AN ENHANCED DISTRIBUTION TRANSFORMING ENCODER (DTE) OF THE HONEY ENCRYPTION SCHEME FOR REINFORCING TEXT-BASED ENCRYPTION

ABIODUN ESTHER OMOLARA

UNIVERSITI SAINS MALAYSIA

2020

AN ENHANCED DISTRIBUTION TRANSFORMING ENCODER (DTE) OF THE HONEY ENCRYPTION SCHEME FOR REINFORCING TEXT-BASED ENCRYPTION

by

ABIODUN ESTHER OMOLARA

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

March 2020

ACKNOWLEDGEMENT

I wish to convey my wholehearted gratitude to my supervisor, Assoc. Prof. Aman Bin Jantan for his time, helpful comments and invaluable guidance in the fine-art of scientific writing and research throughout the development of this dissertation. He supplied me with so many academic resources and his expertise, lectures, and insightful guidance were pivotal in completing this dissertation.

I am immensely thankful to my committee members, Prof. Bahari Belaton, Prof. Azman Samsudin, Dr. Zarul Fitri Zaaba, Dr. Manmeet Kaur Mahinderjit Singh and Dr. Mohammed FR Anbar for their invaluable help and guidance during my thesis development and presentations. Their extensive professional guidance, hard questions and expertise perspectives helped in reshaping and developing my research.

I wish to express sincere appreciation to the Center of Cyber Safety and Education for awarding me the $(ISC)^2$ Women Cybersecurity scholarship, 2017/2018. Thank you for believing in me and availing me the great opportunity to be part of the voice to fight Cyber Crime in our ever-evolving digital world.

I wish to express my admiration to my dear husband, Isaac Oludare Abiodun and beloved son, Tiwa Oludare for their unconditional love and tireless support. I am most thankful to my father, Elder. E.O Dada (JP) and my siblings, especially my twin sister, Kemi Dada, for their indefatigable love, prayers, and support. I remember with gratitude my late mother and I appreciate her effort for raising me to always go after my goals, you shall never be forgotten, *mama*.

I want to thank beautiful Ms. Sheela Muniandy who occupies a special place in my heart and have helped throughout the rigours of the logistics of this thesis. She has single handedly arranged and cater for all my official needs from the beginning of the development of this thesis until the end and I am really grateful for her kindness. Finally, I take this moment to sincerely thank the Dean and all the faculty members, lecturers, and my colleagues at the School of Computer Science, Universiti Sains Malaysia for providing me with an excellent academic environment throughout my Ph.D. study.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AES	Advanced Encryption Standard
AI	Artificial Intelligence
Alice	A placeholder name commonly used for the sender of a message
ASCII	American Standard Code for Information Interchange
BC	Before Christ
BiLSTM	Bidirectional LSTM
BiRNN	Bidirectional RNN
Bob	A placeholder name commonly used for the receiver of the message
CCA	Chosen Ciphertext Attack
CNN	Convolutional Neural Network
CS	Computer Science
DES	Data Encryption Standard
DH	Diffie Hellman
DLP	Discrete Logarithm Problem
DSA	Digital Signature Algorithm
DSM	Domain Specific Model
DSS	Digital Signature Standard
DTE	Distribution Transforming Encoder
ECC	Elliptic Curve System
ECDLP	Elliptic Curve Discrete Logarithm Problem
Eve	A placeholder name for the adversary
FPGA	Field Programmable Gate Array
GPG	Gnu Privacy Guard
GPU	Graphical Processing Unit
HE	Honey Encryption
HMM	Hidden Markov Model
ICM	Intent Classification Model
KMA	Known Message Attack
LSTM	Long Short-Term Memory
MR	Message Recovery
NIST	National Institute of Standards and Technology

NLBE	Natural Language Based Encoder
NLM	Neural Language Models
NLP	Natural Language Processing
NLTK	Natural Language Tool Kit
NSA	National Security Agency
PBE	Password-Based Encryption
PGP	Pretty Good Privacy
PKC	Public Key Cryptography
RNN	Recurrent Neural Network
RSA	Rivest-Shamir Adleman
S.MIME	Secure/Multipurpose Internet Mail Extensions
SPN	Substitution Permutation Network
TDNM	Target Distribution Non-malleability
TDSS	Target Distribution Semantic Security
UTF-8	Unicode Transformation Format 8
XOR	Exclusive-OR

