CRYPTOGRAPHY

Cryptography is the study of secret (crypto) writing (graphy) concerned with developing algorithms which may be used to

- conceal the context of some message from all except the sender and recipient (privacy or secrecy), and/or

- verify the correctness of a message to the recipient (authentication)

- form the basis of many technological solutions to computer and communications security problems

CIPHERS IN HISTORY

Cryptology could be be considered as one of humanity's oldest professions.

have a history of at least 4000 years

ancient Egyptians enciphered some of their hieroglyphic writing on monuments

The clay of Phaistos (Cretan-Minoan, 17th century BC), still unciphered.

Herodotus describes how encrypted messages were trasported by messangers

ancient Hebrews enciphered certain words in the scriptures

2000 years ago Julius Ceasar used a simple substitution cipher, now known as the Caesar cipher

Roger Bacon described several methods in the 1200s

Geoffrey Chaucer included several ciphers in his works

Leon Alberti devised a cipher wheel, and described the principles of frequency analysis in the 1460s

Blaise de Vigenere published a book on cryptology in 1585, described the polyalphabetic substitution cipher

Increasing use, esp in diplomacy, war over centuries

One-time pads: Widely used in diplomacy

David Kahn in his beautiful book "The Codebreakers" has numerous such examples.

STEGANOGRAPHY

Methods of concealing text.

Character marking: Selected letters of text are overwritten in pencil. The marks are not visible unless the paper is held at an angle to bright light.

Invisible ink: Substances can be used that leave no visible trace until heat or some chemical is applied.

Pin punctures: Small pin punctures on selected letters are not ordinarily visible unless paper is held in front of light.

Typewriter correction ribbon: Used between lines typed with a black ribbon; results of typing visible only under a strong light.

These techniques have modern analogies (e.g. pixel transformations). Unfortunately, steganography in general requires a lot of overhead. See http://cacr.math.uwaterloo.ca/ dstinson/visual.html for an interesting image.

Steganography has become a modern subject with applications to security. Some important topics include

1. Mimicry (reading between the lines, compression and decompression techniques, endoding and decoding algorithms)

2. Anonymous remailers

3. Secret broadcasting (Dining Cryptographers problem)

MACHINE CIPHERS

Jefferson cylinder: developed in 1790s, comprised 36 disks, each with a random alphabet, order of disks was key, message was set, then another row became cipher

Wheatstone disc: invented by Wadsworth in 1817, but developed by Wheatstone in 1860's, comprised two concentric wheels used to generate a polyalphabetic cipher

Hagelin machine: A truly pioneering machine.

Enigma Rotor machine: one of a very important class of cipher machines, heavily used during 2nd world war, comprised a series of rotor wheelswith internal cross-connections, providing a substitution using a continuosly changing alphabet

See my web page for pictures.

BASIC CONCEPTS

cryptography: the art or science encompassing the principles and methods of transforming an intelligible message into one that is unintelligible, and then retransforming that message back to its original form

plaintext: the original intelligible message

ciphertext: the transformed message

cipher: an algorithm for transforming an intelligible message into one that is unintelligible by transposition and/or substitution methods

key: some critical information used by the cipher, known only to the sender receiver

encipher (encode): the process of converting plaintext to ciphertext using a cipher and a key

BASIC CONCEPTS

decipher (decode): the process of converting ciphertext back into plaintext using a cipher and a key

cryptanalysis: the study of principles and methods of transforming an unintelligible message back into an intelligible message without knowledge of the key. Also called **codebreaking**

cryptology: both cryptography and cryptanalysis

code: an algorithm for transforming an intelligible message into an unintelligible one using a code-book

COMMUNICATION SECURITY

As information highways expand cryptographic techniques will play an important role in satisfying "user-privacy" requirements. Important aspects of security include

- **Authentication**
- **Communication Security**
- **Data Distribution**
- **Digital Cash**
- **Electronic Mail**
- **Electronic Voting**

REQUIREMENTS

In **communication security** it is the security of real-time electronic links, local and wide area networks, link encryption, cellular and ordinary telephony, and faxes.

In **data distribution** it is conditional access (e.g., TV), software distribution, information bulletin boards.

In **digital cash** it is the creation of an electronic system that replaces paper money and is more flexible than credit cards.

In **electronic voting** it is secure distributed computation, elections in shareholders meetings.

GOALS

In **communication security** they include: message privacy, sender and recipient authentication, and nonrepudiation.

In **data distribution** they include broadcast and multicast operations, message privacy, and selective reception.

In **digital cash** they include anonymity, untraceability, transferability, fairness, off-line operations, and universality.

In **electronic voting** they include anonymity, fairness, and accountability.

