A SURVEY ON THE PROVABLE SECURITY USING INDISTINGUISHABILITY
NOTION ON CRYPTOGRAPHIC ENCRYPTION SCHEMES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EMRE AYAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CRYPTOGRAPHY

FEBRUARY 2018

Approval of the thesis:

A SURVEY ON THE PROVABLE SECURITY USING INDISTINGUISHABILITY NOTION ON CRYPTOGRAPHIC ENCRYPTION SCHEMES

submitted by **EMRE AYAR** in partial fulfillment of the requirements for the degree of **Master of Science in Department of Cryptography, Middle East Technical University** by,

Prof. Dr. Ömür Uğur
Director, Graduate School of **Applied Mathematics** _____

Prof. Dr. Ferruh Özbudak
Head of Department, **Cryptography** _____

Assoc. Prof. Dr. Ali Doğanaksoy
Supervisor, **Cryptography, METU** _____

Dr. Onur Koçak
Co-supervisor, **TÜBİTAK - UEKAE, İstanbul** _____

Examining Committee Members:

Assoc. Prof. Dr. Murat Cenk
Cryptography, METU _____

Assoc. Prof. Dr. Ali Doğanaksoy
Department of Mathematics, METU _____

Assist. Prof. Dr. Fatih Sulak
Department of Mathematics, Atılım University _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: EMRE AYAR

Signature   :

ABSTRACT

A SURVEY ON THE PROVABLE SECURITY USING INDISTINGUISHABILITY NOTION ON CRYPTOGRAPHIC ENCRYPTION SCHEMES

Ayar, Emre

M.S., Department of Cryptography

Supervisor : Assoc. Prof. Dr. Ali Doğanaksoy

Co-Supervisor : Dr. Onur Koçak

February 2018, 44 pages

For an encryption scheme, instead of Shannon's perfect security definition, Goldwasser and Micali defined a realistic provable security called semantic security. Using indistinguishability notion, one can define security levels according to the polynomial time adversaries' capabilities such as chosen plaintext attacks (CPA) and chosen ciphertext attacks (CCA) for both symmetric and asymmetric encryption schemes in addition to the hard mathematical problems the algorithms based on. Precautions to prevent these attacks, however, differ for symmetric and asymmetric schemes in some aspects. In a symmetric encryption scheme, designer needs to impose a secure mode of operation to the cipher whereas in asymmetric encryption schemes padding and hash- based structures are used to provide security. In this thesis, we first give the descriptions of CPA and CCA security under indistinguishability notion for both symmetric and asymmetric encryption schemes. Then we analyse the security of widely used algorithms with respect to these security models.

Keywords : Indistinguishability, CPA security, CCA security

ÖZ

ŞİFRELEME ALGORİTMALARINDA AYIRDEDİLEMEZLİK KAVRAMI KULLANILARAK GÜVENLİK TANIMI

Ayar, Emre

Yüksek Lisans, Kriptografi Bölümü

Tez Yöneticisi   : Doç. Dr. Ali Doğanaksoy

Ortak Tez Yöneticisi   : Dr. Onur Koçak

Şubat 2018, 44 sayfa

Bir şifreleme sistemi için, Shannon'un mükemmel güvenlik tanımı yerine, Goldwasser ve Micali gerçekçi senaryolara dayanan semantik güvenliği tanımlamıştır. Ayırdedilemezlik kavramını kullanarak, saldırganın yeteneklerine bağlı, seçili mesaj saldırıları ve seçili şifreli mesaj saldırıları olarak farklı güvenlik seviyeleri tanımlanabilinmektedir. Bu tanımlamalar, kullanılan algoritmaların dayandığı matematiksel zorluktaki problemlere ek olarak tanımlanmıştır. Gizli-anahtar ve açık-anahtar şifreleme sistemlerinde bu tanımlar bazı değişiklikler içermektedir. Gizli anahtar şifreleme sistemlerinde tasarımcı güvenli blok şifreleme modlarını kullanabilirken, açık-anahtar şifreleme sistemlerinde dolgu algoritmaları ve özüt fonksiyon tabanlı dönüşümler kullanılabilmektedir. Bu tezde, öncelikle güvenlik seviyelerinin tanımları yapılmış olup, daha sonra sık kullanılmış ve standart haline gelmiş şifreleme sistemlerinin güvenlik analizleri ele alınmıştır.

Anahtar Kelimeler : Ayırdedilemezlik, IND-CPA, IND-CCA, şifreleme algoritmaları için ispatlanabilir güvenlik

To My Family

ACKNOWLEDGMENTS

TABLE OF CONTENTS

xvii

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Cryptography is a science that contributes to the information security. It aims to provide secrecy, integrity, authentication and non-repudiation to secure information between parties. Regarding this, it benefits cryptographic primitives like encryption and decryption algorithms that are constructed by using mathematical functions. An encryption scheme is defined with the key pair (e, d), encryption function Enc, decryption function Dec. One uses Enc and the encryption key e to produce the corresponding ciphertext c for the message m, $Enc(e, m) = c$. Upon receiving the ciphertext c, the original message m can be generated using the decryption key d and Dec algorithm: $Dec(c, d) = m$. If there exist a polynomial time algorithm P such that $d = P(e)$, then E is called a symmetric key scheme. Otherwise, E is said to be an asymmetric, or public key, scheme. For each scheme, perfect security is defined by Shannon as follows: For a uniformly random key pair (e, d) and given thwo messages $m_1 \neq m_2$ and a ciphertext c such that either $c = Enc(m_1)$ or $c = Enc(m_2)$. If an observer cannot guess the right plaintext corresponding to the ciphertext c with a probability greater than $\frac{1}{2}$ then the scheme is a perfectly secure encryption scheme. However, this level of security definition requires to have the keys having same length along with the messages which is quite impractical[7].Secure algorithms are not practical and the attacker is assumed to has limitless time and computational power which is not the case in real world. Therefore, we need a more realistic security model. We need to consider practical encryption and decryption algorithms and the adversaries having reasonable capabilities and the computation powers. Polynomially secure model is more applicable since the attacker has polynomial time computation powers and the algorithms are considered to work in polynomial time. Encryption scheme is said to be polynomial time indistinguishable, if the adversary can not find a polynomial time algorithm to distinguish the right plaintext corresponding to ciphertext in polynomial time for any message pair. In a semantically secure encryption scheme, adversary has additional information about the plaintext not negligibly more than random guess by knowledge of ciphertext. Two definitions were shown to be equivalent[16]. Using polynomial indistinguishability definition, real world attack scenarios can be modeled by allowing the adversary to perform encryptions and/or decryptions before or after receiving the challenge ciphertext.

In this thesis, we give formal definitions of security games and analyse the security levels of well-known private and public encryption schemes along with tools to upgrade security level of these cryptosystems under various assumptions.

In the first chapter, definition of indistinguishability and security levels defined by this notion are introduced. Constructions are defined by "games" and decided over adversarials capabilities on the cryptosystem. First, the chosen plaintext attacks where the adversary has access to the encryption oracle and encrypts messages of her choice then the chosen ciphertext attacks where she has also access to the decryption oracle are considered. Some definitions are divided where the adversary may use the oracles adaptively or not.

In the second chapter, we analyse the security of symmetric encryption schemes. Block cipher modes of operations and authenticated encryption schemes are considered. We will see that deterministic modes of operations are secure only when they used once with the same key and other modes of operations with initial values (IV) are secure with random IVs rather than nonces. After those well known and used modes of operations, authenticated encryption (AE) will be considered. In AE, besides some security requirements on based modes of operation, underlying message authenticated codes (MAC) are supposed to satisfy some security definitions.

In the third chapter, the security of asymmetric encryption schemes are considered. They are classified according to the mathematical hard problems they are based on. Firstly, schemes based on discreet logarithm problem are considered. El-gamal cryptosystem and its successful extension Cramer-Shoup cryptosystem are analysed. Secondly, encryption schemes which are based on integer factorisation problem are considered. RSA is the first encryption scheme based on this problem and it is followed by Rabin, Goldwasser-Micali and Paillier cryptosystems which can also be reducible to integer factorization problem. Thirdly, McEliece cryptosystem and its variatons and extensions which are based on the general linear code problem are examined. Finally, some methods and transformations are considered which aim to upgrade security levels of the above cryptosystems.

Before the detailed description of the security games, we need some basic definitions and descriptions related to symmetric and asymmetric encryption schemes as follows:

1.1 PRELIMINARIES

1.1.1 Symmetric Encryption Algorithms

Symmetric-key encryption scheme basically consists of three algorithms which are key generation, encryption and decryption algorithms.

Key generation is a randomized algorithm Gen, probabilistically chooses a key k from the keyspace $K : \{0,1\}^n$.

Keys generated by key generation algorithm are generated uniformly at random from the key space K (the set of all possible keys outputted by the algorithm) in almost all generation prodecure.

Encryption algorithm Enc_k takes the key k and a message $m \in \{0,1\}^n$, the set of all

"legal" messages supported by the encryption scheme, and outputs a ciphertext $c \in \{0,1\}^n$ such that $Enc_k(m) = c$.

Decryption algorithm Dec_k is a deterministic algorithm such that for given c outputted by Enc_k, $Dec_k(c) = m$ and since the ciphertexts are constructed using plaintexts and keys, plaintext and key spaces define a set $C = \{0,1\}^n$ of all possible ciphertexts.

The scheme provides correct decryption if given any $k \in K$ and $m \in M$,

$$Pr[Enc_k(m) = c : Dec_k(c) = m \ or \ c \ is \ invalid] = 1.$$

The encryption algorithm may be probabilistic(randomized), stateful or both. In order for the encryption algorithm to be random, some random value is used in the encryption process while generating the ciphertext. For each message, the encryption algorithm adds another random value (not used before) and compute the new ciphertext. If the encryption algorithm is stateful, it takes an input value called state, that is initialised in a predefined way. The message, the key and the current state become inputs of the encryption algorithm. Once an encryption is done, the state is updated and stored for the next encryption process. This enforcements on encryption schemes prevents from encrypting the same plaintext to the same ciphertext every time.

On the other hand, decryption algorithm has inputs defined in the encryption process hence can not be randomized or stateful.

Most of the encryption schemes bound the set of plaintexts they encrypt. For example, only the multiple of some block length n is accepted as a length of the plaintext or there exist a limit for the length of the message. If one encryption session does not meet the requirements, encryption algorithm returns an error symbol $\perp$, means not a valid input is feed. Stateless schemes have the set of invalid plaintexts such that it returns $\perp$. On the other hand, stateful schemes return $\perp$ depending not only message length but also possibly the state value. For example, exceeding the pre-defined bound of a counter-like state.

Unless the encryption algorithm outputs $\perp$, decryption algorithm processes correctly and gives the corresponding message m.

1.1.2   Asymmetric Encryption Algorithms

In this part, the communication over a public network is considered, namely asymmetric encryption scheme.

To construct an asymmetric encryption scheme, one can use deterministic one-way trapdoor functions or probabilistic bit by bit encryption algorithms.

One-way function is a function $F : Y \rightarrow Y$ which is computed easily but hard to compute the inverse. There are variety of one-way functions but we need an extra

feature called trapdoor to construct encryption schemes. Trapdoor is a knowledge that helps to invert the function efficiently and must be kept secret, however it must remains hard to invert the function without that trapdoor. Unlike the case in the symmetric one, in public-key cryptosystems we have two or more parties. Each user has its own $<P,S>$ key pair associated with the user. Public key P is reachable by everyone who wants to send a message and only the user knows the private key S. To communicate in this scheme, first the keys are generated by a key generation algorithm and then, to send message m, anyone uses the same encryption algorithm based on a trapdoor one-way function and encrypts the message m by computing $E(m, P)$ and sends the resulting ciphertext c to that user. The user gets the message m by computing $D(c, S)$ where D is the corresponding publicly known decryption algorithm. We need the equation

$$D(S, E(P, m)) = m.$$

holds for the system to work properly. One can formally define a general asymmetric encryption scheme(or trapdoor function scheme) (Gen, Enc, Dec) as follows:

Key generation : The generator G outputs the pair (f, t_f) where f is a chosen trapdoor function and t_f is the associated trapdoor information.

