CSA E0 235: Cryptography January 11, 2016

Scribe for Lecture 2

Instructor: Arpita Patra *Submitted by: Jayam Modi*

1 Discrete Probability Background

- **Probability Distribution** -A probability distribution over a finite set U is a function $Pr : U \rightarrow [0, 1]$ such that $\sum_{x\ in\ U} Pr(x) = 1$. E.g. In a uniform probability distribution on U, $Pr(x) = \frac{1}{|U|}\ \forall x$.
- **Event** - An occurrence of one or more elements of U is called an event.
- **Union Bound** - $Pr[A_1 \cup A_2] \leq Pr[A_1] + Pr[A_2]$ (valid for more than two events also).
- **Conditional Probability** - It gives the probability that one event occurs given that some other event has occurred. $Pr[A|B] = Pr[A \cap B]/Pr[B]$. If A and B are independent, $Pr[A|B] = Pr[A]$ and hence, $Pr[A \cap B] = Pr[A] \times Pr[B]$.
- **Law of Total Probability** - If $E_1, \ldots, E_n$ are a partition of all possibilities of events, then $Pr[A] = \sum_i Pr[A \cap E_i] = \sum_i Pr[A|E_i] \times Pr[E_i]$.
- **Bayes Theorem** - $Pr[A|B] = (Pr[B|A] \times Pr[A])/Pr[B]$
- **Random Variable** - It is a variable that takes on discrete values from a finite set with certain probabilities.
- **Probability Distribution of a Random Variable** - It specifies the probabilities with which a random variable takes the possible values of a finite set. All probabilities must be between 0 and 1 & all probabilities such sum up to 1.

2 Secure Communication in Private Key Setting

Modern cryptography started evolving from 1980's. The blunders made by classical approach to cryptography led to significant improvements in the modern era.

Secure communication in modern cryptography is done using a 3 step approach. The steps are as follows -

- **Formulate a formal definition** - A formal definition of a crypto-system is essential in its efficient and secure design. The definition must clearly indicate what we're trying to protect, what type of violations are considered as break of the system, what are the threats to the system and how much power does the threat/adversary have.

- **Identify assumptions and build a construction** - The construction of crypto-systems often relies on several number theoretic assumptions or conjectures. Their choice affects the security of the system. They need to be identified before the actual construction of the crypto-system.
- **Prove the construction** - Lastly, we need to formally prove that the crypto-system that has been constructed is secure so that we can give a guarantee for the security of the system.

2.1 Syntax of Secret/Private Key Encryption (SKE)

SKE is used for communication between two parties over an untrusted channel. Since the channel is untrusted, transferring data as plain text is not safe since the adversary can be eavesdropping on the channel. The encryption technique is called *private* since the keys are known only to the sender and receiver as opposed to public key encryption where a component of the key is known to everyone. It is called *symmetric* since the same key is used for encryption as well as decryption.

The diagram 1 depicts various entities and symbols in an SKE scheme.

