CSA E0 235: Cryptography (17th January 2016)

Lecture 2

Instructor: Arpita Patra *Submitted by: Gaurav Sharma*

1 Introduction

The period of cryptography after 1980 is considered to be as Modern Cryptography era because during this period cryptographers started thinking on different lines from those that were basis of classical cryptography.Whereas the classical approach was more focussed on art than on science the modern cryptography has evolved more into science. In the late 19th century Auguste Kerckhoff argued against several design principles of classical cryptography. He gave several arguments such as "cipher method must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience".Thus Modern cryptography involved into a subject that involves scientific study and development of crypto schemes using mathematical constructions to secure digital information,transactions and distributed computations against adversarial attacks.The schemes are now being developed in a more systematic manner with an ultimate aim being to give a rigorous proof that validates that the given scheme is secure. After learning from the blunders of the classical approach we would now focus on carrying out secure communication in a modern way. We would thus now be formulating a formal definition by identifying threat and break model,Identifying different assumptions, constructing a scheme and then in last proving it to be secure. The emphasis given to definitions, assumptions and proofs differentiates the modern cryptography from the classical one. we would now study these three principles in more details.

2 Secure Communication in Private Key Settings

One of the fundamental issue in field of cryptography is secure message communication in which two parties are safely communicating with each other using an untrusted channel between them. An adversary sitting on the channel has the capability to tap the messages and has unbounded computational powers.In the symmetric key setting, we consider the following assumptions :-

- That the secret key is a-priorly exchanged between the sender and receiver and is private between sender and the receiver(**Private(secret) Key Encryption**).
- That the same key(**Symmetric Key Encryption**) is used for both encryption and decryption.
- 'm' is the plain text which is being exchanged.
- 'c' is the cipher text (scrambled message or encoded message)

2.1 Syntax of Secret Key Encryption

Formally, a symmetric key encryption scheme is a tuple of probabilistic polynomial time algorithms(Gen,Enc,Dec) as well as a specification of finite message space 'M'. These algorithms are further defined as in subsequent paragraphs.

2.1.1 Key Generation Algorithim Gen()

The key generation algorithm Gen takes as an input 1^n and outputs a key 'k' according to some distribution.We denote by 'K' the finite key space, i.e the set of all possible keys that can be output by Gen
Note : Gen() is a randomized algorithm.

2.1.2 Encryption Algorithm Enc()

The encryption algorithm Enc takes as an input a key $k \in K$ and a plain text message $m \in M$ and outputs a cipher text 'c'.We denote it by$Enc_k(m)$ i.e the encryption of the plain text message m using the key k.
Note :$Enc_k(m)$ is a deterministic/randomized algorithm.

2.1.3 Decryption Algorithm Dec()

The decryption algorithm Dec takes as input a key 'k' and a cipher text 'c' and outputs a message 'm'.We denote it by $Dec_k(c)$ i.e the decryption of the cipher text message c using the key k.
Note :$Dec_k(c)$ We assume that Dec is deterministic and outputs m i.e the original plain text message. $m = Dec_k(c)$
The following three spaces are associated with a crypto-system:

- The message space M, from which the messages come from. While designing a crypto system, we generally try to make no or minimal assumptions about the distribution of M.
- The key space K, which is fully determined by Gen.
- The cipher text space C, which is determined by K,M and Enc.

2.2 Formal Definition of Security

The formal definition of security is essential for the proper design,study,evaluation and usage of cryptographic primitives.It is very important as it gives clear description of the scope of threats and thus what security guarantees are desired from a particular scheme.This not only helps in better design of the crypto schemes but also throws more clarity on requirements right at the design stage.This further helps the designers to avoid any major post facto design changes once the design is freezed.If we consider evaluation and analysis aspects then too clear formal definition certainly helps in better evaluation and analysis of the proposed scheme against the desired guarantees.In certain cases with formal definition itself we can prove whether the given construction is secure or not by demonstrating that

it meets the definition.
Using the formal definition we can do meaning full comparisons of different possible schemes and can compare them as per the intended context of its use.While designing we can thus analyse different schemes and can choose a weaker scheme against a stronger definition while considering several trade-off's.we can thus choose the best possible scheme.
Any definition of the security consists of two distinct components: a threat model(i.e the specification of the assumed power of the adversary) and a security guarantee(usually specified by describing what constitutes a 'break model' of the scheme).
The **threat** can further be elaborated as:

- Identifying the potential threats i.e who is our threat from which you want to protect yourself from.
- cultivating your adversary.It means knowing more about him .
- look out at the practical scenarios and think like an adversary would think to break the model.