Encryption : For any message $m \in M$, $E(f, m) = f(m)$.

Decryption : For an encrypted message $c \in E(f, m)$ and t_f, $D(t_f, c) = f^{-1}(c) = f^{-1}(f(m)) = m$.

Another way to construct a public-key encryption scheme is to use probabilistic bit by bit encryption algorithms. However, due to their computational costs, they are not widely used.

Single-Bit Encryption Schemes   Besides encrypting the message by dividing some small messages according to the scheme, one can encrypt messages bit by bit by using above trapdoor functions and some hardcore predicates.

A hard core predicate of a function f is a boolean predicate B such that computing the boolean predicate of a given x is efficient but for a given $f(x)$, hard to compute the predicate of x in polynomial time. Formal definition is as follows:

Definition 1.1. B is a hard-core predicate $B : \{0,1\}^* \to \{0,1\}^*$, if

- There exist a probabilistic polynomial time algorithm A such that for all x values $A(x) = B(x)$.

- For all probabilistic polynomial time algorithm G and constants c, there exist a k_0 such that,

$$Pr[G(f(x) = B(x)] < \tfrac{1}{2} + \tfrac{1}{k^c}$$

for random choices of x of length $k > k_0$.

A trapdoor predicate is a hard-core predicate except the trapdoor information is used for calculating the predicate function value.

A public key encryption scheme can be defined with these tools as follows:

Key Generation : Gen chooses (i, t_i) by running algorithm S_1 where i and t_i are public and private keys respectively.

Encryption : Let m be binary message, Encryption algorthm $Enc(i, e)$ selects $x \in D_i^m$ and the ciphertext becomes x by using algorithm S_2.

Decryption : $Dec(c, t_i)$ computes $B_i(c)$ using a polynomial time algorithm and gets the binary message m.

An advantage of the single bit encryption schemes is that every bits security depend on a random value in the domain of the predicate hence no information can leak for any individual bit of the bitstring.

In some public-key encryption schemes, manipulation feature on ciphertext is needed. In that case homomorphic encryption algorithms are used to satisfy that need.

1.1.3 Some Security Properties

For some attack scenarios, encryption schemes need to provide some properties like one-wayness and malleability. Definitions are as follows:

One-wayness
A function is one-way if it is easy to compute and hard to find the inverse. Regarding encryption schemes, one wayness is defined for the algorithm that is used for encryption rather than arbitrary functions. Formally;

Definition 1.2. An encryption algorithm $Enc : \{0, 1\}^* \rightarrow \{0, 1\}^*$ is one way if

- There exist a probabilistic polynomial time (PPT) algorithm to calculate $c = Enc(m)$ for a message m,

- For every PPT algorithm A, there exist a negligible function $n(k)$ where k is the security parameter such that,

$$Pr[Enc(m) = c; A(1^k, c)] \leq n(k).$$

for sufficiently large k.

Malleability

Malleability is being able to produce another ciphertext from given ciphertext such that the new ciphertext has corresponding plaintext related to the corresponding plaintext of the given ciphertext.

CHAPTER 2

INDISTINGUISHABILITY

2.1 Security Definitions

For encryption schemes, one can define security in various ways. According to Shannon, an encryption scheme is perfectly secure against all adversaries having unlimited time and computational power if cphertext reveals no information about the corresponding encrypted plaintext in the plaintext space. Formally,

Definition 2.1. For an encryption scheme $E(Enc, Dec)$ defined over key-space K, message-space M and ciphertext-space C, consider a probabilistic experiment in which the random variable k is uniformly distributed over K. For all $m_0, m_1 \in M$, and $c \in C$, the following equality holds

$$Pr[Enc(k, m_0) = c] = Pr[Enc(k, m_1) = c].$$

This definition can be summarized as a ciphertext reveals no information about the encrypted plaintext for adversaries who have unlimited computation power and time. This definition guaranties the security, however, no encryption scheme satisfies this security level except one time pad, no encryption scheme satisfies this security level due to its impractical assumptions. In real world, adversaries have limited time and computing power, and the encryption algorithms should be polynomial time. Hence, security definition can be as powerful but not applicable as Shannon's perfect security definition or it can be defined only by attacks that are known so far. Security of an encryption scheme needs to give confidence as in the case of perfect secrecy but needs to be efficient and applicable to that days encryption schemes. From this point of view, security definitions turns out to be equivalent to computational security.

Before giving a generic security definition, environments that the parties communicate over have to be well defined. In symmetric and asymmetric encryption schemes, there are different types of problems that need to be considered.

In private key encryption schemes, security definition is about the privacy of the keys and the plaintexts. In an ideally secure scheme, adversary shouldn't extract any information about the plaintext even a partial information. In reality, this is not the case since at least the message space like English words or the length of the plaintexts are

known in most of the schemes. Hence, the security notion is defined with the adversaries who have limited computing power. Adversary may reveal valuable information about the plaintext but not feasibly in efford or time.[15]

A public-key encryption scheme needs several properties to be considered as secure. As a starting point, properties that have minimum requirement can be considered first and security notions build up on them. First of all, the private key recovery shouldn't be possible by accessing the public key. Secondly, with the public key and ciphertext of a message from a message space shouldn't be recovered entirely with a high probability. Furthermore, ciphertext shouldn't leak any useful facts about the messages. Thirdly, an adversary shouldn't be able to compute any valuable information about the message traffic, for example she shouldn't recognize the two messages of same content were sent.

There are some security problems special to the schemes with trapdoor functions. First one is that the invertion will become not that hard any more if the message space consists of restricted elements like English language or like M consists of only 0 and 1. Distinguishing $f(0)$ and $f(1)$ become not that hard in that case. Secondly, chosen one way trapdoor function does not neccesarily guarantees to hide all the information about m and hence from any information leaked can become a start point of an attack. Lastly, for the encryption algorithms that are deterministic, every message has a unique corresponding ciphertext under the same key pair. Encryption scheme leaks information when sending same message twice or to multiple recipients. Also, relevant messages have corresponding relevant ciphertexts in this set up which is not a desired property.

To deal with these problems in real world, Goldwasser and Micali first introduced semantically secure encryption schemes in their paper probabilistic encryption[16]. They constructed security definition on adversary who has polynomial time computation power instead of unlimited power.

However for provable security with the notion indistinguishability is widely accepted because of the applicability over the encryption schemes.

2.2 Indistinguishability

Indistinguishability is the concept that modern cryptography use to construct strong pseudorandom generators, secure encryption schemes, commitment schemes and more. Indistingushability is used in definitions of computational indistinguishability or polynomial indistinguishability considering X and Y over $(0,1)^{l(n)}$.

In this thesis we deal with the polynomial indistinguishability in encryption schemes, i.e. given a ciphertext out of two plaintexts, it is not possible to distinguish which plaintext is the one encrypted. More formally;

Definition 2.2. Let m_0 and m_1 be two messages and $E(Gen, Enc, Dec)$ be an encryption scheme. Let c is encryption of one of the messages chosen uniformly at random by E. Then the cryptosystem E is polynomial time indistinguishable if for all PPT M, A, and for any polynomial Q, for all sufficiently large k

$Pr(A(1^k, e, m_0, m_1, c) = m|(e, d) \leftarrow G(1^k); \{m_0, m_1\} \leftarrow M(1^k); m \leftarrow \{m_0, m_1\}; c \leftarrow E(e, m)) < \frac{1}{2} + \frac{1}{Q(k)}$.

The definition is used to define stronger attack models indistinguishability under chosen plaintext attacks, called IND-CPA and indistinguishability under chosen ciphertext attacks, called IND-CCA security on both symmetric and asymmetric cryptosystems rather than passive attacks like known plaintext or known ciphertext attacks. Active attacker has capabilities that she has access to the encryption and/or decryption algorithms whereas in passive attack types the attacker only listens and captures plaintexts and/or ciphertexts. One of the de facto requirement of a secure encryption scheme is one can not relate any ciphertext with its plaintext pair, i.e. no information is leaked by only seeing the ciphertext.

2.3 Attack Types Defined by Indistinguishability Notion

Attack scenarios are designed according to attackers capabilities during the attack. One can costruct various scenarios by playing with this capabilities considering real-world examples or not.

First the definition of primitives are needed in order to construct the games played.

Left or right encryption oracle
It is an oracle such that for a given encryption scheme and two messages of the same length, it encrypts one of the two messages uniformly at random and returns the corresponding ciphertext.

Left-or-right encryption oracle (LR-Oracle) of an encryption scheme $E(K, E, D)$ with keyspace K, encryption and decryption algorithms, E and D respectively, is defined as follows;

Algorithm 1 LR-Oracle

Oracle $E_K(LR(M_0, M_1, b))$ // $b \in \{0, 1\}$ and $M_0, M_1 \in \{0, 1\}^*$,

    if $|M_0| \neq |M_1|$ then return $\bot$,
    $C \leftarrow E_K(M_b)$,
    return C.

Decryption oracle
Decryption oracle, upon receiving a ciphertext, first checks whether it is legitimate or not. Legitimate means it is not previously returned by LR-Oracle. Then returns corresponding message if it is legitimate otherwise returns an error $\bot$.

Decryption oracle of an encryption scheme $E(K, E, D)$ with keyspace K, encryption and decryption algorithms, E and D respectively, is defined as follows;

Algorithm 2 Decryption Oracle

Oracle $D_K(C)$

$\qquad M \leftarrow D_K(C),$

$\qquad$ if M is previously encrypted then return $\perp$,

$\qquad$ else return C.

Negligible value

It is a value that can be taken as zero for all practical purposes. For example 2^{-100} is a negligible value for today's technology.

2.3.1 Indistinguishability Under Chosen-Plaintext Attack

Consider an adversary who has access to the encryption machine and can perform encryptions of set of messages of her choice. This is trivial when we consider the public-key cryptography. Then she can use these informations to gain some information about the plaintext upon receiving an arbitrary ciphertext. Considering this scenario, an attack model can be costructed so that the attacker may capture the encryption machine and without neccesarily looking inside the encryption process, she may be able to encrypt messages of her choice, called chosen plaintext attacks, and use this information later. In such a situation, not only a secure encryption algorithm but also a secure encryption scheme is needed. This scenario is expressed as a game by Goldwasser and Micali in 1984 to construct a provably secure encryption schemes. It is also showed that the definition is equal to semantic security definition. Semantic security is a property of a cryptosystem where for a given ciphertext of a certain message no probabilistic polynomial time algorithm (PPTA) can determine any partial information on the message with probability not negligibly more than all other PPTA's having only the message length information[16].

According to Goldwasser and Micali, for a given encryption scheme $E(Gen, Enc, Dec)$, indistinguishability under a chosen plaintext attack can be defined as a game where an adversary A first allowed to query the encryption oracle with messages and get back the corresponding ciphertexts(called learning phase). Then she chooses two messages of the same length and gives to the left-or-right encryption oracle. The oracle computes one of the messages and give it back to the adversary(called challenge phase). It is said the encryption scheme provides IND-CPA security if the adversary tells the correct message is the one encrypted in a reasonable amount of time.

More formally, the experiments of the game played are defined as follows respectively;

Algorithm 3 IND-CPA Game

Let A be an algorithm having access to an oracle.

Experiment $Exp_E^{ind-cpa-cg}(A)$

$b \leftarrow \{0,1\} ; K \leftarrow \mathsf{K}$,

$b' \leftarrow A^{E_k(LR(-,-,b))}$,

if $b = b'$ then return 1,

else return 0.