LIST OF APPENDICES

PENGEDARAN PENGEKOD MENGUBALU (DTE) YANG DIPERTINGKAT SKEMA PENYULITAN MADU UNTUK MENGUATKAN PENYULITAN BERASASKAN TEKS

ABSTRAK

Skema Penyulitan Madu (HE) adalah strategi keselamatan maklumat yang digunakan sebagai pengukuhan kepada skema penyulitan konvensional untuk menangani serangan daya kasar khususnya dalam konteks sistem penyulitan berasaskan kata laluan Skema HE bergantung kepada model yang dikenali sebagai Pengekod Mengubah Pengedaran (DTE), yang menumpukan kepada penggunaan penipuan sebagai pendekatan pertahanan utama dalam reka bentuk primitif yang memudahkan komunikasi selamat dengan menghasilkan teks biasa yang palsu tetapi munasabah apabila menyahsulit dengan kunci yang salah Walau bagaimanapun, konsep skema HE dibatasi oleh kesesakan kegunaan dan oleh itu, gagal mencapai kes penggunaan dunia sebenar yang lain. Sebagai contoh, pengekodan mesej buatan manusia, seperti e-mel perlu menyesuaikan skema ke bahasa semula jadi yang bermaksud merangka semula pengekod deterministik semasa untuk bekerja di bawah tetapan sedemikian. Di samping itu, menimbulkan umpan untuk mesej yang dihasilkan oleh manusia memerlukan mesej yang masuk akal yang boleh menipu penyerang. Masalah ini masih belum dapat diselesaikan kerana beberapa penyelidikan untuk menambahbaik DTE gagal untuk menghasilkan mesej menipu yang munasabah. Tambahan pula, mereka gagal untuk memperkenalkan kerahsiaan pada mesej palsu sebagai kata kunci atau serpihan mesej teks biasa asas didedahkan semasa penyahsulitan, seterusnya membolehkan sistem itu menjadi serangan siferteks terpilih (CCA). Dua sumbangan utama dalam kerja ini adalah tindak balas terhadap dua masalah ini. Pengekod berasaskan bahasa semulajadi (NLBE) telah dibangunkan dan pendekatan untuk menyembunyikan kata kunci dari teks biasa dan menghasilkan mesej menipu yang munasabah diperkenalkan. Analisis eksperimen menggunakan simulator manusia sebagai piawaian emas menunjukkan kadar tidak dapat dibezakan adalah 94% dalam kes yang paling teruk di mana seorang musuh yang tidak terkawal dapat meneroka sistem dan kadar tidak dapat dibezakan 100% apabila diuji untuk simulator manusia dengan maklumat sampingan.

AN ENHANCED DISTRIBUTION TRANSFORMING ENCODER (DTE) OF THE HONEY ENCRYPTION SCHEME FOR REINFORCING TEXT-BASED ENCRYPTION

ABSTRACT

Honey Encryption (HE) is a cryptosystem used as a reinforcement to the conventional encryption scheme to address brute-force attacks specifically in the context of password-based encryption systems. The HE scheme relies on a model called the Distribution Transforming Encoder (DTE), which focuses on the use of deception as a key defensive approach in the design of primitives that facilitate information security by yielding plausible-looking but fake plaintext during decryption using an incorrect key. However, the concept of the HE scheme is limited by the bottleneck of applicability and thus fails to reach other real-world deployment use-cases. For instance, encoding a human-generated message such as email requires adapting the scheme to natural language which means re-designing the current deterministic encoder to generate plausible or realistic decoys message that can fool the attacker. This problem remains unsolved because of the few researches on enhancing the DTE fails to produce plausible decoy messages in human-language. Furthermore, they fail to introduce secrecy on the fake message as keywords or fragments of the underlying plaintext message are revealed during decryption, thus, enabling the system to a chosen-ciphertext attack where an attacker may use the results from prior decryption to inform their choices of which ciphertexts have decrypted. The two main contributions of this work are its responses to these two problems. A natural language-based encoder (NLBE) was developed and an approach for concealing the underlying plaintext and producing plausible decoy messages is presented. Experimental analysis using human simulators as the gold standard shows a 94%

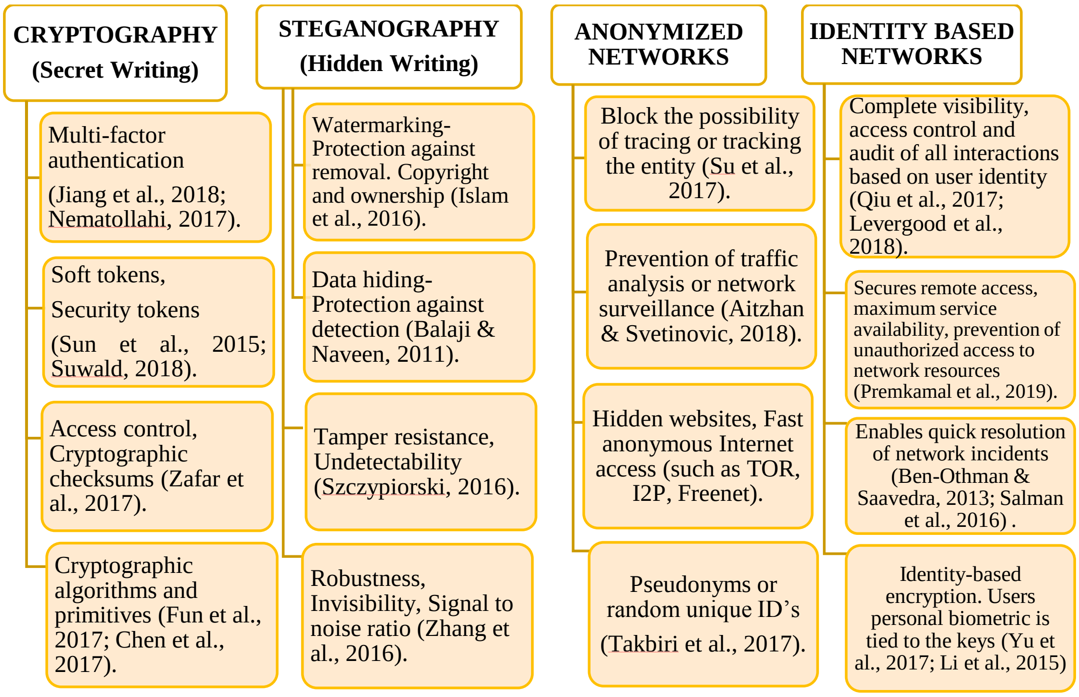
indistinguishability rate in the worst case where an unbounded adversary can explore the complete Oracle and a 100% indistinguishability rate when the keyspace is large enough.

