:

$$0, \ 1$$

quantum bits, *qubit*: a superposition of $|0\rangle$ and $|1\rangle$

$$\alpha|0\rangle + \beta|1\rangle,$$

Representation of Qubits

Let $|0\rangle$ and $|1\rangle$ be a basis of the Hilbert space $\mathcal{H}$.

Elements of $\mathcal{H}$ is usually denoted by

$$\alpha|0\rangle + \beta|1\rangle,$$

where α and β are complex numbers with

$$|\alpha|^2 + |\beta|^2 = 1.$$

When measured with $\{|0\rangle, |1\rangle\}$,

1. the probability of obtaining $|0\rangle$ is $|\alpha|^2$, and
2. the probability of obtaining $|1\rangle$ is $|\beta|^2$.

Properties of Qubits

1. Infinite many information can be represented by a qubit.
2. However, when measured, it will give only one bit of information, either 0 or 1.
3. After measurement, the qubit will change its superposition state to either $|0\rangle$ or $|1\rangle$, depending on the outcome of the measurement.
4. It is impossible to examine a qubit to determine its quantum state. (Only if infinite many identical qubits are measured would one be able to determine the values of α and β.)

Efficient Quantum Algorithms

- (1992) Deutsch-Jozsa's algorithm for testing whether a Boolean function is constant or balanced needs only 1 evaluation of the function.
  A classical algorithm needs $2^{n-1}+1$ evaluations of the function.

- (1997) Bernstein-Vazirani's algorithm for determining the value of $a \in \mathbf{Z}_2^n$ in $f_a(x) = a \cdot x$ needs only 1 evaluation of the function.
  A classical algorithm needs n evaluations of the function.

- (1994) Simon's algorithm for determining the period of a function $f : \mathbf{Z}_2^n \mapsto \mathbf{Z}_2^n$ needs only $O(n)$ (expected) evaluation of the function.
  A classical algorithm needs 2^n evaluations of the function.

Efficient Quantum Algorithms

- (1994) Peter Shor's integer factorization algorithm runs in $O(\log^3 n)$ time.

  The best-known classical algorithm needs $O\left(e^{(64/9)(\log n)^{1/3}(\log\log n)^{2/3}}\right)$ time.

- (1995) Lov Grover's search algorithm needs only $\sqrt{n}$ queries.

  Traditional algorithm needs n queries.

Post Quantum Cryptography

1. Based on Factoring: RSA

2. Based on Discrete Logarithm Problem: ElGamal

3. Use groups defined by elliptic curves

4. Based on solving shortest non-zero vector in a lattice

5. Based on error correction code

6. Based on composition of multivariate functions

7. Based on quantum information

Quantum Entanglement

Let $\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_n$ be quantum systems with underlying Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2 \ldots, \mathcal{H}_n$, respectively.

The global quantum system $\mathcal{Q}$ is *entangled* if its state

$$|\phi\rangle \in \mathcal{H} = \bigotimes_{j=1}^{n} H_j$$

cannot be written in the form

$$|\phi\rangle = \bigotimes_{j=1}^{n} |\phi_j\rangle$$

An Example of Entanglement

$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \neq |\phi\rangle \otimes |\varphi\rangle$ for any $|\phi\rangle$ and any $|\varphi\rangle$.

$(\alpha|0\rangle + \beta|1\rangle) \otimes (\alpha'|0\rangle + \beta'|1\rangle) =$
$(\alpha\alpha'|00\rangle + \alpha\beta'|01\rangle + \beta\alpha'|10\rangle + \beta\beta'|11\rangle)$

Entanglement

1. The measurement outcome of entangled qubits are correlated.

2. Entanglement is defined only for pure ensembles, entanglement for mixed ensembles has not been well understood yet.

Quantum Cryptography

1. If the eavesdropper measured the quantum bits, there is a high probability that it will be detected.
2. In 1984, Charles Bennett and Gilles Brassard proposed a quantum key distribution protocol which has been shown to be unconditionally secure.
3. All quantum computations are reversible, some cryptographic primitives, such as two-party secure computation, have been shown to be impossible in quantum settings without additional assumptions.

Thank You