Tools include: key-agreement protocols, private-key cryptosystems, public-key cryptosystems, digital signatures, certificates, secure hardware, untraceability protocols,..., and **beautiful mathematics**.

Research in cryptography is a diverse and mathematically sophisticated practice. Topics include

- Design and Analysis of Cryptographic Algorithms
- Design and Analysis of Cryptographic Protocols
- Hardware and Software Implementations
- Applications of Cryptography

CAUSES OF SYSTEM VULNERABILITY

In typical applications workstations are attached to LANs. The user can reach other hosts, workstations, and servers in the same LAN that are interconnected via bridges and routers.

Transmissions from station to station is visible on the LAN to all stations. Data is transmitted in the form of packets which contain source/destination IDs, and other information.

On this basis, an eavesdropper can monitor and capture traffic packets. Eavesdropper need not be a local LAN user; it could be anyone to whom the LAN offers a dial-up capability.

Eavesdropping may also occur in any of the communications links which provide connectivity to the system, e.g., by tapping wires used for transmission, attaching a low-power radio transmitter and pick up resulting signals. This problem becomes worse in WANs.

TWO BASIC APPROACHES TO SECURITY

Link Encryption: Each vulnerable coomunication link is equipped on both ends with an encryption device. The main disadvantage is that it is effective only if all potential weak links from source to destination are secured.

End-to End Encryption: Data is encrypted only at the source node and decrypted at the destination node.

Problem: Data consists of packets. Packets have a header portion and a content portion. **You cannot encrypt the header!** (because it would be impossible to route the data). It follows that although user data is secure the traffic pattern is not!

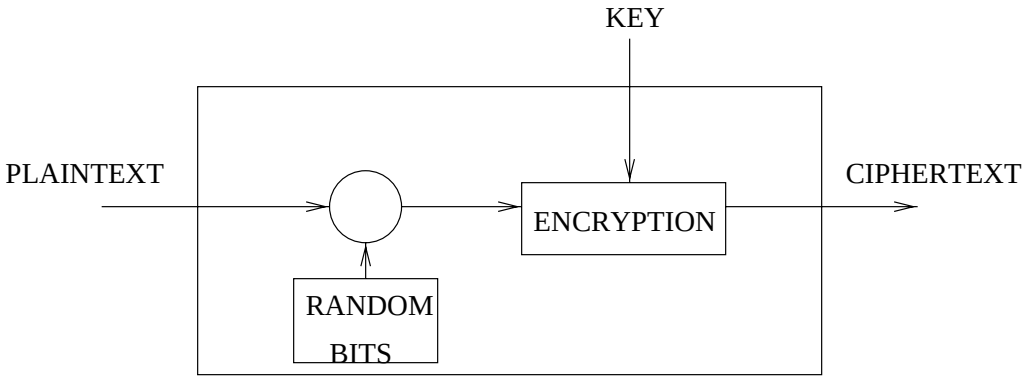
Solution: Use a combination of Link and End-to-End encryption.

PLACEMENT OF SECURITY FUNCTION

In the communication hierarchy, Link security is at a low level, while End-to-End security is high level.

Link encryption occurs at the physical or link layers of the Operarting System.

End-to-End encryption occurs at a Front End Processing unit and the header bypasses encryption in intermediate stages.



TRAFFIC SECURITY

It is usually necessary to conceal

- Identities of partners,
- How frequently two users communicate,
- Message patterns, e.g., length, quantity, time, etc.
- Events that correlate with special communications.

Link encryption conceals headers thus reducing the probability of effective traffic analysis. End-to-End encryption limits defence possibilities.

DISTRIBUTING KEYS

In conventional encryption a key must be shared by the two communicating users. Therefore any conventional cryptographic system is as good the method employed for distributing keys.

- A key can be delivered by one user to the other either directly (e.g., physically) or indirectly (e.g., physically by an intermediary).

- A new key can be delivered by encrypting it with an older key and either using a direct secure connection or an indirect secure connection via an intermediary.

The first option is awkward. Some form of the second option us widely accepted.

KEY CONTROL

Hierarchical: A hierarchy of Key Control Centers is established. Each center responsible locally for a small system. Control is passed to a higher level for external communication.

Key Lifetime: Same key is used only for a limited lifetime.

Decentralized Key Control: Full decentralization is not practical. However some form of decentralization limits abuses by a central authority.

Key Usage: It is useful to classify keys on the basis and type of usage. E.g., Data Encryption keys (for general communication), PIN keys (for Personal Identification Numbers), File keys (for encrypting files). This method limits potential damage caused by compromises in type of transmission.

PSEUDO RANDOM GENERATION

Random numbers find numerous uses in cryptography, especially in authentication schemes, session key generation, in conventional as well as public-key cryptography.