It is assumed that the adversary interact with the oracles without knowing the inner working process and has no control over it. She rather use it as a blackbox and gets the return values. Note also that random choices of the $LR-Oracle$ are independent from the random values or states in the randomized or stateful definitions of the encryption algorithms respectively. We define IND-CPA advantage of the adversary A as follows:

$$Adv_E^{ind-cpa}(A) = 2Pr[Exp_E^{ind-cpa-cg}(A) = 1] - 1.$$

$Pr[Exp(a) = 1]$ represents the probability of the successive guesses and $Adv(a)$ measure the success of the adversary A in the game. $Pr[Exp(a) = 1]$ is assumed to be greater or equal to $\frac{1}{2}$ since the adversary guesses random bit b half of the time at worst guesses like simply chosing randomly or chosing always 1 or 0.(otherwise adversary A has another advantage of not knowing uniformly random b value more than $\frac{1}{2}$)

2.3.2 Indistinguishability Under Chosen-Ciphertext Attack

Assume that the attacker has more capabilities over the scheme than IND-CPA scenario namely access to the decryption oracle as well as the encryption oracle. This type of attack become realistic only with some restrictions on use of decryption oracle. This scenario is thought as adversary has access to the decryption oracle temporarily and the oracle refuses to decrypt previously produced ciphertexts. In a way, the security of ciphertexts that are produced after the adversary has access to decryption oracle is considered rather than the security of fortcoming ones. A ciphertext is called legitimate if it is produced by the lr-encryption oracle. Otherwise it is called legitimate.

For a given symmetric encryption scheme $E(Gen, Enc, Dec)$, in addition to her abilities in IND-CPA definition, the attacker A also has access to the decryption oracle. She can feed the decryption oracle with any legitimate ciphertext before the challenge phase and get the corresponding plaintext.Then gives two challenge messages of her choice. lr-encryption oracle returns a value uniformly random.After this phase adversary may or may not query the decryption oracle. If she does then it is called adaptive, otherwise called non-adaptive. Consider the following experiment:

Algorithm 4 IND-CCA Game

Let A be an algorithm having access to both encryption and decryption oracles and $b \in \{0, 1\}$

Experiment $Exp_E^{ind-cca-b}(A)$

$K \to K$,

$b \to A^{E_k(LR(-,-,b)),D_K(-)}$,

A queried $D_K(-)$ on a ciphertext previously returned by $E_K(LR(-,-,b))$,

then return 0,

else return b.

The advantage $Adv(a)$ of the adversary is defined as

$$Adv_E^{ind-cca}(A) = Pr[Exp_E^{ind-cca-1}(A) = 1] - = Pr[Exp_E^{ind-cca-0}(A) = 1].$$

Here the conventions that are not mentioned are the same as in the case of IND-CPA.

The encryption scheme is indistinguishable under non-adaptive chosen ciphertext attack (IND-CCA1) if an attacker can not gain a significant advantage while playing the game non adaptively defined in a reasonable usage of her resources. It is called indistinguishable under adaptive chosen ciphertext attack (IND-CCA2) if the attack is adaptive. The IND-CCA notion is so strong that the most of the modes of operations and even one time pad algorithm can be broken under IND-CCA assumptions.

In asymmetric cryptosystems, adversary has additional power i.e. public key. Hence in public-key encryption, adversary is more powerful than in the private-key encryption. One can clearly see that, according to the advantages of the adversary in the games, an encryption scheme that satisfy IND-CCA2 satisfies both IND-CCA1 and IND-CPA. Also IND-CCA1 security implies IND-CPA. However, inverse is not always true.

CHAPTER 3

IND-CPA and IND-CCA Security of Symmetric Encryption Schemes

In this chapter, security of widely used and standarized symmetric encryption schemes are considered under indistinguishability notion. After considering stream cipher-based encryption schemes, block cipher-based encryption schemes will be discussed under block cipher with confidentiality modes of operations and authenticated encryption schemes.

3.1   Security of Stream Cipher-Based Encryption Schemes

Stream ciphers are mostly XOR-based ciphers. In XOR-based cipher, plaintext bits are XOR-ed with the keystream generated by the state of the underlying stream cipher. Well known stream ciphers are RC4, Salsa, ChaCha. XOR-based encryption schemes using stream ciphers as key stream generator are IND-CPA secure if the underlying stream cipher is strong pseudo random generator (PRG)[7].

$$Adv_E^{ind-cpa}(A) \leq Adv_E^{prf}(B).$$

However, IND-CCA security is trivially fails since adversary may XOR challenge ciphertext with any plaintext and wins the game with advantage 1.

3.2   Examples of Modes of Operations in Symmetric Encryption Schemes

In this section, widely excepted block cipher modes of operations and their security levels with respect to IND-CPA and IND-CCA assumptions are considered. First six modes of operations discussed below are called as confidentiality modes. The last three are called authenticated-encryption modes.

3.2.1 Confidentiality Modes

This six modes of operations, except the electronic code book (ECB), have initial value (IV) which can be either a random number or a nonce value. In IV-based modes of operations, choices of IV as nonce or random play important role in security. If the message length is not a multiple of the block size, the message must be padded to a multiple of the block length in all these six modes.

3.2.1.1 Electronic Code Book

Definition 3.1. Electronic code book (ECB) is a block cipher mode of operation that is approved in various standards [26]. It is the only deterministic mode of operation. The details of encryption and decryption using ECB is given in 3.1 [8]
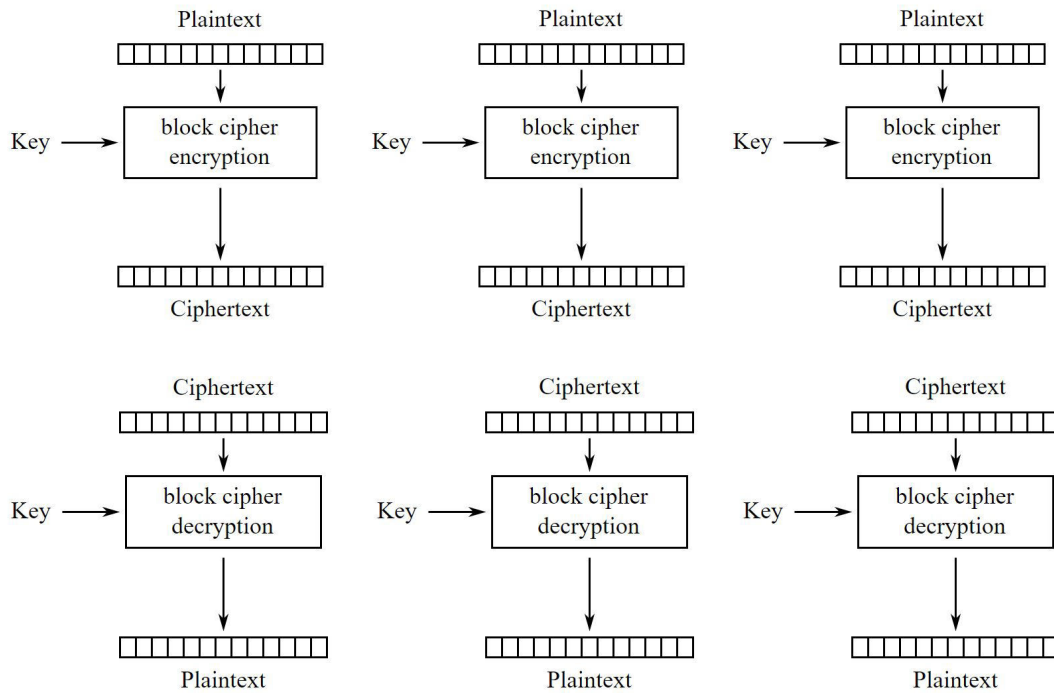
Figure 3.1: ECB encryption and decryption

Since ECB is a deterministic mode of operation, it satisfies neither IND-CPA nor IND-CCA security. An attacker has a full advantage on knowing the challenge ciphertext c_b if she may ask m_1 or m_2 before the challenge. This mode is recommended to be used only with messages having length equal to the block size of the algorithm and should be used only once in an encryption scheme.[26]

14

3.2.1.2 Cipher Block Chaining

Definition 3.2. Cipher Block Chaining (CBC) mode of operation takes block cipher $E : K \times \{0,1\}^n \to \{0,1\}^n$ and a known initialisation vector IV as input. The details of encryption and decryption using CBC is given in 3.2
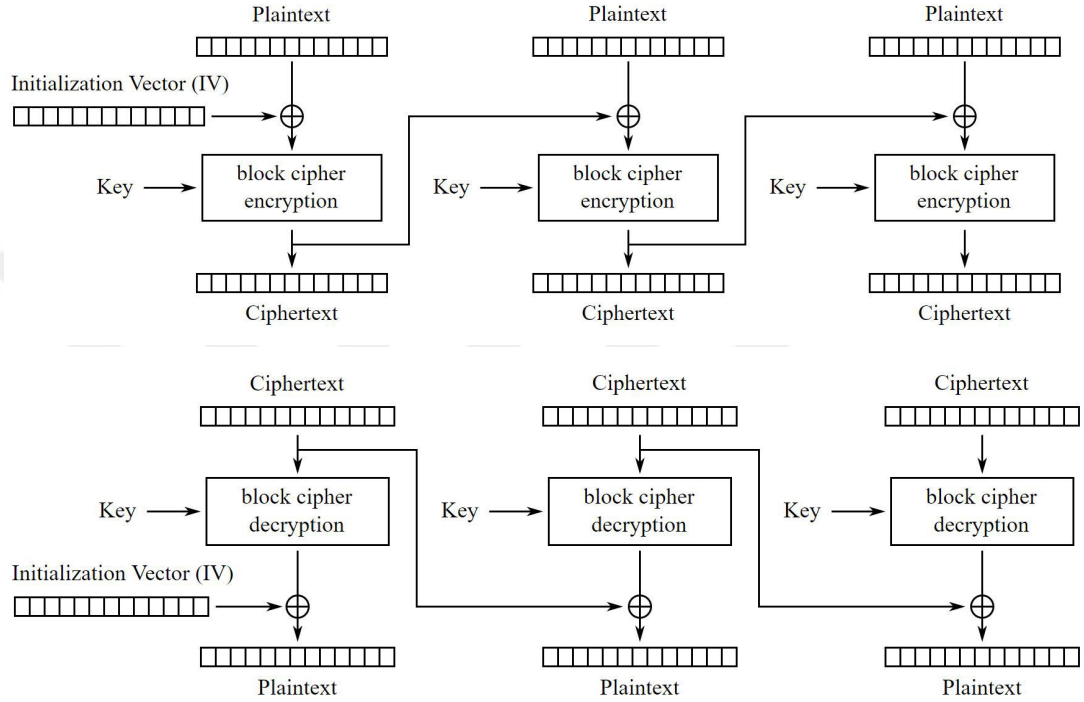

Figure 3.2: CBC encryption and decryption

CBC mode with random IV is IND-CPA secure while the CBC mode with nonce IV is not. Consider an attacker attacking CBC with random IV. She queries $p_1, ..., p_q$ plaintexts having length a positive multiple of n-block size. Then the advantage of the adversary is less than $\frac{q^2}{2^{n+1}}$.

When the IV is nonce-based, Adversary asks oracle to encrypt (n_1, p_1) and (n_2, p_2) where $n_1 = p_1 = 0^n$ and $n_2 = p_2 = 1^n$. For the challenge plaintexts p_1 and p_2, the adversary outputs zero if $c^1 \neq c^2$, and one if $c^1 = c^2$.