Similarly the Break can also be further elaborated as:

- what you are afraid of loosing. i.e what is the important or vital for you.Its identification is important as the break model would depend on it.
- What do you want to protect . i.e what do you want to protect most. i.e you should prioritize your information.
- When you don't know what to protect then it would be difficult to know whether you are protecting it or not.
- There would be different methods to to extract different types of information. Analysis of this is thus also important

Based on these aspects only we design our threat and break models.

2.2.1 Threat Model

The threat model thus considers various aspects such as the computational power of the adversary.It is always safe to have no assumption about the computational power and to assume that adversary is having unbounded computational power.We assume it to have such computational powers that , given any hard problem he will solve it in no time.The kinds of attacks that we are considering are of passive sniffing i.e eavesdropping on the packets during their transit. That is only basic Cipher-text-only attack(**COA**)are possible. Thus our threat model becomes

The Threat Model

- Randomized
- Unbounded Powerful.
- COA.

2.2.2 Break Model

The break model is based on the what the secure encryption scheme should guarantee to us.The different things that we consider a given encryption scheme should guarantee can be thus :
Definition 1 :Secret Key ,It should be impossible for attacker to recover key : We know that if the attacker gets to know about the key shared between two parties then that scheme is no longer secure.Consider , e.g , the scheme where $Enc_k(m) = m$. The ciphertext leaks no information about the key yet the message sent in the clear. thus inability to recover the key is not sufficient for security.
Definition 2: Entire Message should be be impossible to recover the message : This definition is better but this definition would consider an encryption scheme secure if its cipher text revealed 90% of plaintext as long as 10% remains hard to figure out.This is also not sufficient considering cases of bank passwords.
Definition 3: No additional information to be leaked: . This definition states that regardless of any information an attacker already has cipher text should leak no additional information about the plain text.This definition seems to be correct. Now to formulate the precise formulation of this definition we need to refresh up our memory of discrete probability theory before we formalise the definition

Basics of Discrete Probability Theory

- U is a finite set, e.g.$\{0,1\}$

- **Probability distribution**: Probability distribution Pr over U is a function $Pr : U \longrightarrow [0,1] \quad such \quad that \quad \sum_{x \quad in \quad U} Pr(x) = 1$

- **Event:** Occurrence of one or more elements of U is called an event
  - e.g Consider Uniform Distribution on$U = \{0,1\}^4$
  - Let A = occurrence of elements of U with msb two bits as 01
  - Pr(A) = $\frac{1}{4}$

- **Union Bound:** For events A1 and A2

  $Pr\left[A1 \cup A2\right] \quad \leq \quad Pr\left[A1\right] + Pr\left[A2\right]$ (extend for more than 2)

- **Conditional probability:** probability that one event occurs, assuming some other event occurred
  - $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$
  - For independent $A, B : Pr(A|B) = Pr(A) \quad and \quad Pr(A \cap B) = Pr(A).Pr(B)$

- **Law of total probability:** Let$E1, \cdot\cdot, En$ are a partition of all possibilities of events. Then for any event A:
  $Pr[A] = \sum_i Pr[A \cap E_i] = \sum_i Pr[A|E_i] \cdot Pr[E_i]$

- **Bayess Theorem:** $If Pr(B) \neq 0$ then
  $Pr(A|B) = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)}$

- **Random Variable:** variable that takes on (discrete) values from a finite set with certain probabilities (defined with respect to a finite set). We can also define it as a real valued function defined on the sample space.

- **Probability distribution for a random variable:** specifies the probabilities with which the variable takes on each possible value of a finite set
  - Each probability must be between 0 and 1
  - The probabilities must sum to 1

If our aim is to protect the secret key then it is safe as encrypted message does not reveals the key and key is safe.Thus $Enc_k(m)$ should not reveal the message. This is acceptable in case of short messages but in the case of long message the encryption with such a large key might not be possible.Even if we think of protecting only a part of message then it is not a good approach considering bank passwords in which even part of numbers should also not be revealed.Thus the break model should ensure that No additional information should be revealed irrespective to the prior information.Thus the break model becomes

The Break Model

- Thus to conclude our break model we can say that:No additional information about the message should be leaked from the cipher text irrespective of the prior information that the adversary has.

2.3 Identifying Assumptions

Most of the modern cryptographic constructions cannot be proven secure in an unconditional manner.Security often depends on some widely believed(although unproven) assumptions. Modern cryptographic assumptions states that any such assumptions must be clearly stated and unambiguously defined.Here it is to be noted that these assumptions should be chosen with care and popular well studied conjectures(ex : well known conjectures in number theory) should always be favoured over the adhoc ones.Trade-offs discussed as in the case of formal definition apply here as well: too weak an assumption thus may hamper the efficiency of the system, while assumptions that are too strong involves the risk of being eventually proven wrong.