CHAPTER 1

INTRODUCTION

1.1 Background of Research

Information security is a broad area which involves the security of diverse applications and categories of infrastructure. Various strategies of information security have been developed to achieve covertness of communications and alleviate the chances that the communicating parties have their conversation intercepted or eavesdropped on. Such strategies include ***secret writing*** referred to as ***cryptography***; ***hidden writing*** referred to as ***steganography***; anonymized networks; identity-based networks as depicted in Figure 1.1.



CRYPTOGRAPHY (Secret Writing)
Multi-factor authentication (Jiang et al., 2018; Nematollahi, 2017).
Soft tokens, Security tokens (Sun et al., 2015; Suwald, 2018).
Access control, Cryptographic checksums (Zafar et al., 2017).
Cryptographic algorithms and primitives (Fun et al., 2017; Chen et al., 2017).
STEGANOGRAPHY (Hidden Writing)
Watermarking- Protection against removal. Copyright and ownership (Islam et al., 2016).
Data hiding- Protection against detection (Balaji & Naveen, 2011).
Tamper resistance, Undetectability (Szczypiorski, 2016).
Robustness, Invisibility, Signal to noise ratio (Zhang et al., 2016).
ANONYMIZED NETWORKS
Block the possibility of tracing or tracking the entity (Su et al., 2017).
Prevention of traffic analysis or network surveillance (Aitzhan & Svetinovic, 2018).
Hidden websites, Fast anonymous Internet access (such as TOR, I2P, Freenet).
Pseudonyms or random unique ID's (Takbiri et al., 2017).
IDENTITY BASED NETWORKS
Complete visibility, access control and audit of all interactions based on user identity (Qiu et al., 2017; Levergood et al., 2018).
Secures remote access, maximum service availability, prevention of unauthorized access to network resources (Premkamal et al., 2019).
Enables quick resolution of network incidents (Ben-Othman & Saavedra, 2013; Salman et al., 2016) .
Identity-based encryption. Users personal biometric is tied to the keys (Yu et al., 2017; Li et al., 2015)

Figure 1.1 Current Measures to Address Communication Problems

Each one of the measures above (Figure 1.1) provides some form of information security against eavesdropping and interception during communication. Cryptography is one of the most important measures applied in the field of information security.

Cryptography is the science of designing secure communication schemes in the presence of third parties usually referred to as adversaries (Ahmed et al., 2012; Rivest, 1991). In this technique, confidential data/information is transmitted through an open network such that only the receiver who has the shared secret key can peruse/read the message which may be documents, phone conversations or data in other forms.

However, most conventional cryptographic schemes are susceptible to brute-force attacks given enough time and resources, such as an increase in computational power, skilled teams, money and advance in research. A brute-force attack is a trial and error technique used by attackers to decrypt a ciphertext. This attack searches for a single key, which returns plausible plaintext by computing all keys in the keyspace.

Messages are transmitted as bytes which are encoded in American Standard Code for Information Interchange (ASCII), Unicode Transformation Format-8 (UTF8) characters and so on. Consequently, an adversary that intercepts an enciphered text and tries to decrypt the message using incorrect keys can determine when he/she has recovered the message based on the output of the decrypted message (Kaliski, 2000; Bellare et al., 2012; Bonneau, 2012; Jo & Yoon, 2015; Kim & Yoon, 2016; Beunardeau et al., 2016; Lindholm, 2019).

Furthermore, conventional encryption schemes employ an n-bit key, where the security of encryption increases with the size of the key. Nevertheless, if a plaintext P is a 16-digit Mastercard number encoded via ASCII and the conventional encryption scheme is used for enciphering the card number, the probability that any Pi≠P is a valid

ASCII encoding of a 16-digit string which is negligible, at t $(10/256)^{16} < 2^{-74}$. Consequently, an adversary will discard incorrect messages and recover P with a high probability.

The advent of high-performance computing and high-processing tools such as GPU, FPGA also makes the brute-force attack to be highly successful as attackers become empowered with high computational power (Zonenberg, 2009; Cho et al., 2011; Moradi et al., 2011; Cousins et al., 2017; Coutinho et al., 2019).

Recently, an encryption scheme called Honey Encryption (HE) has been introduced by Juels & Ristenpart (2014a and 2014b) to reinforce the information-theoretic security of the conventional encryption scheme, i.e., a guaranteed threshold level of security regardless of the computational power of the attacker. HE is an encryption scheme that supplies a valid-looking but fake plaintext on an attempt to decrypt a ciphertext using incorrect keys. Consequently, an intruder employing a brute-force attack gain no information from guessing and checking of keys. The diagram, illustrated in Figure 1.2 describes a scenario where (a) the conventional encryption scheme and (b) the honey encryption scheme was used for the secure transmission of message respectively.