It can be shown that it CBC is not IND-CCA secure by the following attack scenario:

The attacker asks the oracle to encrypt a random plaintext $p = p_1 p_2 p_3$ having $p_i's$ lengths equal to n and then receives the output $c = (IV, c_1 c_2 c_3)$. Then for a random IV' and ciphertext $c' = c_1' c_2 c_3$ of the same length (3n), adversary ask the decryption oracle to decrypt $(IV', c_1' c_2 c_3)$ and gets $(p' = p_1' p_2' p_3')$. For challenges p and p', if $p_3 = p_3'$, then she knows $e(p) = c$ or $e(p') = c$ otherwise. This shows that CBC is not IND-CCA secure.

3.2.1.3 Cipher Feedback

Cipher Feedback (CFB) mode has the same input and parametrization with the CBC mode, however, unlike the CBC mode it works as a self-synchronous stream cipher as in the figure 3.3
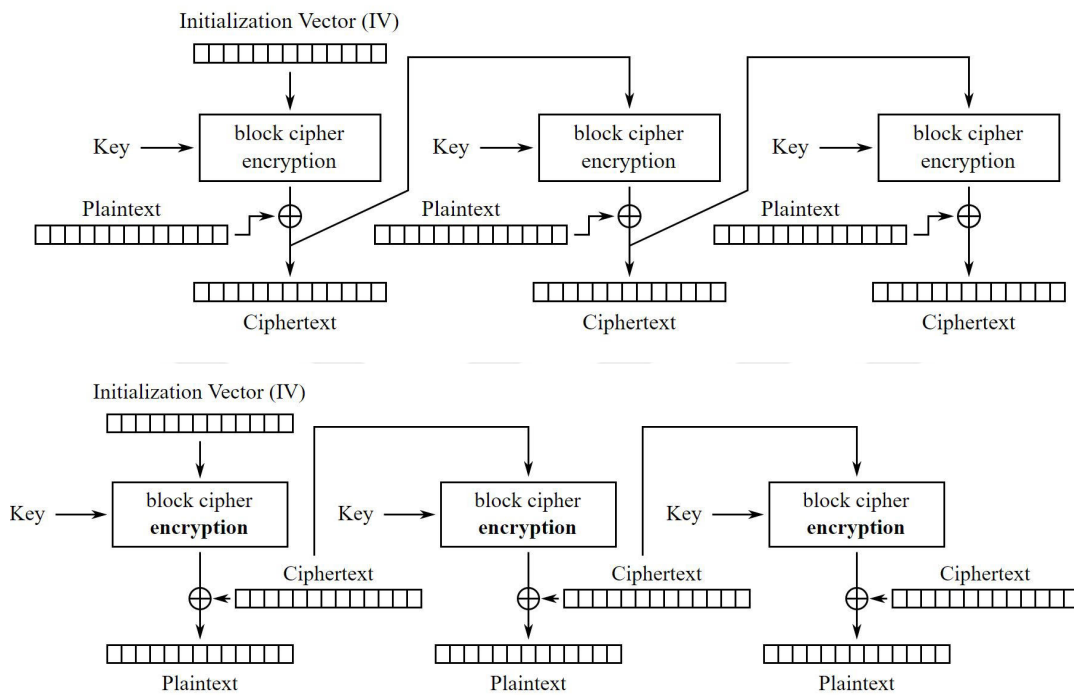
Figure 3.3: CFB encryption and decryption

Like CBC mode, IND-CPA security depends on the choice of the IV. If the IV is chosen randomly for each encryption, then the attacker has negligible advantage as in the CBC mode with random IV except that ρ is the r bit data segments.

To break CFB mode with nonce-based IV using IND-CPA, first the attacker sends the following challenges to the oracle: A 2-block plaintext $p_1 = 0^{2n}$ with a nonce value $n_1 = 0^n$. After receiving the challenge ciphertext $c_1 = c_11c_12$, she asks the encryption of $(n_2 = c_11, p_2 = 0^n)$ and gets c_2. If $c_2 = c_12$ then $E(p_1, N_1) = c_1$ otherwise $E(p', N_1) = c_1$. However, CFB mode is not IND-CCA secure by considering the same attack scenario in CBC mode.

3.2.1.4 Output Feedback

Output Feedback (OFB) mode has same parameters and inputs with CBC and CFB modes. It works as a synchronous stream cipher as in figure **??**
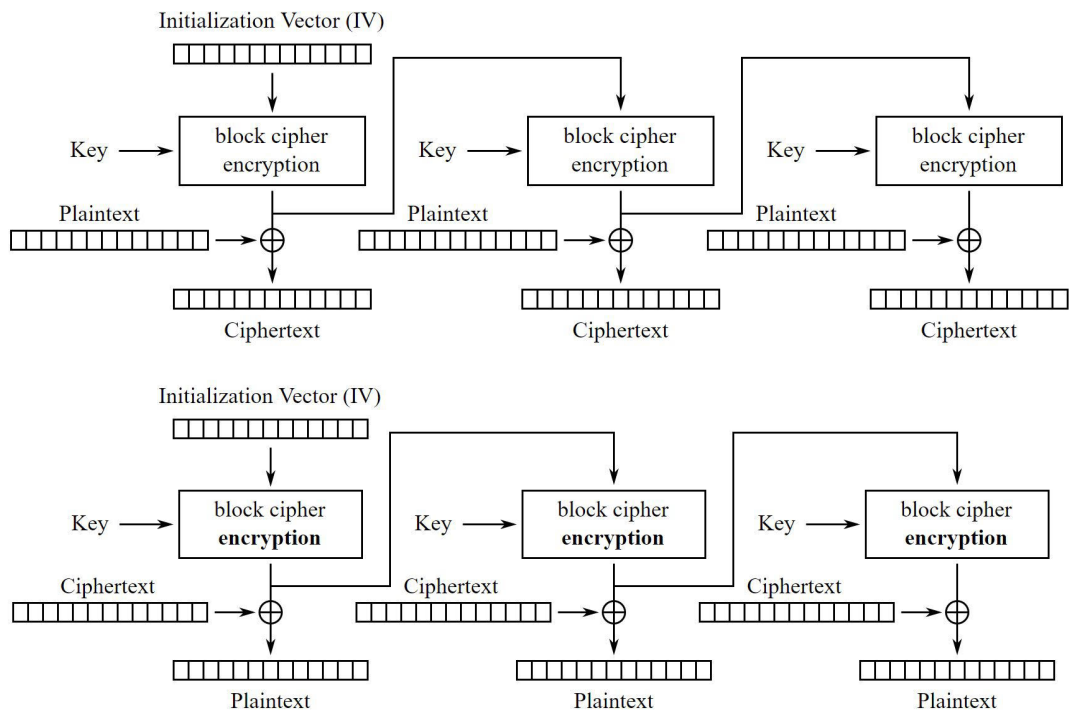
Figure 3.4: OFB encryption and decryption

OFB mode also shares the same security levels up to constraints on IV as CBC and CFB. To break OFB with nonce IV, same same attack scenario can be used as in CBC [26] in order to show that it is not IND-CPA secure. For random (or nonce) IV version of OFB can also be shown that it is not IND-CCA secure by for a given ciphertext c produced from the message m with initial value IV, adversary can ask the decryption of c' where c' is the bitwise complement of c.

3.2.1.5 Counter

Counter (CTR) mode is also parametrized by any block cipher $E : K \times 0, 1^n \rightarrow 0, 1^n$ and takes key, plaintext and a nonce (counter) value. Different from CBC, CFB and OFB, the nonce value has exactly the same length as the plaintext. It also has a property that the decryption algorithm is exactly the same as the encryption algorithm and only encryption algorithm of the corresponding blockcipher is sufficient. This property allows to use pseudorandom functions, not neccesarily a permutation, as the blockcipher.
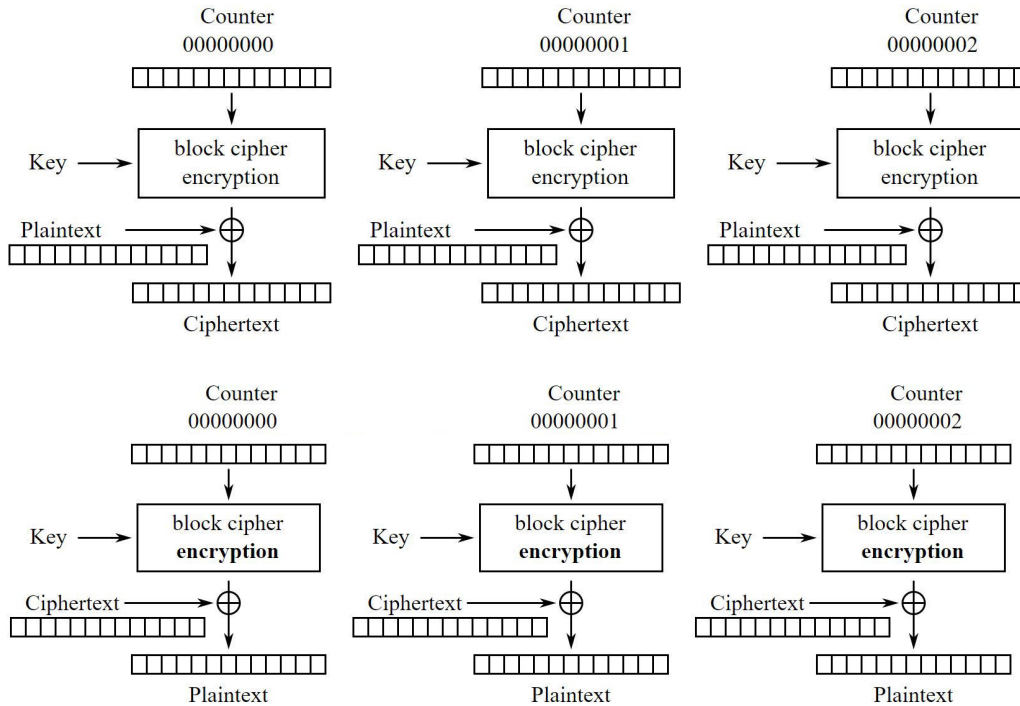
Figure 3.5: Counter mode encryption and decryption

According to Rogaway[26], using pseudo-random function instead of pseudo-random permutation brings "perfect" IND-CPA security to CTR mode.

Let A be an adversary in IND-CPA game that runs in time at most t and asks at most q queries, these totaling at most ρ n-bit blocks. Then there exists an adversary B (attacking the underlying block cipher) such that the advantage of the attacker A is less than or equal to the advantage of B where B runs in time at most $t' = t + O(q + 2n\rho)$ and asks at most ρ oracle queries.

3.2.1.6 Xor-Encrypt-Xor-Based Tweaked-Codebook

XEX-based tweaked codebook mode with ciphertext stealing (XTS) is designed to used as full disk encryption mode. Other uses are not recommended in standards as IEEE Standards [26] for its possible weaknesses. Security definition with indistinguishability is not applicable[26] for this mode although the underlying mode of operation XEX2 is IND-CCA secure.

Figure 3.6: XTS Encryption

3.2.2 Authenticated Encryption Modes

None of the symmetric encryption schemes with modes of operation satisfy IND-CCA security so far. In order to create more secure modes of operations 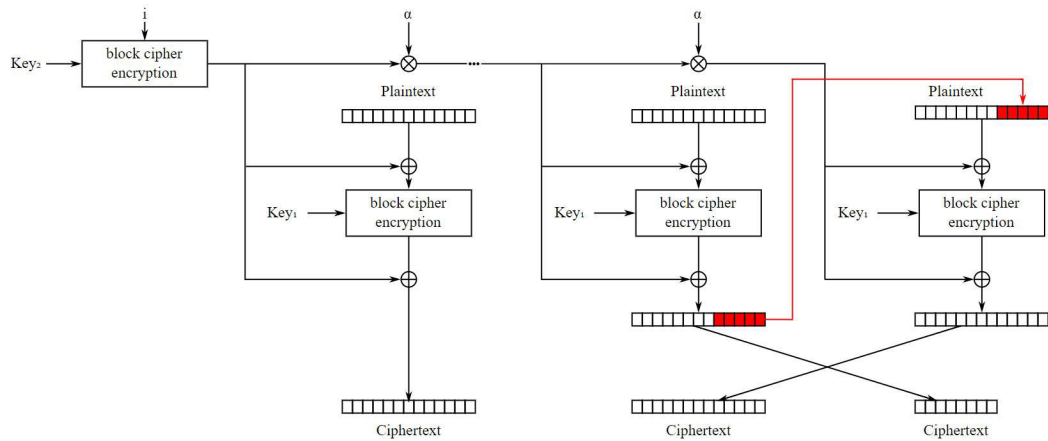one can use a message authentication code, MAC. A deterministic MAC is system of two algoritms (S, V) called signing and verification algorithm. Signing algorithm S generates a string of bits called tag and verification algorithm V verifies the tags. Formally;

- S is a deterministic or a probabilistic algorithm having inputs as a key k and a message m and outputs t, called tag.

