COMPSCI 630 Systems **Spring 2017**

Lecture 11: Encryptions & RAID

Lecturer: Emery Berger *Scribe(s): Tsung-Yu Lin, Anirudha Desai*

11.1 Encryptions

A typical scenario used in encryption: Alice communicates with Bob through Internet. There are other people present in the middle of message passing: Eve sits and listens to the messages between Alice and Bob; Mallory, a malicious attacker, tries to modify and interrupt the messages. An approach to avoid the messages being acquired by Eve is using encryption. Messages are encrypted and keys are used to decode the messages.

11.1.1 Symmetric key encryption

Symmetric key encryption uses the same key as shared secret at both sides of encoding and decoding. Imagine Alice communicate plain text with Bob. They want to encode the text so that even Eve intercepts the message, he cannot read the encoded message.

Rot13 Cipher using Rot13 is an example of symmetric key encryption. The transformation simply encodes messages by shifting alphabets (Transposition cipher). Figure 11.1 (a) gives an example. There are only 26 possible mappings and thus it is easy for Eve to decode it by trying all possibilities. ENIGMA is an example of Cipher machine.

Substitution cipher Better than Rot13, substitution cipher encodes the text by replacing each alphabet with another without preserving the order. Figure 11.1 (b) gives an example. There are factorial of 26 possible such mappings and it is difficult to break by exhaustive search. However, given that the frequency of alphabets is not distributed uniformly, for example, ”e” is most common and ”z” is rare in English, people who intercept the message can decode the message by *frequency analysis* to recover the mapping. Additional information using bigram, for example, ”t” and ”o” very likely appear together, makes the frequency analysis easier.

One-time pad One-time pad cipher has position independent random substitution cipher for each character. When Eve observes the encoded text, each alphabets appear equally likely, so that the cipher is secure. But the drawback with this is that it does not scale well. For example, in the case of submarines where frequent communication is not possible, randomly choosing a cipher will not be feasible.

11.1.2 Asymmetric key encryption

Public key encryption use a pair of keys: public key known to everyone, and private key known only to the recipient of the message. Alice encodes the message using public key and Bod decodes the message using his private key. The asymmetric key approach relies on one-way functions where computing the inverse of the

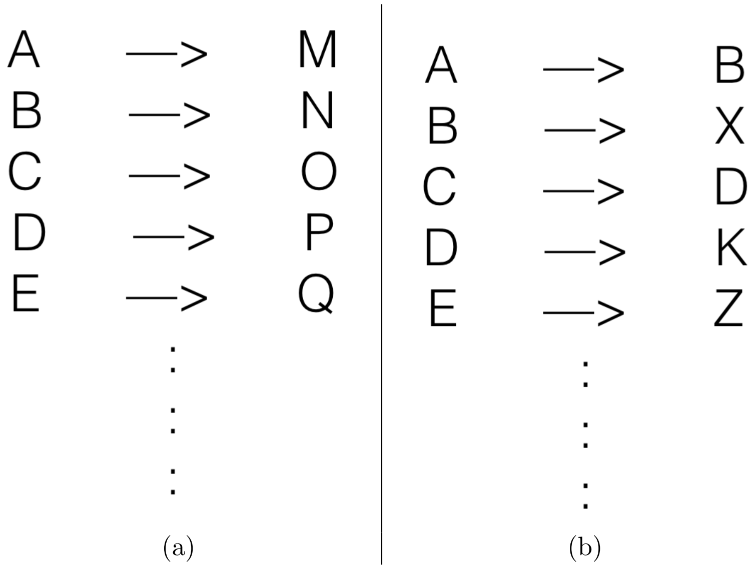

Figure 11.1: (a) Rot13 cipher (b) Substitution cipher

function is hard. The commonly used one-way function is the multiplication of two large prime numbers. The reverse direction is to compute prime factorization which is not easy to achieve for large numbers. Reference is made about Diffie-Helman key exchange algorithm for symmetric key exchange and about RSA Algorithm for asymmetric key exchange.

11.1.3 Dealing with mallory problem

To deal with mallory problem, a common approach is to create a separate document to verify the encoded document is not forged. The document is referred as digest which can be generated by a hash function, for example md5sum. The recipient can take the received message and compute the output of hash function to the digest. It is almost impossible to have forged document matched to the digest. Digest needs to be elsewhere otherwise Mallory can forge the digest.

11.2 RAID

11.2.1 Hard drive disk

The structure of HDD is shown as Figure 11.2. A head is attached on the arm reading the data from disk. The head can move along the arm inward or outward to read data from different tracks (seek latency) and the disk can rotate such that the sector with required data is under the head (rotation latency). Due to these two operations, the latency of HDD is high. To reduce the latency, file system tries to put files in consecutive blocks. Comparing solid state drive (SSD) to HDD, SSD has the same throughput but has no seek and rotation latency. Because there is no moving part in SSD, SSD generate less heat and is widely used nowadays.

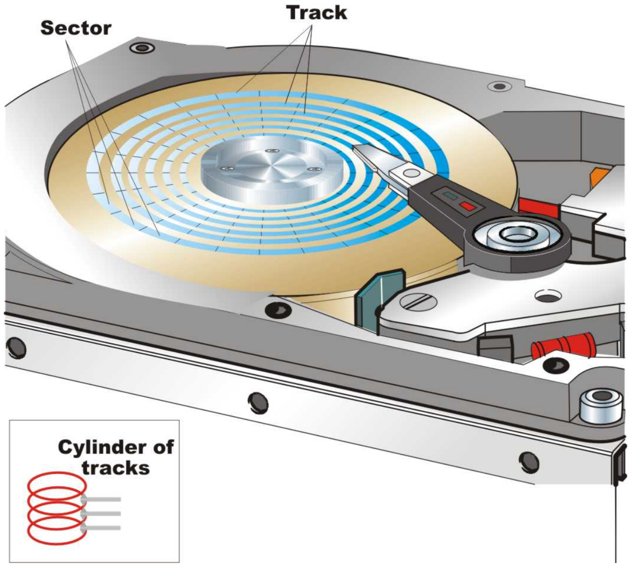

Figure 11.2: Example of a Hard drive disk

11.2.2 Redundant array of independent disks

RAID 0 It is also known as striping. It breaks up the data and separates them across several disk drives to achieve parallelism for speedup. Data can remain intact only when the data on all disks are not corrupted. This approach sacrifices the robustness to enhance the performance.

RAID 1 Mirroring: replicating the data to enhance the durability.

RAID 2 Using Hamming code for error correction.

Problem of RAID RAID is great until it fails. Generally, people are still using tapes for backup. To restore the data from failure, RAID needs to copy the whole disk form one to another. Copying a whole disk is almost 100% to fail. The is the reason why rebuilding RAID usually takes forever. On the other hand, the assumption of independent failure rate of disks is not valid. Because the array of disks are usually on the same racks, use the same power supplies, under same temperature, and purchased at the same time, if one of them fails, the others are likely to fail as well.