The sender encrypts the plaintext M with a key K, i.e., C = Enc(M, K). The encrypted message which is now the ciphertext C is sent to the receiver via a transmission channel. The receiver acquires the plaintext by decrypting the ciphertext using the correct key and a decryption algorithm M = Dec(C, K). During transmission, an adversary may intercept the ciphertext C with a definitive goal of learning the content of the plaintext M. He/she tries to get the message by guessing the keys used in securing the plaintext. The attacker in both instances gets a false plaintext when he/she uses a false key but can only tell if the attack was successful based on the output.

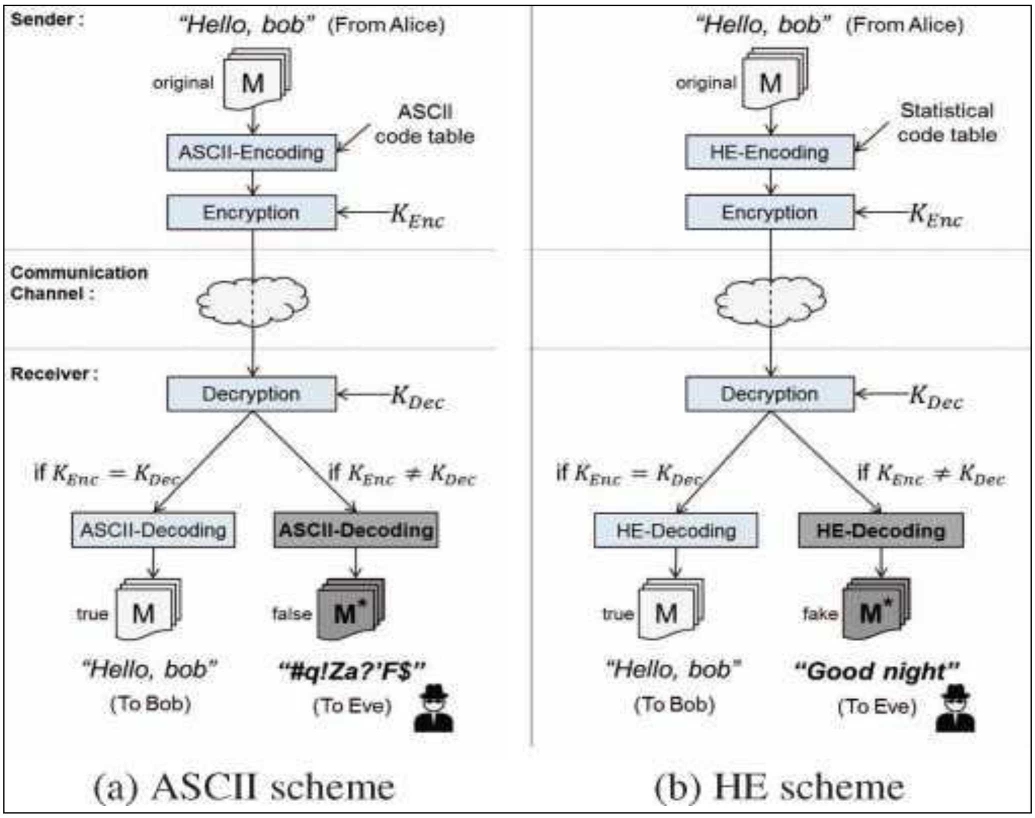


Figure 1.2 The Conventional Encryption and Honey Encryption Scheme

Adapted from Kim et al., (2016)

In the scenario described in Figure 1.2(a), the sender encodes a plaintext “Hello bob” using the conventional encryption scheme and sends it to the receiver. An attacker, Eve observing the transmission from the open channel intercepts the ciphertext and attempts to learn the content of the plaintext by brute-forcing the keys gets an invalid message *“#q!Za?’F$”*. This non-uniform structure (gibberish) is an indication to Eve that the key she is decrypting with is incorrect. Eve continues her quest for finding a readable message structure/distribution which is the plaintext by continuously submitting random passwords to the Oracle until she succeeds. In Figure 1.2(b), where the HE scheme is used to secure the message, Eve gets *“Good night”* when she tries an incorrect key. The fake but valid-looking structure (readable uniform distribution) of the message confuses Eve, making it impossible for her to tell if the attack was successful or not. Thus, she is stuck with bogus data in the event where she traverses the entire key space or Oracle. She cannot ascertain which of the bogus data is the plaintext, especially where she has no idea of the target message. Eve wins the game if she can determine the plaintext from the plausible (but decoy) messages accumulated during the message recovery attack.

1.2 Research Motivation

The major motivators of pursuing this research are as follows:

Reinforcing the security of conventional ciphers: Conventional schemes, for instance, the Advanced Encryption Standard (AES) is widely considered as one of the best and secure cryptosystem (Daemen & Rijmen, 2013; Shao et al., 2010). AES is responsible for a substantial amount of the information security that we benefit from daily. It is employed in the security of applications and products by giant institutions such as the NSA, Apple, Microsoft et cetera.