  $S : K \times M \rightarrow \{0, 1\}^\tau$

- V is a deterministic algorithm that takes a key k, a message m and a tag t, and outputs either accept or reject.

  $V : K \times M \times T \rightarrow r$ where r is either accept or reject.

It is required that for all keys k and messages m,

$$Pr[k, m, S(k, m) = accept] = 1$$

where K is key space, M is message space and T is tag space for given values.

If the signing algorithm is deterministic then it is called deterministic MAC system and the tags generated are unique for specified key and message. MAC systems can be randomized by choosing signing algorithm S randomized.

Most of the deterministic MAC algorithms uses signing algorithm as CBC mode of operation encryption algorithm.The basic CBC-based MAC takes a key and a message as input and returns the final block of the ciphertext as tag.
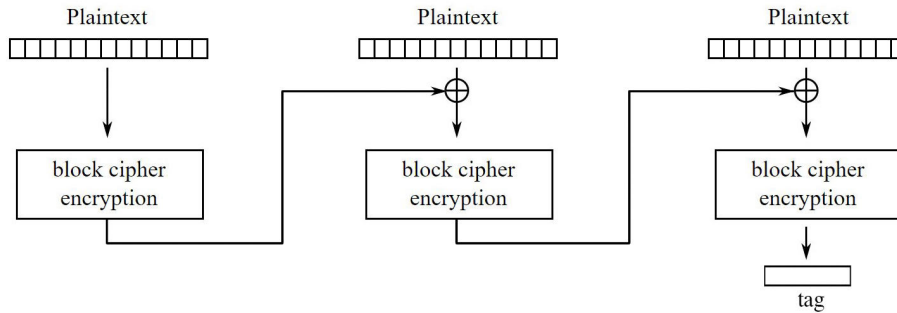
19

Figure 3.7: CBC-based MAC

In ISO/IEC 9797-1 standards, there are 6 different version of MAC which all are based on CBC-mode encryption. For detailed information reader refer to [15]. There are also hash based MAC systems.Mostly used ones are HMAC,standardized in NIST FIPS 198-1, RFC 2104, IEEE 802.11, SSL 3.0, TLS 1.0–1.2, SSH, and S-HTTP[15].

Provable security of a MAC algorithm is defined [7] by existential unforgeability under an adaptive chosen-message attack. Basically it means the adversary should not be able to forge a valid tag on any message having capability that she can obtain any tags corresponding to arbitrarily chosen messages during her attack. Details of the attack scenario, ie game and formal definition, reader refer to [7].

Authenticated-encryption modes (AE) are simply uses modes of operations and MAC systems to deliver both privacy and authenticity. They can be combined either as composition or can be used integrated. This partitioning commonly called one-pass and two-pass AE schemes. CCM, Counter-mode encryption with CBC-MAC authentication, and GCM, Galois counter-mode with Carter-Wegman MAC authentication are two examples of one-pass AEs whereas EAX is a two-way AE scheme. A message can be send with an AE scheme in 3 ways:

Authenticate-then-encrypt (AtE) Tag is first computed then the message and the tag produced are encrypted together.

Encrypt-then-authenticate (EtA) First the message is encrypted and then tag is computed over the produced ciphertext.

Encrypt-and-authenticate (EandA) Encryption and tag generation processes independently.

For detailed version, reader refer to [7]

Not all the above approaches are secure with even with secure encryption and MAC algorithms. Among these, EtA approach is IND-CCA secure if the encryption scheme is IND-CPA secure and MAC algorithm is unforgeable. Security of some widely used and standarized authenticated encryption, modes of operations, are as follows:

3.2.2.1 Counter with CBC-MAC

Counter with CBC-MAC (CCM) has parameters a blockcipher E, a formatting function $Format$, a counter-generation function $Count$ and tag length $\tau \in [32, ..., 128]$. CCM works in authenticate-then-encrypt type.
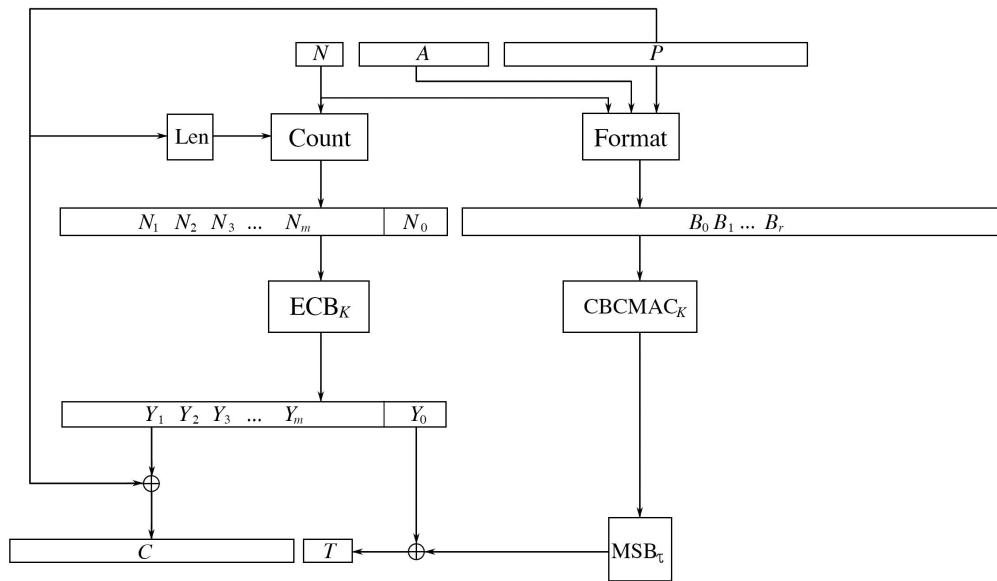
Figure 3.8: CCM encryption

Jacob Jonsson proved[17] that CCM mode of operation is IND-CCA secure with the advantage of the adversary

$Adv_{CCM} \leq \dfrac{\rho^2}{2^n}$ where ρ is the number of calls adversary made and blockcipher has input length n.

3.2.2.2 Galois Counter Mode

Galois Counter Mode (GCM) is parametrized by a 128-bit blockcipher (suggested as AES in SP 800-38D [12]) and τ-tag length $\in [32, 64, 96, 104, 112, 120, 128]$.GCM works in encrypt-then-authenticate type.

21
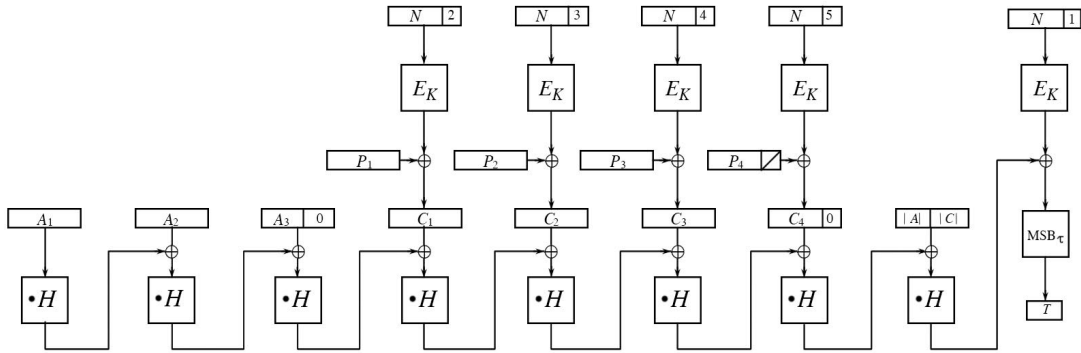
Figure 3.9: GCM encryption

McGrew and Viega find[20] a bound to the advantage of an attacker in an IND-CCA game as

$$Adv_{GCM} \leq \frac{0.5(\rho + 2q) + 0.5(\rho + 2q)(l_N + 1)}{2^n} + \frac{q(l + 1)}{2^\tau}$$

where l_N bounds the queried nonce N, l bounds the sum of the block lengths for each (A,C) pair that arises during the adversaries queries, q is a bound for number of queries, ρ bounds the whole blocklength of the processed plaintexts.

3.2.2.3 Encrypt-Then-Authenticate-Then-Translate

Encrypt-Then-Authenticate-Then-Translate (EAX) authenticated encryption mode is proposed by Bellare et. al.[3] as an alternative to CCM. It is parametrized by n-bit blockcipher E and tag length $\tau \in [0, ..., n]$.
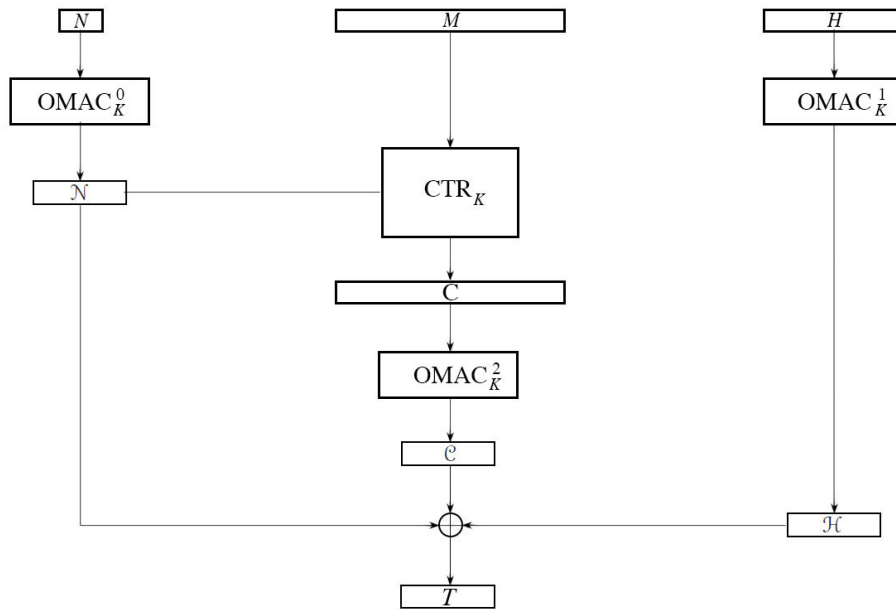
Figure 3.10: EAX encryption

EAX authenticated encryption is IND-CCA secure and Bellare et. al.[3]show that the adversary has advantage in EAX mode Adv_{EAX} less than or equal to

$$Adv_{EAX} \leq \frac{9.5\rho^2}{2^n} + Adv_E$$

CHAPTER 4

IND-CPA and IND-CCA Security of Asymmetric Encryption Schemes

In this section the encryption schemes considered under the mathematical problems they are based on and some well-known padding algorithms and transformations are considered.

4.1 Based on Discreet Logarithm Problem

There are various equivalent of Diffie-Hellman Decisional problem but the following one can be considered to construct an encryption scheme.

Definition 4.1. Let G be a group of a prime order q, for the following distributions R and D:

R of random quadruples (g, g^x, g^y, g^z) element of G_4;

D of quadruples (g, g^x, g^y, g^{xy}) element of G_4, where $g \in G$, $x, y, z \in Z_q$. g, called base, can be either random or fixed.