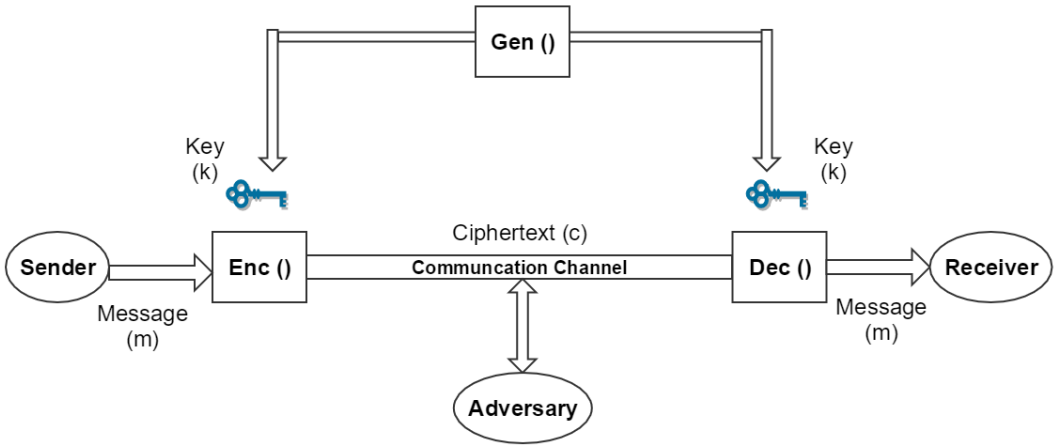

Figure 1: Secret Key Encryption

Definition 1 An SKE scheme is a collection of the following 3 algorithms [2]:

- Key Generation Algorithm (**Gen()**) - This algorithm generates the secret key k used for communication. This key is chosen according to some probability distribution that is pre-defined in the system. Gen() is a randomized algorithm.

  A point to be noted here is that according to Kerchoff's principle, the security of the system should hold even when all the algorithms are public. If Gen() is a deterministic algorithm, then the adversary will know which key will be output by Gen() and hence no security can be achieved.

- Encryption Algorithm (**Enc()**) - This algorithm takes as input a plaintext message m & a key k and outputs a ciphertext c. This algorithm can be deterministic or randomized. When it is randomized, it may output different ciphertexts for the same inputs. This randomness is indicated by writing $c \leftarrow \mathsf{Enc}_k(m)$. On the other hand, when the algorithm is deterministic, it will output the same ciphertext every-time for the same combination of the inputs (message and key). This is indicated by writing $c := \mathsf{Enc}_k(m)$.
- Decryption Algorithm (**Dec()**) - This algorithm takes as input a ciphertext c & a key k and outputs a plaintext message m. If $\mathsf{Dec}_k(c) = m$ with probability 1 for a ciphertext c generated by $\mathsf{Enc}_k(m)$, then the scheme is called *perfectly correct*. Almost all the schemes studied and used are perfectly correct.

$\diamondsuit$

Corresponding to the three algorithms, 3 spaces are generated:

- Key Space ($\mathcal{K}$) - This is the set of all keys generated by the Gen() algorithm.
- Plaintext/Message Space ($\mathcal{M}$) - This is the set of all valid messages that can be taken as input by the Enc() algorithm.
- Ciphertext Space ($\mathcal{C}$) - This is the set of all ciphertexts output by the Enc() algorithm.

Thus, any SKE is specified by using the triplet (Gen, Enc, Dec) and $\mathcal{M}$. This is because, $\mathcal{M}$ does not depend on the triplet while $\mathcal{K}$ and $\mathcal{C}$ are defined by output of Gen and Enc.

The probability distribution of $\mathcal{M}$ depends on external factors while that of $\mathcal{K}$ depends on the Gen() algorithm. Thus, these two are independent to each other. But the probability distribution of $\mathcal{C}$ depends on that of $\mathcal{M}$ and $\mathcal{K}$. Also, all the distributions are known to the adversary.

2.2 Threat and Break Model

To formulate a formal definition of SKE, we need to identify the threats to the system and specify what will be considered as break of the system.

2.2.1 Threat Model

Three major aspects are considered in reference to an adversary-

Computing Power

- No assumption on the computing power of the adversary.
- He can solve any hard problem (like factorization or discrete logarithms) in no time and hence becomes the strongest adversary in terms of computing power.
- A scheme secure against an attack by such a powerful adversary is a very strong security scheme

Power of Randomness

- Our encryption algorithm can be randomized or deterministic.
- If the adversary does not have the power to toss a coin, i.e. does not have the power of randomness, then he will be at a disadvantage.
- Thus, giving the power of randomness to the adversary is essential to build stronger encryption schemes.

Access Capabilities

- What kind of attacks can be launched by the adversary ?
- What components of our communication Protocol are accessible to him ?
- In a *Ciphertext only Attack (COA)*, the adversary can only eavesdrop/tap the ciphertext during its transmission. This is a passive attack.
- There are other more stronger attacks like Chosen Plaintext Attack and Known Ciphertext Attack.