Standard security bodies such as the National Institute of Standards and Technology (NIST) validated and encouraged the use of the AES algorithm. However, it is susceptible to brute-force attack given enough time and resources (Alghazzawi et al., 2014; Deshpande et al., 2009; Grassi, 2018; Mukhopadhyay, 2009). Additionally, key recovery attacks (otherwise called biclique attacks) on AES which is faster than brute-force attacks by a factor of about 4 have seen the light of the day (Biryukov et al., 2010; Derbez et al., 2013; Bogdanov et al., 2011; Bogdanov et al., 2013). While the AES scheme is currently being used in securing various applications, it is only a matter of time before a complete cryptanalysis of the algorithm will be possible as computational power continues to increase.

The HE scheme provides resilience against brute-force attack (Juels and Ristenpart, 2014a & 2014b) and can be used alongside AES to reinforce conventional ciphers. For instance, in AES decryption, the cryptanalyst may have the foreknowledge of what a correctly decoded plaintext looks like since incorrect key outputs unreadable message while a correct key yields the plaintext which is readable (regardless of what the payload actually is). Additionally, the security of large chunks of data stored in the cloud remains a big challenge. The unification of AES and HE will be an alternative as it can provide a better security for cloud-based applications.

Application of HE for solving other economic problems: Honey Encryption has demonstrated its potential to reinforce conventional encryption scheme. However, there are limited applications of the scheme in addressing some economic problem. Enhancing the scheme will create new opportunities for addressing other problems. For instance, the eavesdropping attack on instant messaging (IM) system. Popular IM applications, such as WhatsApp, Telegram enforces end-to-end encryption using conventional encryption schemes. According to Statista in 2019, WhatsApp has emerged as the most popular IM service on Internet-enabled devices with over 1.5 billion monthly active users (Statista, 2019). Nevertheless, it fails to provide security as an eavesdropper can distinguish plausible chat message from random gibberish based on the keys he/she supplies during decryption. This flaw is exploited and backed by recent reports of loopholes in Whatsapp which potentially allows third parties to eavesdrop on encrypted Group chats, thereby learning the message being communicated (Greenberg, 2019; Paganini, 2019). Hence, adapting the HE scheme for encoding human-generated/human produced message will be a step ahead to allow the application of the scheme to other industrial use-cases for addressing other problems.

The increase in data-breaches: There have been massive cases of data breaches; from large scale to small-scale industries, large chunks of sensitive data fall into the hands of cybercriminals daily (Statista, 2019; Maria, 2019). According to recent analysis undertaken by Statista as of May 2019, the biggest online data breaches worldwide occurred from 2016 to date (Statista, 2019). Yahoo announced the most massive data breach in history in 2016, affecting 3 billion accounts. The year 2017 is not an exception as 143 million accounts were compromised, costing the company, Equifax a huge 90 billion US dollars (Komando, 2016; Verge, 2017; Statista, 2018; Statista, 2019) as shown in Figure 1.3.

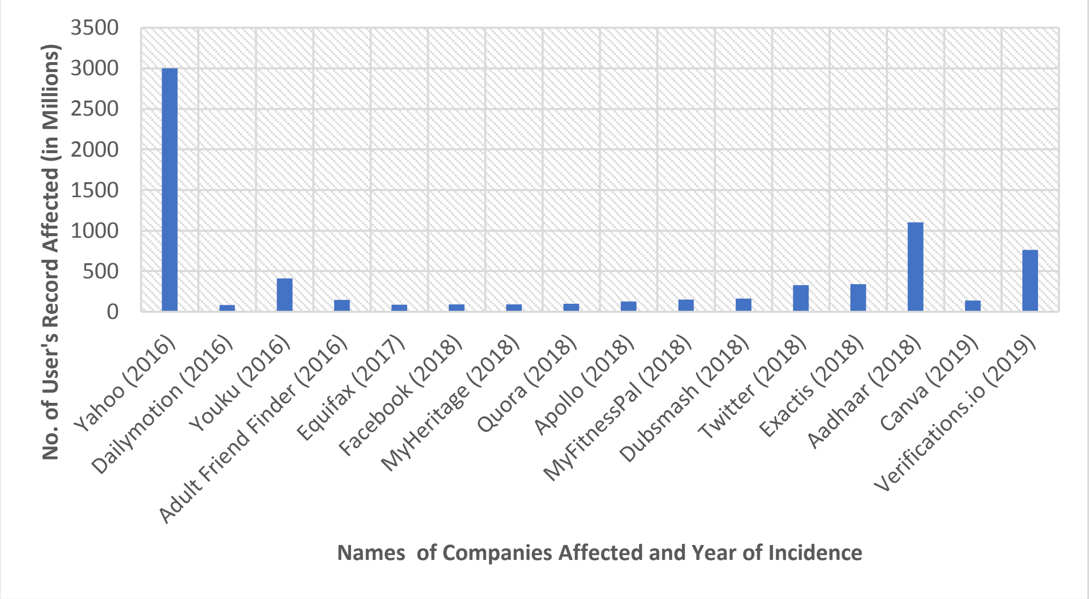


Figure 1.3 Geometric Increase in Data-breaches

Protection of data during phone calls, messaging, e-commerce, crypto-currency systems such as bitcoin, litecoin (Farhangi et al., 2010), smart grids such as smart appliances, smart meters (Vigna et al., 2016) and the Internet-of-Things remains a daunting task as security breaches continues to be on the increase. Thus, the compelling need and motivation to explore other strategies of enhancing the current measures used to ensure information security during communication.