Decisional Diffie-Hellman problem is considered to be hard if there exists no polynomial time statistical test that can distinguish R and D adequately.

Alternatively, for given a quadruple coming from R or D, the difference between the probabilities that the algorithm guesses correctly for quadruples from R and for quadruples from D is non-negligible.

4.1.1 El-Gamal Cryptosystem

Definition 4.2. Built on this hard problem, basic El gamal encryption scheme is as follows;

Key generation :

Alice chooses a cyclic group G of order $(q-a)$ large prime and g be a generator element of G.

Then she chooses a random $x \in \{1, ..., q-1\}$ and computes $h = g^x$

Publishes (G, q, g, h) as her public knowledge.

Encryption :

In order to send message m to Alice with her publics (G, q, g, h), Bob chooses a random $y \in \{1, ..., q-1\}$ and calculates $c_1 = g^y$

Calculates $s = h^y$ which is equal to g^{xy}

Bob hides his secret message by calculating $c_2 = ms$

Finally he sends $(c1, c2) = (g^y, mg^{xy})$ to Alice

Decryption :

Upon receiving the pair (c_1, c_2) Alice calculates $s = c_1^x$

Then computes $m = c_2 s^{-1}$ where s^{-1} is the inverse of s in the group G.

This scheme is IND-CPA secure since even the attacker encrypts the same message m or similar messages m, in every encryption fresh random y is used and g^{xy} is indistinguishable from a random element in the group G. This implies mg^{xy} is indistinguishable if the DDH problem holds in the group G. Key point to gain semantic security is to choose a new random number y for each encryption. When it comes to the IND-CCA security, El-gamal Encryption Scheme fails as follows:

The attacker simply sends its challenge pairs (m_0, m_1) and LR-Encryption oracle chooses $b \in \{0, 1\}$ uniformly at random and gives (c_1, c_2) for a random y'. Then the attacker chooses a random number y'' and calculates $(c_1' = c_1 g^{y''}, c_2' = c_2 g^{y''})$ to decryption oracle. This is a valid action since $(c_1', c_2') \neq (c_1, c_2)$. Since $c_1' = g^{x+y}$ and $c_2' = m_b g^{x+y}$, oracle decrypts and show the message m_b by computing $s' = c_1'^x$ and $m_b = c_2' s'^{-1}$. Then attacker's advantage is 1 on guessing the bit b.

4.1.2 Cramer-Shoup Cryptosystem

In 1998, provably secure encryption scheme based on the same problem was developed by Ronald Cramer and Victor Shoup as an extension of El-gamal cryptosystem, largely called Cramer-Shoup Cryptosystem. This scheme proved to be IND-CCA secure[9]. The algorithm uses universal one way family of hash functions to reach this security level. The basic scheme proposed in the original paper is as follows:

Definition 4.3. Key generation :

Find a group G of order large prime q and choose two random elements $g_1, g_2 \in G$.

Then choose also secret randoms $x_1, x_2, y_1, y_2, z \in Z_q$.

Compute $c = g_1^{x_1} g_2^{x_2}$, $d = g_1^{y_1} g_2^{y_2}$ and $h = g_1^z$

Next from the family of universal one way functions, choose a hash function H and give (g_1, g_2, c, d, h, H) as public tuple.

Encryption :

One can send message $m \in G$ by first choosing a random $r \in Z_q$ and computing

$u_1 = g_1^r$, $u2 = g_2^r$, $e = h^r m$, $\alpha = H(u_1, u_2, e)$, $v = c^r d^{r\alpha}$

Then the ciphertext corresponding to m and random r becomes (u_1, u_2, e, v).

Decyrption :

If we have ciphertext tuple (u_1, u_2, e, v), decryption process as follows:

First α is computed $\alpha = H(u_1, u_2, e)$ and check whether $u_1^{x_1+y_1\alpha} u_2^{x_2+y_2\alpha} = v$ holds.

If this condition is not satisfied,output reject else

$m = e u_1^{-z}$

This scheme can be verified as an encryption scheme by showing

$$D(E(m,r)) = m$$

as follows:

Since $u_1 = g_1^r$ and $u_2 = g_2^r$ and we have $u_1^{x_1} u_2^{x_2} = g_1^{rx_1} g_2^{rx_2} = c^r$.

Similarly, $u_1^{y_1} u_2^{y_2} = d^r$ and $u_1^z = h^r$, so test phase of decryption algorithm passes and output will be $m = e h^{-r}$.

This encryption system is IND-CCA secure under the assumptions that the hash function comes from a universal one-way hash function family and the DHCP is hard in the group G. This security level is gained by the decryption oracle's test phase,i.e. rejecting the invalid ciphertext tuples at least with probability $1 - \frac{1}{q-i+1}$) which can be considered as negligible enough to the security definition.

4.2 Based on Integer Factorization Problem

4.2.1 RSA

RSA function is the first trapdoor function which is used in PKE proposed by Rivest, Shamir and Adelman. The function depends on the integer factorization problem which is NP and suspected that not to be in NP-complete. But since no polynomial time algorithm has been published that can factor all integers despite the fact that it is widely studied, cryptographic schemes rely on this problem are used extensively. Formal definition of RSA-based encryption scheme is as follows:

Definition 4.4. RSA cryptosystem consists of 3 algorithms:Key generation, encryption and decryption.

Key generation :

First choose two distinct very large prime numbers p and q (large term depends on the computational power of that days technology)

Compute $n = pq$

Compute $\phi(n) = lcm(p-1, q-1)$

Choose an integer $e \ni 1 < e < \phi(n)$ and $gcd(e, \phi(n)) = 1$

Determine d as $e^{-1} \mod \phi(n)$

Encryption :

To send a message M

First turn M into an integer m using reversible $padding\,algorithm \ni 0 < m < n$.

Then compute ciphertext c as $c = m^e \mod n$

Decryption :

For a given ciphertext c;

Compute $c^e = (m^e)^d \mod n$ and use revers of the same padding algoritm to recover M.

One can clearly see that without a probabilistic feature, RSA based encryption algorithm is not IND-CPA secure since the encryption process is deterministic. Encryption of m_b is always equal to the c_b under the same encryption set up so IND-CPA and IND-CCA games become trivial.

4.2.2 Rabin Cryptosystem

Rabin cryptosystem is very similar to RSA cryptosystem except that it relies on the problem which is proved to be as hard as integer factorization problem.

Definition 4.5. Rabin cryptosystem works as follows:

Key generation :

First choose two distinct very large prime numbers p and q, preferably $p, q \equiv 3 \mod 4$

Compute $n = pq$, then the pair (p, q) is the private key pair and n is public key.

Encryption :

To send a message m convert it into an integer in the message space $M = \{0, 1, 2, ..., n-1\}$

Compute ciphertext c as $c = m^2 \mod n$

That is c is the quadratic residue of $m \mod n$

Decryption :

For a given ciphertext c;

Compute the square roots

$m_p = c^{\frac{1}{4}(p+1)} \mod p$ and $m_q = c^{\frac{1}{4}(q+1)} \mod q$

Then by applying extended Euclidean algorithm one can find y_p and y_q satisfying $y_p p + y_q q = 1$.

Then use chinese remainder theorem and find four square roots in the set $0, ..., n-1$.

Hence one of the square roots $\mod n$ will be the integer form of the message m.

Additional computation costs occur because of te decryption process. Hence this cryptosystem is not widely accepted. Because of the deterministic feature like RSA cryptosystem, Rabin cryptosystem is not IND-CPA secure. Encryption of m_b is always equal to the c_b under the same encryption set up so IND-CPA and IND-CCA games become trivial.

4.2.3 Goldwasser-Micali Cryptosystem

Goldwasser-Micali Cryptosystem[16] is developed by Shafi Goldwasser and Silvio Micali in 1982. It is the first provably secure probabilistic asymmetric encryption scheme under standard cryptographic assumptions. It is based on the quadratic residuosity problem which is defined as follows:

Definition 4.6. Let $a \in \mathbb{Z}_N^*$ where N is a positive integer.

a is said to be a quadratic residue modulo N if there exist an $x \in \mathbb{Z}_N^*$ such that $x^2 = a \mod N$ and x is a square root of $a \mod N$. If no such x exists, then a is called a quadratic nonresidue modulo N. The set of all quadratic nonresidues modulo N is denoted by Q_n and the set of all quadratic residues by $\bar{Q}_n$

Definition 4.7. Let $N = p$ be an odd prime, a is an integer such that $gcd(N; a) = 1$. Then the Legendre symbol is defined to be

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{if } a \in Q_p \\ -1 & \text{if } a \in \bar{Q}_p \end{cases}$$

It holds that for a given $a, b \in \mathbb{Z}_p^*$ and a prime p

29

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

Definition 4.8. For a composite $N = pq$ where p and q are primes, the Jacobi symbol is defined such that

$$\left(\frac{a}{N}\right) = \left(\frac{a}{p}\right)\left(\frac{a}{q}\right).$$

Definition 4.9. Quadratic residuosity problem which is known to be hard defined for a given $N = pq$ and $a \in \mathbb{Z}_N^*$ with

$$\left(\frac{a}{N}\right) = 1$$

desicion of a being quadratic residue modulo N or not.

Note that computing the jacobi symbol of a given pair $[a, N]$ is in polynomial time. So if p, q are known, then the problem is solved in polynomial time.

Definition 4.10. Depending on above problem, the encryption scheme is defined as follows:

Key generation :

First choose two distinct very large prime numbers p and q

Compute $N = pq$ and choose a quadratic nonresidue $y \in \mathbb{Z}_N$ satisfying $\left(\frac{y}{N}\right) = 1$

Then the public key p_k is the pair (N, y) and private key s_k is pair $p, q]$)

Encryption :

A message M having length n is encrypted bit by bit.

If $M = m_1 m_2 .... m_n$, then to encrypt m_i for $1 \le i \le n$

Choose a random value $r \in \mathbb{Z}_N^*$ such that the ciphertext $c = y^m r^2 \bmod N$

Decryption :

For a given ciphertext c;

Decrypt m_i as 0 if c is a square, otherwise equal to 1 using factors of $N = pq$ and the jacobi symbol

$$\left(\frac{y^m r^2}{N}\right) = \left(\frac{y^m r^2}{p}\right)\left(\frac{y^m r^2}{q}\right).$$

The GM cryptosystem is not an efficient encryption scheme since the encryption process is bit by bit but it satisfy IND-CPA security since for any bit of the plaintext randomly chosen value r from the group of units modulo N allows plaintext to have different ciphertext values for each time. [not sure]But if the message space is chosen different than 0,1 , scheme is not IND-CCA secure since one can simply for given plaintext p_1 and p_2 and challenge ciphertext c_b ask the last bit flipped version of c_b and see the p_b.

Later version of GB cryptosystem is Blum Goldwasser Cryptosystem which is proposed by Manuel Blum and Shafi Goldwasser in 1984[5]. Blum-Goldwasser cryptosystem is also CPA secure but not CCA secure since for given plaintext p, an XOR-based stream cipher using the Blum Blum Shub (BBS) pseudo-random number generator generates a different random sequence for each encryption process. For given plaintext p_1 and p_2 and challenge ciphertext $\bar{c}$, y,the attacker ask for decryption of $\bar{a}$, y and decides the challenge.

4.2.4 Paillier Cryptosystem

The Paillier Cryptosystem is a probabilistic asymmetric encryption scheme created by Pascal Paillier in 1999 based on the difficulty of computing n-th residue classes[13] which is the same in Okamoto-Uchiyama[24] and Naccache-Stern[21] Cryptosystems. It also assumes the decisional composite residuosity assumption to prove that encryption and decyrption algorithms work properly. It is a additive homomorphic cryptosystem which means for given a public key and messages m_1 and m_2, $m_1 + m_2$ can be encrypted.This feature allows it to be used in systems like electronic voting and threshold cryptosystems.