2.2.2 Break Model

The break of a system is defined in terms of what we're trying to protect.

- If we try to protect only the secret key and nothing else, then a scheme that has the property $\mathsf{Enc}(m) = m$ is secure but leaks out the entire message.
- If we don't try to protect the entire message, then a scheme leaking most significant 10 bits is secure but transferring m : *bank password* | *amazon password* is secure but may leak out the entire bank password ($<= 10$ bits).
- If the adversary gains no additional information about the message except that which was known prior to the attack, then this notion captures perfect security.

According to this model, the adversary gains nothing by tapping over the communication channel. He might as well go and enjoy his life by playing games or watching cricket.

2.3 Formal Definition of Perfectly Secure Encryption

The threat and break model specifies clearly the nature and power of adversary as well as what we want to protect. Mathematically,

$Pr[M = m]$ captures the prior information of the attacker about m since the probability distribution of $\mathcal{M}$ is well known. It is the **a-priori** probability that m might be communicated.	The attacker's knowledge after looking at the ciphertext is $Pr[M = m \mid C = c]$. This is the **posterior** probability that m is encrypted in c.

Definition 2 An encryption scheme (Gen, Enc, Dec) over a plaintext space $\mathcal{M}$ is *perfectly secure* if for every probability distribution over $\mathcal{M}$, every plaintext $m \in \mathcal{M}$ and every ciphertext $c \in \mathcal{C}$, the a-priori probability should be equal to the posterior probability, i.e.

$$Pr[M = m \mid C = c] \;=\; Pr[M = m]$$

♢

This was the first formal definition of security given by Claude Shannon in 1949 [1].

3 Vernam Cipher (One-time pad)

This cipher was patented by Vernam in 1917 even before Shannon defined perfect security. Later on, it was proved by Shannon that Vernam's cipher was perfectly secure. It works as shown in fig 2.

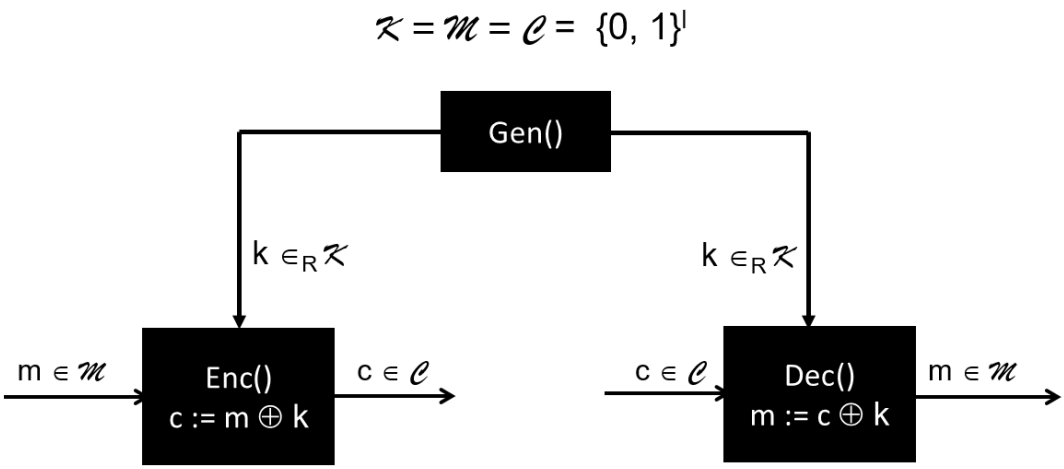

Figure 2: Vernam Cipher

This correctness of this scheme can be proved using the fact that $\mathsf{Dec}(\mathsf{Enc}(m)) = \mathsf{Dec}(k \oplus m) = k \oplus k \oplus m = m$.

3.1 Proof of Perfect Security

Intuitively, in the Vernam cipher, for each message m and ciphertext $c = \mathsf{Enc}(m)$, we can find a unique key $k = c \oplus m$ such that $k \in \mathcal{K}$. Since each key k is chosen uniformly at random from $\mathcal{K}$ and each message m is equally likely to be encrypted, the adversary can know nothing about the plaintext after looking at ciphertext c.