1.3 Research Problem

The framework for building the HE scheme composes of a distribution transforming encoder (DTE) which is a pair of (encode/decode) algorithm followed by a pair of (encrypt/decrypt) algorithm using a conventional symmetric cryptographic scheme. A DTE is a randomized model of the encoding scheme tailored to the target distribution. It maps the space of plaintext messages to a seed space of n-bit strings by considering a probability distribution of the message space of the plaintext and specifies a proportional ratio of bit strings to the message (Juels & Ristenpart, 2014a; Juels & Ristenpart, 2014b; Tyagi et al., 2015).

The DTE construction presented by Juels and Ristenpart (2014) is specific for passwords protecting passwords (or passwords protecting keys). More precisely, low-entropy keys protecting high-entropy keys. Hence, it is more suitable for password management systems (Juels & Ristenpart, 2014a; Juels & Ristenpart, 2014b; Jaeger et al., 2016). While the current HE provides a modest and additional security to passwords, it creates a bottleneck of applicability. Thus, limiting the concept of HE from reaching other real-world deployment use-cases. For instance, encoding a reasonable-sized human-generated message, such as email, written-documents requires adapting the scheme to natural language which means re-designing the current deterministic encoder to work under such settings (Juels & Ristenpart, 2014a; Juels & Ristenpart, 2014b; Beunardeau et al., 2016). Indeed, the authors who proposed HE left as an open challenge, its application into other domains such as its adaptability to natural language message when they made this statement, "***...for human-generated***

messages (password vaults, e-mail, etc.) (...) is interesting as a natural language processing problem," (Juels & Ristenpart, 2014a; Juels & Ristenpart, 2014b).

Chatterjee et al. (2015) proposed the natural language encoders (NLE) to broaden the use cases of the HE scheme for its application to other fields such as the natural language. In 2016, Golla et al. improved Chatterjee's work by presenting the adaptive encoders. The presented models worked relatively well for password management systems such as PINS, passwords, biometrically extracted keys (as initially intended by the founders of HE). The significant contribution of Chatterjee and Golla was improvising the standard HE system by allowing its application for password vaults/manager. However, the proposed models fail to capture the empirical properties of language when extended for text encryption in human-generated messages. Hence, they yield messages that are semantically void and, in some cases, fail in the context of syntax. This invariably leads to messages that do not scale well and not realistic (Mainguy, 2014; Beunardeau et al., 2016; Deng et al., 2018).

Message recovery (MR) in HE describes the process where an unbounded adversary exhaustively tries likely keys to decrypt the ciphertext in order to learn partial information about the underlying plaintexts or to usefully maul the ciphertexts to recover a portion or all of the plaintext. However, MR security in HE has several inadequacies from the perspective of modern security goals for conventional symmetric encryption, and notwithstanding for the applications for which cryptographers have investigated the applicability of the HE scheme. Current HE scheme is susceptible to a Chosen-Ciphertext Attack (CCA). A CCA is an adversarial model for cryptanalysis in which the attacker collects information, at least in some part, by selecting a ciphertext and decrypting it by trying random or likely keys. This attack is guaranteed to be successful when parts or fragments of the plaintext message is revealed during decryption of the ciphertext. A practical attacker with a random Oracle can easily use the revealed parts of the underlying plaintext message to form other parts of the message to recover the plaintext and the key. This kind of attack is more successful when the attacker has some information about the plaintext. The application of state-of-the-art HE to the security of genomic materials by Huang et al., (2015) leaks most of the genome data when the adversary tries random keys based on the key distribution. All this begs for in-depth research into the HE scheme to verify if there are concrete constructions to help achieve better confidentiality for the underlying plaintext during a brute-force analysis process.

To achieve an MR security in HE, an adversary with the knowledge of the target distribution of the message and key distribution is given a challenge message. The challenge message is drawn from the target distribution, encoded under a key, and the subsequent ciphertext is given to the adversary. The adversary wins if it can output the challenge message. This security is achieved in a well-defined distribution such as PINS, Credit card numbers, and so forth. However, this whole situation becomes very difficult when applied to human-generated messages when important keywords of the structural information of the underlying plaintext is leaked during the brute-force analysis process.

An encoder that must convince an adversary requires the generation of a fake/honey message which is semantically and contextually realistic natural language good enough to fool humans and automated tools from telling them apart from real messages with high probability. Such encoder should be able to grasp the concept of recursivity, implying it must reflect to a large extent the syntactic and semantic correlation of human language to fit the long-range forward and backward interactions as used in human language.

Additionally, sizeable context-relevant information is needed as the source of data to fetch information and train the encoder to generate the decoy message. Contemporary cryptography hinges on Kerchoff's principle of obscurity on the key but not in the algorithm, implying that the source of decoy message must be public knowledge. Thus, the data source used must be exposed to the public. This leaves room for an attacker who has knowledge of the distribution of the data source.

1.4 Research Questions

The importance of this study will be summarized by four fundamental questions:

1. What method can be used to enhance the DTE of honey encryption for encoding/decoding of human-generated message?
2. What techniques can be incorporated into the HE scheme to prevent adversaries from learning partial information of the underlying plaintext which may lead to a chosen-ciphertext attack (CCA)?
3. What strategies can be employed to help generate convincing honey/decoy messages that can fool an attacker?
4. Is the proposed encoder semantically secured when applied in real-world systems?

