Introduction to Cryptography

Vivek Agarwal <me@vivek.im>

June 16, 2013

1 Introduction

Cryptography is everywhere. Its used for web traffic encryption (HTTPS), wireless (WPA2, GSM, Bluetooth), files on disk (EFS, TrueCrypt), content protection on DVD/Bue-ray (CSS, AACS), user authentication, etc.

In Cryptography, both confidentiality and integrity have to be preserved. So, no eavesdropping and no tampering.

Symmetric Encryption is the building block of data encryption. It is called symmetric because both the encryption and the decryption function use the same key. Encryption algorithm is (and should be) publicly known. Never use a proprietary cipher. Encryption algorithms in the public domain are tested by thousands of security experts and are much more safer to use than proprietary ones which are tested and designed by only a handful of people.

$E(k, m) \Rightarrow c$ and $D(k, c) \Rightarrow m$

where E and D are Encryption and Decryption functions, m is plain text, c is cipher-text and k is the secret key. In symmetric encryption, secret key is known only by the sending and the receiving parties.

The secret key can be used to either encrypt just one message and then change the key or it can be used to encrypt multiple messages.

Cryptography is-

- A tremendous tool
- The basis of many security mechanisms

Cryptography is not-

- The solution to all security problems

- Reliable unless implemented and used properly
- Something you should try to invent yourself

The two parts of core cryptography - secret key establishment and secure communication.

Application of cryptography-

- Digital signatures
- Anonymous communication
- Anonymous digital cash
- Elections
- Private auction

Secure Multi-Party computation

Suppose we are implementing an online voting system, we could have a trusted authority to whom each voter sends his encrypted vote and the authority keeps everything but the winning party's name secret. But what if the trusted authority is not longer trust-able? In cryptography, there is a theorem that says that anything that can be done with trusted authority can also be done without.

So, a protocol can be implemented such that every voter talks to one another anonymously and at the end of the protocol, desired results are outputted.

Zero Knowledge (proof of knowledge)

Suppose Alice knows $N = p * q$, where p and q are fairly large numbers so that getting p and q factors from knowing just about N is very difficult. Now, if Alice wants to prove to Bob that she knows the factors without letting Bob know about any of the details, she can do it with zero knowledge proof.

Three steps in Cryptography

- Precisely specify threat model
- Propose a construction
- Prove that breaking construction under threat model will solve an underlying hard problem

2 Examples of Symmetric Ciphers

2.1 Substitution Cipher

An alternate character mapping table is created which contains the alternate mappings for every character in the language. When the encryption is done, every character is replaced by its alternate version from the table and the reverse process is followed for decryption.

A popular example of substitution cipher is Caesar Cipher. In Ceaser Cipher, no key is used and the substitution is fixed, namely every character is shifted by say 3, so a becomes d, b becomes e and so on. This is very easy to break.

Breaking Substitution cipher

- Use frequency of English letters - using this, keys for some of the most common letters can be known
- Use the frequency of pair of letters (digrams) - using this, rest of the key table can be found out

This is the worst sort of encryption algorithm since given only the cipher text, attacker can get the plaintext.

2.2 Vigener Cipher

A key is selected, say 'CRYPTO', then the key is repeated to cover the entire length of the message to get the final key. Now the each corresponding letter from both the final key and the message are added (and modulo 26'ed) to form the ciphertext.

Message can be broken easily if the key size is known (and even if the key size is unknown, it can be determined). The ciphertext is broken into blocks of the size of key. Then we know that every first letter of each block was substituted by the same key letter and so on for every letter in the blocks, these blocks are transposed to form new set of blocks in which first block contains first letters from every previous block, second block contains second letters from every previous block and so on. Each of these new blocks are solved individually as substitution cipher and then transposed back to form the plaintext.

2.3 Rotor Machines

(The Herbern Machine) It's yet another form of substitution cipher. Rotor Machine is a type writer which contains the substitution table hardcoded in it in a rotor and every time a letter was pressed on the machine, the rotor rotated by one character. Again, this cipher was also broken using frequency tables.

Another rotor machine is The Enigma which is much more complex version of it containing multiple rotors and the key was initial setting of the rotors. It was used during WW2. But it was broken as well by British cryptographers.

2.4 Data Encryption Standard

In 1974, a group at IBM set together a cipher that was adopted to be used as a standard cipher by the government and was called Data Encryption Standard. Key space for DES is 2^{56} and block size is 64 bits which was good enough in 1974. These days, it can very easily be broken using bruteforce attacks.