Definition 4.11. Original cryptosystem consists of a probabilistic key generation algorithm Gen, a probabilistic encryption algorithm Enc and a deterministic decryption algorithm Dec and work as follows:

Key generation :

> Choose two large primes p and q and calculate $n = pq$.
>
> Then select a semi-random, nonzero integer, $g \in \mathbb{Z}_{n^2}$ such that the order of g is a multiple of n in $\mathbb{Z}_{n^2}^*$.
>
> Set (n, g) as public key and (p, q) as private key pairs.

Encryption algorithm :

> To encrypt a message $m \in \mathbb{Z}_n$, choose a random number $r \in \mathbb{Z}_n^*$
>
> Compute ciphertext c as $c \equiv g^m r^n \bmod n^2$

Decryption algorithm :

> First calculate inverse of $\frac{(g^{\lambda(n)} \bmod n^2) - 1}{n}$ in $\mathbb{Z}_n^*$ call it μ.
>
> Then for a given ciphertext c, $m \equiv \frac{(c^{\lambda(n)} \bmod n^2) - 1}{n} \mu \bmod n$

The cryptosystem provides IND-CPA security assuming the decision of composite residuosity is hard since for given challenge messages m_1 and m_2 and challenge ciphertext c_b, the attacker can not have any advantage of extract information without making any decryption call since messages are encrypted with a fresh random number r in each encryption. However because of its homomorphic properties, the original system is not secure against chosen ciphertext attacks. The attacker asks for decryption of $c_i k^n$ for a small k and get m_i from the decryption oracle. However Pierre-Alain Fouque and David Pointcheval proposed an improvement that provides IND-CCA security[13]. In 2000, Ivan B. Damgård and Mads J. Jurik proposed[10] a generalised version, so called the Damgård–Jurik cryptosystem, and the previous system become a special case of the latter one by taking $n = 2$ instead of $s + 1$ for $s \in N$. This cryptosystem has also the same security levels with the original system.

4.3 Based on General Linear Code Problem

4.3.1 McEliece Cryptosystem

McEliece cryptosystem is developed by Robert McEliece in 1978 as an asymmetric encryption algorithm based on the hardness of decoding a general linear code which is an NP-hard problem. Algorithm is not widely used due to its key sizes(220kb for 128 bits of security)[4] but is a candidate for post-quantum cryptography since it is known to be resistant to attacks applying Shor's algorithm and measuring cost states. An error-correcting code (especially Binary Goppa codes in the original algorithm) is used as private key because of their easy decoding prodecure and by disguising the private key as a general linear code, the public key is constructed. There are variety of cryptosystems using the algorithm with different types of codes but most of them are insecure due to their structure of decoding. Wang proposed a secure version of the cryptosystem with security parameters depend on the hardness of decoding random linear code and also its dual version is introduced[22] by Herald Niederreiter in 1986.

Definition 4.12. The basic structure of key generation, encryption and decryption algorithms is as follows;

Key generation ;

Generate a k x n generator matrix G' of an irreducible binary Goppa code, where we assume that there is an efficient error-correction algorithm $Correct$ which can always correct up to w errors.

Generate a k x k random non-singular matrix S.

Generate a n x n random permutation matrix P.

Set $G = SG'P$.

Output public key $pk = (G, w)$ and secret key $sk = (S, G', P)$.

Encryption algorithm :

Take a plaintext $m \in \{0,1\}^k$ and the public key pk as input and outputs ciphertext $c = mG \oplus e$, where e En,w.[not clear]

Decryption algorithm :

Given a ciphertext c and secret key sk as input,

Compute $cP^{-1} = (mS)G' \oplus eP^{-1}$, where P^{-1} denotes the inverse matrix of P.

Compute $mS = Correct(cP^{-1})$.

Output $m = (mS)S^{-1}$.

The original cryptosystem is known to be insecure against chosen plaintext attacks since for the challenge ciphertext c_b for given challenge plaintext pair m_0, m_1, attacker computes $m_0G \oplus c_b$ and checks whether it is equal to w or not. In 2008 Nojima et al. present a IND-CPA secure construction by using padding with random value under standard assumptions. In this construction for every message m, a fresh random value r is concatenated, $r\|m$, hence for IND-CPA game, attacker is not able to decide the challenge ciphertext by only querying the encryption oracle. However the construction is not IND-CCA secure. Attacker can ask $c' = r'c = r'rm_iH$ to decrypt and get $r'm_i$ and hence m_i. Döttling et al. proposed an IND-CCA secure version of the scheme by their k-repetition PKE construction.[11]

4.4 Based on Elliptic Curve

4.4.1 Elliptic Curve Integrated Encryption Scheme

Elliptic curve integrated encryption scheme (ECIES) is proposed by Abdalla, Bellare, and Rogaway[1] using the following functions:

- Key Agreement (KA): Function used for the generation of a shared secret by two parties.

- Key Derivation Function (KDF): Mechanism that produces a set of keys from keying material and some optional parameters.

- Encryption (ENC): Symmetric encryption algorithm.

- Message Authentication Code (MAC): Data used in order to authenticate messages.

- Hash (HASH): Digest function, used within the KDF and the MAC functions.

ECIES has variations which are standarized[19] in ANSI X9.63 , IEEE 1363a , ISO/IEC 18033-2 and SEC 1. The scheme is proven to be IND-CCA secure in the original paper.

4.5 Methods Work For All Encryption Schemes

4.5.1 Optimal Asymmetric Encryption Padding (OAEP)

To deal with the security concerns in systems with RSA function, some padding algorithms were introduced. This solutions work for IND-CCA security as well as IND-CPA security. First of these padding algorithms which is provably secure was proposed by Bellare and Rogaway named Optimal Asymmetric Encryption Padding (OAEP)[2].This padding algorithm is an example of a All or Nothing Transform (AONT). AONT which is first introduced as a mode of operation for block ciphers, allows ciphertext to be understood only if the whole data is known.

Definition 4.13. Details of the algoritm is as follows:

Let f be trapdoor permutation, possibly RSA, k be the number of bits being encrypted as a block of total message M, and k_0 be chosen $\ni$ adversary's running time is notably smaller than 2^{k_0} steps.

The length of the message to encrypt $n = k - k_0$ is fixed and shorter messages can be padded to this fixed number by adding zeros.

Two hash functions G and H are chosen from the set of standard cryptographic hash functions such that

$$G : \{0,1\}^{k_0} \to \{0,1\}^n;$$

$$H : \{0,1\}^n \to \{0,1\}^{k_0};$$

To encrypt $m \in \{0,1\}^n$ random k_0-bit r is chosen and encryption processes as follows:
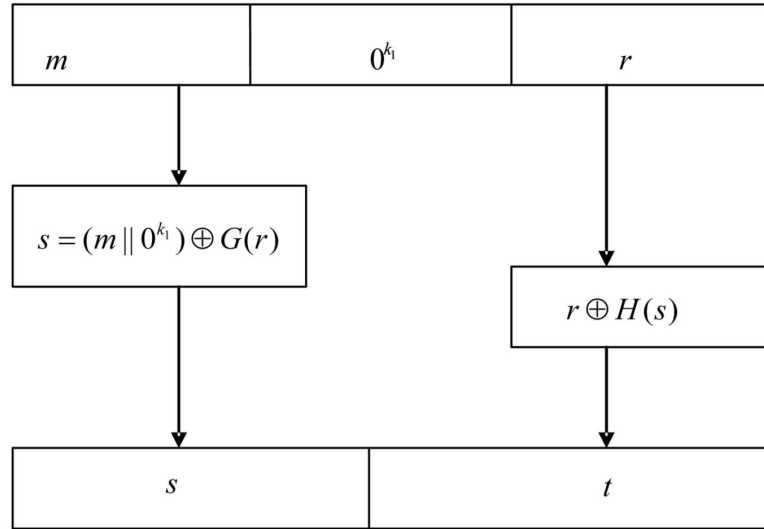
$$Enc^{G,H}(m) = f(s \,||\, r \oplus H(s) \text{ where } s = x \oplus G(r)$$

Concatenation is denoted by '$||$' and the decryption algorithm $Dec^{G,H}$ is defined as follows:

$$Dec^{G,H}(c) = s \,||\, t \text{ where } t = r \oplus H(s)$$

Calculate $H(s)$ to get r from t and then $G(r)$.

It follows that $s \oplus G(r)$ gives the message m.

It provides IND-CPA security for all trapdoor one-way functions and provides IND-CCA security if it is used with RSA function. [14]

After 7 years, Shoup point out a gap in OAEP security proof and introduce a new version of OAEP called OAEP+, along with complete security proof.He showed that OAEP is still secure though when using with RSA.[28]

Details of the algoritm is as follows:

Definition 4.14. Let f be a one-way trapdoor permutation, possibly RSA, k be the number of bits being encrypted as a block of total message M, and k_0 and k_1 be chosen such that adversary's running time is notably smaller than 2^{k_0} and 2^{k_1} steps.

The length of the message to encrypt n is fixed and shorter messages can be padded to this fixed number by adding zeros.

Three hash functions G, H and H' are chosen from the set of standard cryptographic hash functions such that

$$G : \{0,1\}^{k_0} \to \{0,1\}^n;$$
$$H' : \{0,1\}^{n+k_0} \to \{0,1\}^{k_1};$$
$$H : \{0,1\}^{n+k-1} \to \{0,1\}^{k_0};$$

To encrypt $m \in \{0,1\}^n$ random k_0-bit r is chosen and computes the followings:

$$s(G(r) \oplus x)\|H'(r\|x),$$
$$t = H(s) \oplus r,$$
$$w = s\|t,$$
$$y = f(w).$$
Then the ciphertext is y.

To decrypt a given ciphertext y, the decryption algorithm computes the followings:

$w = g(y),$
$s = w[0...n + k_1 - 1],$
$t = w[n + k_1...k],$
$r = H(s) \oplus t,$
$x = G(r) \oplus s[0...n - 1],$
$c = s[n...n + k_1 - 1]$

Then checks $c = H'(r\|x)$ or not and outputs x accordingly.

Simplified OAEP[6] introduced by Boneh and 3-round OAEP[25] by Pahn and Pointcheval are two more variations of OAEP where the first one is proved to be IND-CCA secure if the underlying asymmetric encryption algorithm is RSA or Rabin and the second one is IND-CCA secure with more encryption algorithms, like El-gamal,Paillier, with a relaxed IND-CCA assumption[18].

4.5.2 Fujisaki-Okamoto Transform

In 1999, Eiichiro Fujisaki and Tatsuaki Okamoto proposed[14] the first generic transformation for asymmetric encryption systems. rnsformation is the first generic transformation from an arbitrary one-way public-key encryption scheme to an IND-CCA secure asymmetric encryption scheme. But due to its inefficcient decryption process it is not widely used.

Definition 4.15. Consider a one-way asymmetric encryption scheme (Gen, Enc, Dec), two hash functions G and H having output lengths k_1-bits and k_2-bits respectively. Corresponding knowledge extractors T_G, T_H contains triples (R_j, m_j, H_j) and (R_i, m_i, G_i) respectively where $H_j = H(R_j, m_j)$ and $G_i = G(R_i, m_i)$.Let IND-CPA symmetric encryption algorithm $Enc_K^{symm}(m)$ having key length k_1 and message length k_1. Then the new scheme $(Gen^{fo}, Enc^{fo}, Dec^{fo})$ as follows:

Key generation :

Gen^{fo} generates key pair (s_k, p_k) uniformly random

Encryption algorithm :

Encryption algorithm $(Enc^{fo}(m; R, r))$ works for any k_1-bit message m and random values $R \in M$-message space.