1.5 Research Objectives

The ultimate objective of this research was to enhance the DTE of the honey encryption scheme for text encryption. Therefore, the research objective is divided into the following four sub-objectives:

1. To propose a method for developing a natural language-based encoder (NLBE) for secure encoding/decoding of human-generated message.

2. To present a technique to enhance message recovery security and to resist chosen-ciphertext attack (CCA). The proposed technique will prohibit an attacker from learning partial information of the underlying plaintexts or from usefully mauling ciphertexts.
3. To develop a strategy for generating convincing honey/decoy messages that can fool an adversary.
4. To evaluate the proposed models as proof of concepts to illustrate how the proposed natural language-based Honey encryption scheme can be realized within real-world systems.

1.6 Research Methodology

The key problem this research is addressing is enhancing the encoder of the HE scheme for the support of encoding natural language messages. The problem is divided into sub-problems and later integrated into one as a solution to the identified problem. A divide and conquer technique is used to solve each subproblem, after which the whole models are integrated. The study is divided into precisely five (5) phases which as presented in Figure 1.4.

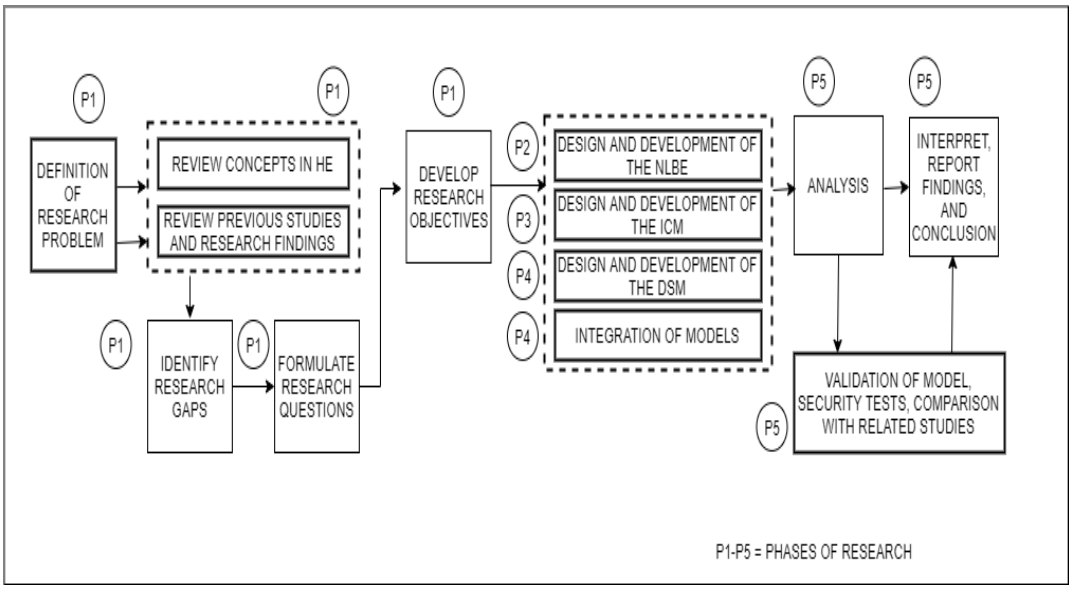

P1
DEFINITION OF RESEARCH PROBLEM
P1
REVIEW CONCEPTS IN HE
REVIEW PREVIOUS STUDIES AND RESEARCH FINDINGS
P1
IDENTIFY RESEARCH GAPS
P1
FORMULATE RESEARCH QUESTIONS
P1
DEVELOP RESEARCH OBJECTIVES
P2
DESIGN AND DEVELOPMENT OF THE NLBE
P3
DESIGN AND DEVELOPMENT OF THE ICM
P4
DESIGN AND DEVELOPMENT OF THE DSM
P4
INTEGRATION OF MODELS
P5
ANALYSIS
P5
INTERPRET, REPORT FINDINGS, AND CONCLUSION
P5
VALIDATION OF MODEL, SECURITY TESTS, COMPARISON WITH RELATED STUDIES
P1-P5 = PHASES OF RESEARCH

Figure 1.4 Overview of Research Methodology

Phase 1 is the preliminary investigation of the literature and problem formulation. This phase closes with finding the gaps from the research problem and developing the research objective of the study.

Phase 2 answers the research question 1 by proposing the natural-language-based encoder (NLBE). The NLBE allows the application of the encoder to a complex domain such as human-generated messages.

Phase 3 answers the research question 2 by proposing an intent classification model (ICM). The objective is to strengthen message recovery security in HE to allow resistance to a chosen ciphertext attack (CCA).

Phase 4 addresses research question 3 by proposing a domain-specific model (DSM). This phase is an extended research effort that is designed in par with the ICM in Phase 3 to further strengthen the natural language based HE scheme by restricting the decoy messages to a specific domain to completely fool the adversary.

Phase 5 addresses research question 4 by experimentally showing the feasibility of the application of the proposed NLBE in real-life systems using security tests, validations and evaluations of the proposed models.