Theorem 1 *Vernam Cipher is Perfectly Secure [2]*

Proof *To prove this theorem, according to Shannon's definition, we need to show that* $Pr[M = m \mid C = c] = Pr[M = m]$.

Now, for any arbitrary c *and* m,

$$\begin{aligned} Pr[C = c \mid m = M] &= Pr[M \oplus K = c \mid M = m] \\ &= Pr[m \oplus K = c] \\ &= Pr[K = c \oplus m] \\ &= 1/2^l \end{aligned}$$

Also, by the law of total probability,

$$\begin{aligned} Pr[C = c] &= \sum_{m \in \mathcal{M}} Pr[C = c \mid M = m] \times Pr[M = m] \\ &= \sum_{m \in \mathcal{M}} (1/2^l) \times Pr[M = m] \\ &= (1/2^l) \times \sum_{m \in \mathcal{M}} Pr[M = m] \\ &= (1/2^l) \qquad (\textit{the sum is 1 as per the definition of a probability distribution}) \end{aligned}$$

Thus, by Bayes theorem,

$$\begin{aligned} Pr[M = m \mid C = c] &= \frac{Pr[C = c \mid M = m] \times Pr[M = m]}{Pr[C = c]} \\ &= \frac{(1/2^l) \times Pr[M = m]}{(1/2^l)} \\ &= Pr[M = m] \end{aligned}$$

Hence proved. ■

3.2 Drawbacks

Though Vernam cipher is perfectly secure, it has two major drawbacks -

- The key length is equal to the message length. This means that for longer messages, we need to generate longer keys and longer keys are hard to store securely. Also, this implies that we cannot send a message of arbitrary length before exchanging a corresponding key of similar length.

- Re-use of keys for multiple messages of same length is not allowed (thus giving this scheme the name One-Time-Pad). This is because, if the same key k is used to encrypt two messages say m_1 & m_2 to give two different ciphertexts c_1 & c_2 and if the adversary gets hold of both ciphertexts, then he can know the value of $m_1 \oplus m_2$ simply by doing $c_1 \oplus c_2$. This is because

  $$c_1 \oplus c_2 = (m_1 \oplus k) \oplus (m_2 \oplus k) = m_1 \oplus m_2$$

  This means that the adversary gains some knowledge about the messages after looking at the ciphertexts and this violates Shannon's definition of Perfect Security. If the messages are only of a particular language whose frequency distribution is known, then multiple ciphertexts can be analyzed to recover the messages.

- This scheme is secure only against the Ciphertext Only Attack (COA). In case of a known message attack, the adversary who knows a message m and gets hold of its ciphertext c can easily compute the key $k = c \oplus m$ and then decrypt all messages encrypted using this key.

3.3 Historical Importance

In 1923, a secure communication technique using the OTP scheme was developed in Germany. Duplicate paper pads were printed with lines of random number groups with a serial number on each page. There were eight lines with six 5-digit numbers on each line. A page would be used as a work sheet to encode a message and then destroyed. The serial number of the page would be sent with the encoded message. The recipient would reverse the procedure and then destroy his copy of the page. This enabled them to achieve perfectly secure encryption.

In 1963, a Moscow Washington Direct Communications Link (famously known as Hotline or Redline) was established for direct communication between the leaders of US and USSR. It used the teletype technology for communication. The teletype messages were encrypted by the unbreakable OTP scheme. The keypads were exchanged via their embassies in the other country.

References

[1] Shannon, Claude E. *"Communication theory of secrecy systems*."* Bell system technical journal 28.4 (1949): 656-715.

[2] Katz, Jonathan, and Yehuda Lindell. *Introduction to modern cryptography.* CRC Press, 2014.

[3] Arpita Patra. *http://drona.csa.iisc.ernet.in/∼arpita/Cryptography16.html.* Course Materials.