Perfect random generation is impossible by a deterministic device, like a computer. Usually we have to generate pseudorandom numbers with a deterministic source. Resulting numbers must be unpredictable, independent, and uniformly distributed.

Linear Congruence Generator:

$$x \rightarrow ax + b \text{ mod } m$$

Blum-Blum-Shub Generator $p \equiv q \equiv 3 \text{ mod } 4$ are distinct primes.

$$x \rightarrow x^2 \text{ mod } pq$$

VIOLATIONS OF SECURITY

As business and government depend more on computers and networks so grows the threat of commputer crime.

In 1994, V. L. Levin, a Russian computer hacker from St. Petersburgh, managed to inflitrate Citibank and transfer 10 million US dolars over five months to bank accounts in California, Finland, and Germany.

20 years ago computer systems were relatively unavailable. Now the taxonomy of users includes members of crime syndicates, industrial espionage teams, informations thieves, etc.

Only 5% of victim sites are even aware they have been inflitrated.

DETECTING ANOMALOUS PATTERNS

Detecting anomalies can be used for "enhancing" security.

Citibank will not reveal how Levin was caught. However, more is known about IBM's Fraud and Abuse Management System (FAMS).

FAMS separates billing patterns from unusual ones by profiling providers against one another and checking for unusual patterns that pointed to fraud in the past.

Blue Cross/Blue Shield caught a doctor who billed them 1.4 million dollars for bronchoscopies that were never performed. The program noticed that the doctor claimed to perform one operation per patient per week (normally this is performed once or twice in a patient's lifetime).

DETECTING ANOMALOUS PATTERNS

A typical computer user executes a standard pattern of commands.

For example, here is a sequence of commands I normally execute in my UNIX account

```
cd work;
ls -laF;
cd publications;
ls *.tex;
vi myfile.tex;
latex myfile.tex;
dvips myfile;
lpr -Pfaculty myfile.ps;
```

This sequence of commands could be recorded as part of a user's profile. Once created, an anomaly detector continuously compares it to the known profile to obtain a "similarity" score.

ADAPTING OVER TIME

A system can even be taught to adapt over-time. It learns the "usage" patterns of a user and adapts.

In turn, a malicious intruder can try to "fool" the system by teaching it to "accept" an "increasingly aggressive new" usage pattern.

This is of cource hypothetical, but hostile training is a danger.

Detecting anomalies can also be used for "breaking" security.

In 1996, P. Kircher demonstrated how to determine a private key by keeping track of how long takes the computer to decipher messages.

SECURITY IN PRACTICE: 3 METHODS

1. FIREWALLS:

Only certain computers are accessible to the general public (outside the company) forming a special "demilitarized zone" or DMZ.

Potentially dangerous data (e.g. internet, e-mail, etc) are filtered in a proxy server. These are then transfered to proxy programs that can run safely and subsequently delivered to company employees.

A large company or organization may require more than one firewall. As the company grows additional firewalls may need to be installed.

Firewalls also involve packet filtering, thus possibly rejecting packets coming from certain internet addresses. Intruders may of course try to forge trusted source addresses, hence authentication principles play an important role.

2. DIGITAL CERTIFICATES:

To send and receive messages users must have a private as well as a public key (strings of length about 1,000 bits).

Digital signatures are created from the message and the private key and accompany the message.

Signature is verified by using public key.

A trusted authority is being used to create a digital certificate that certifies that a certain public key belongs to a certain person.

3. JAVA SANDBOX:

Unscrupulous developpers could create applets that would interfere with a user's computer system.

Java has a layer of software (called Java Virtual Machine) which executes any applet written in the language.

The virtual machine prevents the program from getting access to the computer's hard drive.

It is like the applet sitting in a child's sandbox (where it can do no damage). It gets out only when the virtual machine verifies that the applet can be trusted.

DEFINITION OF CRYPTOSYSTEM

A cryptosystem consists of the following finite sets

P: plaintext space

C: ciphertext space

K: keyspace

Encryption Function E: For each $k \in K$,

$$E_k : P \to C$$

Decryption Function D: For each $k \in K$,

$$D_k : C \to P$$

Main Property: The functions E_k, D_k are inverses of each other, i.e. for all $p \in P$ and $k \in K$,

$$D_k(E_k(p)) = p.$$

IMPORTANT PROPERTIES

- the encryption and decryption functions are efficiently computable for all keys k, i.e., it should be relatively easy both to encrypt and decrypt, given the key, and

- it should computationally infeasible to decipher the ciphertext, i.e., an opponent upon seeing a ciphertext should be unable to determine either the key k that was used or the original plaintext string.

- Usually assume the cryptographic system is public, and only the key is secret information