Today, there are more advanced ciphers such as AES (2001), Salsa20 (2008) and others.

3 Discrete Probability

Reference - `https://en.wikibooks.org/wiki/High_School_Mathematics_Extensions/Discrete_Probability`

Advanced modern cryptography was developed as a rigorous science where constructions are always accompanied by a proof. The language used to describe the proofs relies on Discrete Probability.

Discrete probability is always defined over a universe which is denoted by U. This universe in our case is always going to be a finite set, very commonly $U = \{0,1\}^n$ a set of all n bit strings which here is denoted by $\{0,1\}^n$. The number of elements in this set is always 2^n.

So, $\{0,1\}^2 = \{00, 01, 10, 11\}$.

Probability distribution P over U is a function $P : U \to [0,1]$ (probability of every element in the set is between 0 and 1). The requirement for this to be true is that sum of all the probabilities be equal to 1.

3.1 Examples of Probability Distribution

Under **uniform distribution**, every element in the universe is assigned exactly the sam weight.

$\forall x \in U : P(x) = 1/|U|$ ($|U|$ means the size of universe/total num of elements)

Point distribution at $x[0] : P(x[0]) = 1, \forall x! = x[0] : P(x) = 0$. Here, all the weight is assigned to $x[0]$ and none to the remaining elements.

3.2 Events

$A \subset U : Pr[A] = \sum_{x \in A} P(x) \in [0, 1]$

Note that $Pr[U] = 1$. The set A here is called event.

3.3 Union Bound

$Pr[A_1 \bigcup A_2] \leq Pr[A_1] + Pr[A_2]$

And if $Pr[A_1 \bigcap A_2] = null$ then $Pr[A_1 \bigcup A_2] = Pr[A_1] + Pr[A_2]$

3.4 Random Variable

A random variable X is a function $X : U \to V$ (from the universe into some set V). Set V is where the random variable takes its value and also defines the distribution upon it.

Example: $X : \{0, 1\}^n \to \{0, 1\}; X(y) = lsb(y) \in 0, 1$

Suppose we have a random variable X, which maps the universe $\{0, 1\}^n$ into the set $\{0, 1\}$. So, the value of X is either 0 or 1. Given a particular n bit string sample y in the universe, random variable will just output the lsb (least significant bit) y.

For the uniform distribution on U: $Pr[X = 0] = 1/2; Pr[X = 1] = 1/2$

3.5 Uniform Random Variable

Let U be some set, e.g. $U = 0, 1^n$ then

$r \xleftarrow{R} U$ denotes a uniform random variable r over U

$$\forall a \in U : Pr[r = a] = 1/sizeof(U)$$

Formally, r is the identity function: $r(x) = x; \forall x \in U$

3.6 Deterministic Algorithm

$y \leftarrow A(m)$

3.7 Randomized Algorithm

It takes input m and has implicit argument r, where r is sampled anewed every time the function is run. n is sampled randomly from set of n bit strings.

$y \leftarrow A(m;r)$ *where* $r \xleftarrow{R} \{0,1\}^n$

3.8 Independent Events

Events A and B are independent if $Pr[A \wedge B] = Pr[A] \cdot Pr[B]$

The occurence of event A tells nothing about B and vice-versa. Similarly, random variables X, Y taking value in V are independent if

$\forall A, B \in V : Pr[X = a \wedge Y = b] = Pr[X = a] \cdot Pr[Y = b]$

4 XOR

XOR of two binary digits is their sum modulo 2.

An important property of XOR

Let Y be a random variable over $\{0,1\}^n$ (distribution maybe non-uniform) and X an independent uniform variable over $\{0,1\}^n$

Then $Z := Y \oplus X$ is a uniform variable over $\{0,1\}^n$

So, if an arbitrarily malicious distribution is taken and XOR'ed with an independent uniform random variable, then the result is uniform random variable. This property is very useful for cryptography.

5 The Birthday Paradox

Let $r_1, r_2, ..., r_n \in U$ be independent identically distributed random variables.

When $n = 1.2 \times |U|^{\frac{1}{2}}$ then $Pr[r_i = r_j | i! = j] \geq \frac{1}{2}$

So, when n number of samples are taken from universe U, then there is a good probability that that two of them are equal.

Example: There are 365 days in a year so if $n = 1.2 \times sqrt(365) = 24$ number of people are taken from a random sample, the probability that two random people from the sample have the same birthday is equal to or more than $\frac{1}{2}$. 24 seems to be such a small number yet this phenomenon is observed.