Then computes $c_1 = Enc(R; H(R, m))$ and $c_2 = Enc_{G(R)}^{sym}(m)$.

The pair $C = (c_1, c_2)$ becomes the ciphertext

Decryption algorithm :

$Dec^{fo}(c_1, c_2)$ first picks up consistent pairs (R, m) from T_G, T_H
By re-encrypting c_1, c_2 using $H(R, m)$ and $G(R)$. If discovered outputs m
Otherwise output 'Reject'.

4.5.3 Rapid Enhanced-Security Asymmetric Cryptosystem Transform (REACT)

Rapid enhanced-security asymmetric cryptosystem transform (REACT)[23] was introduced by Tatsuaki Okamoto and David Pointcheval in 2001. Aim of this transformation is to be a fast generic transformation work with all weak public key cryptosystems like RSA and El-gamal in the random oracle model to reach high security levels as IND-CCA. REACT is integrated with block and stream ciphers to gain high speed rates. The transformation encrypts a session key with asymmetric encryption algorithm to be used in a semantically secure symmetric one in order to encrypt the main message. To reach IND-CCA security, a hash function added to the transformation. Asymmetric encryption algorithm needs to satisfy some conditions in order REACT to give desired IND-CCA security. [23] Okamoto et al. define a new attack model, called Plaintext Checking Attack (PCA), where the adversary can check the validity of the message-ciphertext pair (m, c).

Definition 4.16. First of all, a plaintext-checking oracle is defined so that an adversary feeds this oracle with inputs a message m and a ciphertext c, and the oracle returns 1 if c encrypts m, 0 otherwise. The oracle works fully adaptive,the attacker has always ask about a pair $(m.c)$ whether it is a valid or not.

A decryption oracle defined in IND-CCA games surely more powerful then this oracle and the adversary become too weak in an attack model with this oracle.To satisfy sufficient conditions for the asymmetric encryption algorithm used in the transformation, one-wayness notion is used to reach desired security levels.

One-wayness informally means,from the given ciphertext, one can not recover the whole plaintext.

Definition 4.17. An asymmetric encryption algorithm is said to be one-way if for any adversary A with a bounded running time, its inverting probability is less than a small number.

Then OW-PCA (one-way plaintext checking attack) is defined as follows.

Definition 4.18. Consider an asymmetric encryption scheme E(Enc,Dec),

  The attacker has always access to the PCA-oracle during the attack

  The attacker tries to recover the underlying encrypted message for a challenge ciphertext

  If succeeded with a probability less than a small number,the encryption scheme is OW-PCA secure otherwise not.

This security level is proposed for asymmetric encryption algorithms in the transformation, in order new transformed scheme to be IND-CCA secure.The basic conversion is s follows:

Definition 4.19. Consider an OW-PCA asymmetric encryption scheme (Gen, Enc, Dec), two hash functions G and H having output lengths k_1-bits and k_2-bitsrespectively and an IND-CPA symmetric encryption algorithm $Enc_K^{symm}(m)$ having keylength k_1 and message length k_1. Then the new scheme $(Gen^{react}, Enc^{react}, Dec^{react})$ as follows:

Key generation :

Gen^{react} generates key pair (s_k, p_k) uniformly random

Encryption algorithm :

Encryption algorithm $(Enc^{react}(m; R, r))$ works for any k_1-bit message m and random values $R \in M$-message space and a coin r.

Then computes $c_1 = Enc(R; r)$, $c_2 = Enc_K^{sym}(m)$ and $c_3 = H(R, m, c_1, c_2)$ where $K = G(R)$.

The triple $C = (c_1, c_2, c_3)$ becomes the ciphertext

Decryption algorithm :

$Dec^{react}(c_1, c_2, c_3)$ first calculates $Dec(c_1)$ and get R.

It verifies whether $R \in M$ or not. If so, recovers the session key $K = G(R)$ and $m = Enc_K^{sym}(c_2)$ which is returned if and only if

$c_3 = H(R, m, c_1, c_2)$ and $R \in M$.

Otherwise output 'Reject'.

4.5.4   Alternative Asymmetric Encryption Padding (AAEP)

Besides the preceeding padding algorithms for encryption schemes, REACT, OAEP and its variations, Schartner introduce[27] a low-cost padding algorithm in 2011 called alternative asymmetric encryption padding (AAEP). When concerning low-cost security hardwares and embedded schemes, because of the inefficiency of the implementations of hash functions on low-cost environments, OAEP and REACT are quite slow. Idea is to replace time consuming hash functions with symmetric encryptions which are implemented in hardware frequently.

Basic algorithm simply generates a randomizer r which has length k_0 and use it as a key for a symmetric encryption function Enc in CBC-mode and concatenates with the random r.

Definition 4.20. AAEP transforms the message m into m' as follows:

Generate randomizer $r \in_R \{0, 1\}^{k_0}$

$X = Enc(m)$

$Y = r$

$m' = AAEP(m) = X \| Y$

Basic scheme does not reach the IND-CCA security however since partial decryption is possible with the key r and two succeeding cipher blocks. A variation called $AAEP+$ is introduced in the paper so that AONS is satisfied and IND-CCA security is reached. Also AAEP+ is still faster than OAEP+ in low-cost environment[27].

Definition 4.21. AAEP+ transforms the message m into m' as follows:

Generate randomizer $r \in_R \{0,1\}^{k_0}$

$X = Enc(m)$

$Y = H(X) \oplus r$ for a hash function H having output length k_0

$m' = AAEP + (m) = X\|Y$

CHAPTER 5

CONCLUSION

In this thesis, we give the security definition for an encryption scheme in terms of infinitely many computational power world, namely perfect security by Shannon[7], then for a real world analogy, Goldwasser and Micali's semantic security is defined[16].

In first chapter, using indistinguishability notion and semantic security, the provable security games are introduces according to the adversarials capabilities. These IND-security games allowed us to analyse provable security of encryption schemes.

In the second chapter, we analyse the security of symmetric encryption schemes. Stream and block ciphers and their modes of operations are considered. We see that deterministic modes of operations are secure only for one-block use when they used once with the same key and security of modes of operations with initial values (IV) are dependent on the IV choices. In authenticated encryption modes, we see that they all aim to satisfy IND-CCA security with minimum possible security requirements of underlying confidentiality and authenticity tools.

In third chapter, based on the the mathematical hard problems, security of asymmetric encryption schemes are considered. First, based on discreet logarithm problem, we see that El-gamal cryptosystem satisfy IND-CPA and its successful extension Cramer-Shoup cryptosystem has IND-CCA secure version. Secondly, based on integer factorization problem, RSA, Rabin, Goldwasser-Micali and Paillier Cryptosystems are considered. We see that without a strong hash function involved, they do not satisfy IND-CCA security. Thirdly, McEliece cryptosystem and its variatons and extensions which are based on general linear code problem are considered. Finally, we give definitions of some methods and transformations that aim to upgrade security levels of above cryptosystems.

To conclude, in private-key encryption schemes, underlying blockcipher and IV choices are key point to reach desired security levels whereas in public-key encryption schemes, hash functions need to be carefully chosen together with the padding algorithms.

REFERENCES

[1] M. Abdalla, M. Bellare, and P. Rogaway, Dhaes: An encryption scheme based on the diffie-hellman problem., IACR Cryptology ePrint Archive, 1999, p. 7, 1999.

[2] M. Bellare and P. Rogaway, Optimal asymmetric encryption, in *Workshop on the Theory and Application of of Cryptographic Techniques*, pp. 92–111, Springer, 1994.

[3] M. Bellare, P. Rogaway, and D. Wagner, The eax mode of operation, in *International Workshop on Fast Software Encryption*, pp. 389–407, Springer, 2004.

[4] D. J. Bernstein, T. Chou, and P. Schwabe, Mcbits: fast constant-time code-based cryptography, in *International Workshop on Cryptographic Hardware and Embedded Systems*, pp. 250–272, Springer, 2013.

[5] M. Blum and S. Goldwasser, An efficient probabilistic public-key encryption scheme which hides all partial information, in *Workshop on the Theory and Application of Cryptographic Techniques*, pp. 289–299, Springer, 1984.

[6] D. Boneh, Simplified oaep for the rsa and rabin functions, in *Annual International Cryptology Conference*, pp. 275–291, Springer, 2001.

[7] D. Boneh and V. Shoup, A graduate course in applied cryptography, Version 0.3, from http://cryptobook. net, 2016.

[8] W. contributors, Block cipher mode of operation — wikipedia, the free encyclopedia, 2018, [Online; accessed 28-February-2018].

[9] R. Cramer and V. Shoup, A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack, in *Annual International Cryptology Conference*, pp. 13–25, Springer, 1998.

[10] I. Damgård and M. Jurik, A generalisation, a simpli. cation and some applications of paillier's probabilistic public-key system, in *International Workshop on Public Key Cryptography*, pp. 119–136, Springer, 2001.

[11] N. Dottling, R. Dowsley, J. Muller-Quade, and A. C. Nascimento, A cca2 secure variant of the mceliece cryptosystem, IEEE Transactions on Information Theory, 58(10), pp. 6672–6680, 2012.

[12] M. J. Dworkin, Recommendation for block cipher modes of operation: Galois/counter mode (gcm) and gmac, Technical report, 2007.

[13] P.-A. Fouque and D. Pointcheval, Threshold cryptosystems secure against chosen-ciphertext attacks, in *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 351–368, Springer, 2001.

[14] E. Fujisaki, T. Okamoto, D. Pointcheval, and J. Stern, Rsa-oaep is secure under the rsa assumption, in *Annual International Cryptology Conference*, pp. 260–274, Springer, 2001.

[15] S. Goldwasser and M. Bellare, Lecture notes on cryptography.

[16] S. Goldwasser and S. Micali, Probabilistic encryption, Journal of computer and system sciences, 28(2), pp. 270–299, 1984.

[17] J. Jonsson, On the security of ctr+ cbc-mac, in *International Workshop on Selected Areas in Cryptography*, pp. 76–93, Springer, 2002.

[18] E. Kiltz and K. Pietrzak, On the security of padding-based encryption schemes–or–why we cannot prove oaep secure in the standard model, in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 389–406, Springer, 2009.

[19] V. G. Martínez, L. H. Encinas, and C. S. Ávila, A survey of the elliptic curve integrated encryption scheme, ratio, 80(1024), pp. 160–223, 2010.

[20] D. McGrew and J. Viega, The galois/counter mode of operation (gcm), Submission to NIST Modes of Operation Process, 20, 2004.

[21] D. Naccache and J. Stern, A new public key cryptosystem based on higher residues, in *Proceedings of the 5th ACM conference on Computer and communications security*, pp. 59–66, ACM, 1998.

[22] H. Niederreiter, Knapsack-type cryptosystems and algebraic coding theory, Prob. Control and Inf. Theory, 15(2), pp. 159–166, 1986.

[23] T. Okamoto and D. Pointcheval, React: Rapid enhanced-security asymmetric cryptosystem transform, in *Cryptographers' Track at the RSA Conference*, pp. 159–174, Springer, 2001.

[24] T. Okamoto and S. Uchiyama, A new public-key cryptosystem as secure as factoring, in *International conference on the theory and applications of cryptographic techniques*, pp. 308–318, Springer, 1998.

[25] D. H. Phan and D. Pointcheval, Oaep 3-round: A generic and secure asymmetric encryption padding, in *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 63–77, Springer, 2004.

[26] P. Rogaway, Evaluation of some blockcipher modes of operation, Cryptography Research and Evaluation Committees (CRYPTREC) for the Government of Japan, 2011.

[27] P. Schartner, A low-cost alternative for oaep, in *Proceedings of the International Workshop on Security and Dependability for Resource Constrained Embedded Systemss*, p. 1, ACM, 2011.

[28] V. Shoup, Oaep reconsidered, in *Annual International Cryptology Conference*, pp. 239–259, Springer, 2001.