2.4 Proof of Security

The principles discussed above equip us to achieve our goal of providing a rigorous proof that a given construction satisfies a given definition with some assumptions in place.These proofs become important when we also consider an active attacker who is constantly trying to break the given scheme.Thus proof of security gives us certain guarantee that the scheme is secure.Without a proper proof that no adversary with the specified resources can break some scheme we are only left with an intuition of our scheme being secure.This can be disastrous.

2.5 Threat and Break Model

The threat and break models that we are thus considering for our encryption scheme is

The Threat and Break Model

- **Threat Model**
- -Randomized. can sample randomly.
- -Unbounded powerful adversary
- -cipher text only attacks are possible
- **Break Model**
- -No additional information about the message should be leaked from the cipher-text irrespective of the prior information that the adversary has.
- **Note** In mathematical terms we want to ensure that probability distribution of knowing the plaint text should remain same as that when the cipher text c' has been observed by the adversary.
  i.e $Pr[M = m|C = c] = Pr[M = m]$

2.6 Formal Definition of Security

2.7 Perfectly Secure Encryption : Formal Definition

We can now define the notion of prefect secrecy. We imagine that adversary knows the probability distribution over M, that is the likelihood that different messages will be sent.He also knows the encryption scheme too. The only thing he is not privy to is the key $'k'$ shared between the parties.The adversary has the capacity to eavesdrop on the cipher-text-only attack, where the attacker gets only the cipher-text.Now for this scheme to be perfectly secure,observing the cipher-text should have no effect on the adversary's knowledge regarding the actual message that was sent.In other words the posteriori probability that some message $m \in M$ was sent,conditioned on the cipher-text that was observed should be no different from the a-priori probability that m would be sent. That is by eaves dropping on the cipher text the adversary learns nothing about the plain-text that had been encrypted.Thus formally we can define it as :

Definition An encryption scheme (Gen, Enc, Dec) with message space$M$ is perfectly secret if for every probability distribution over M,every message $m \in M$ and every cipher-text $c \in C$ for which $Pr[C = c] > 0$:

$$Pr[M = m|C = c] = Pr[M = m]$$

where
$Pr[M = m|C = c]$ is the posteriori probability that m is encrypted to c.
$Pr[M = m]$ is the a-priori probability that m might be communicated.

2.8 Perfectly Secure Encryption : Construction

$M = K = C = 0, 1$

The Gen Algorithm

- Gen outputs key $k \in_R K$

The Enc Algorithm

- Enc takes input plain-text message $m \in M$ and key k
- Enc outputs $c = m \oplus$k

The Dec Algorithm

- Dec takes input cipher-text message $c \in C$ and key k
- Dec outputs $m = c \oplus$k

Theorem 1:Vernam Cipher is perfectly Secure
To prove that Vernam Cipher is perfectly secure we need to prove that
$Pr[M = m|C = c] = Pr[M = m]$
We know that for arbitrary c and m, $Pr[C = c|M = m] = Pr[K = c \oplus m] = 1/2^l$
$Pr[C = c] = \Sigma_{m \quad in \ M} Pr[C = c|M = m]Pr[M = m] = \Sigma Pr[M = m]1/2^l$
We also know that $Pr[C = c] == 1/2^l$
substituting these in the Bayes theorem we get
$Pr[M = m|C = c] = \frac{Pr[C=c|M=m]Pr[M=m]}{Pr[C=c]}$
$= \frac{Pr[M=m]1/2^l}{1/2^l} = Pr[M = m]$

2.9 Problems with Vernam Cipher

Vernam Cipher also known as **One Time Pad(OTP)** is an mathematically proven unbreakable encryption algorithim as roved above but it does have certain issues.These are :

(Vernam Cipher Issues)

- Vernam Cipher works good when the key length is small but when the key length is large in case of long messages it is often difficult to agree on the keys among themselves. It would also be difficult to agree upon the keys when the message size is not known in the advance.
- In Vernam Cipher or OTP ,it is not possible to use the same key for multiple message and key unique ness among the multiple messages is kind of a pre-requisite for secure communication. Consider the case when the same key is being used to encrypt two messages m and m'. Here the corresponding cipher-text generated would be say c and$c'(c = m \oplus k \quad \text{and} \quad c' = m' \oplus k)$.Here if the adversary just take two cipher -text message c and c' from the channel and just do a XOR operation then can get the additional information as the difference between the cipher text message and original plain-text message is same$(c \oplus c' \quad = \quad m \oplus m')$. Thus the definition of the perfect security gets compromised here as the adversary gains some additional information by working on the multiple cipher-text message encoded using the same key

Note : In Vernam Cipher key -re usability is not permitted. This issue is independent of the encoding scheme we use and is inherent to any scheme.

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.