1.7 Research Scope

The focus of this study is precisely on the DTE/encoder of the Honey encryption scheme. The encoder was extended to produce semantically secure decoy message for encoding human-generated messages to fool an adversary trying to acquire the plaintext during a brute-force attack. This study particularly considers textual content in English Language.

This research contributes specifically to confidentiality of messages in the domain of cryptography with a sub-domain of honey encryption, decoys and deception. Other moving parts of cryptosystems such as integrity and authentication, are outside the scope of this study. The encoders proposed in the previous studies are studied to investigate their application to other use-case in tackling issues related to information security within the context of this research. This dissertation leverages a composition of some approaches in natural language, deception-based systems to achieve the objectives proposed above. The intuition is to capture the rich semantic and syntactic feature of the human language the way it is being used and model it as a tool along with some techniques to fool the adversary into perceiving and accepting the decoy message as the real message.

1.8 Research Contribution

This research aims to extend the HE scheme to support its adaptation to natural-language. It does not eliminate the conventional encryption scheme but acts as a layer to strengthen the current conventional encryption scheme. To this end, the contributions of this thesis is presented concisely, and the noteworthy details are deferred to the future chapters. Please note that our contributions' descriptions in the following may appear technical and require some preliminary background which have been introduced and discussed in detail in Chapter 2.

1.8.1 Natural Language-based Encoder (NLBE)

Beunardeau et al., (2016) were among the first candidate in the literature to attempt adapting honey encryption to natural language. They consider the Corpus Quotation Distribution Transforming Encoder. Departing from previous works that employed n-gram generative Markov Model and Custom-trained probabilistic grammar model to restrict human activities to cases where decisions are increasingly constrained, for example, passwords (Chatterjee et al., 2015; Golla et al., 2016). However, their approach does not scale well and fail to model concise sentences when extended for real-life use-cases. The proposed natural language-based encoder (NLBE) models human language by allowing the encoding of human-generated messages and documents.

1.8.2 Intent Classification Model (ICM)

Jaeger et al., (2016) discusses how message recovery security in HE is a weak property. They consider the target distribution semantic-security (TDSS) and target-distribution non-malleability (TDNM) security notions. While their approach curtailed brute-force attack and increased message recovery security, it was designed in the context of password-based systems. In addition, they concluded that it is impossible to tighten the security of HE against a known message attack (KMA), where the adversary may have some side information about the plaintext. In this research, this shortcoming was scaled by building an intent classification model. The machine picks a sentence, disambiguates and re-encode it into a humanly understandable message which forms a grammatically and contextually correct message. To this end, the important keyword which an adversary with knowledge of the target information may exploit is extracted and message recovery is reduced significantly.

1.8.3 Domain Specific Model (DSM)

The proposed NLBE models sentences by generating a fake but connected message from the plaintext while preserving the contents and length of the original message. The NLBE gives significant protection to the underlying plaintext by preventing partial information leakage and resisting CCA or KMA attack on underlying plaintext. However, an adversary who submits random keys may find the string of plaintext being generated to be diverse from other decrypted text in a domain-specific system. Thus, the proposed NLBE protects the underlying plaintext but fails to yield convincing honey/decoy message to an informed adversary trying incorrect keys. To this end, if he is computationally unbounded, then he may discard the messages that fall out of his scope of expectation and recover the plaintext. Our proposed solution to this problem was to develop a domain-specific model (DSM).

1.8.4 NLBE in Real-world

The proposed models are tied up with a proof of concept to depict how the system can be realized within a real-world deployment. Aside from using the ICM and DSM to improve the NLBE in terms of security, the models also supplied some features to capture the recursivity of human language and also to capture contexts in the decoy message that will be generated by the encoder. The enhancement made in the proposed encoder increased the use-cases as it can be incorporated into various realistic settings. Towards the end of this study, papers were published as contribution showing the application of the proposed scheme in realistic settings such as, the security of email systems, instant messaging systems, medical systems and other security cases that require human data.

1.9 Summary of Thesis Organization

This dissertation is a collection of the concept, ideas and implementations published in a number of papers and conferences. This section gives a summary of the structure of each chapter presented in this thesis.

Chapter 1 - provided an overview of the research content including research background, motivation, problem statement, and other details. These details are essential to understanding the existing problems that motivated and necessitated this research.

Chapter 2 – presents the necessary background and an overview of related work. This chapter preludes with a general overview of the existing cryptographic schemes and ends with comprehensive details on the target study, honey encryption scheme. Honey encryption is a relatively new concept introduced in the year 2014 and as such, we discussed all the paper found since its inception until now. This chapter illuminates the gap that exists in the current research in Honey encryption which will be filled by the findings of this study.

Chapter 3 –provides the entire step-by-step details employed in accomplishing this research along with the algorithms developed in the course of the research.

Chapter 4 - discusses the implementation details of the proposed model.

Chapter 5 –presents the analysis and result of this study with regards to the research questions and research problems which this study has attempted to solve. This phase is important to ensure the validity of the proposed encoder.

Chapter 6 - concludes this dissertation with open questions pointing to the future directions of the research.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This chapter consists of two main sections: the essential background of cryptography and literature review of the HE scheme. We briefly discuss the two basic cryptographic primitives and narrow it down to Honey encryption which is the central focus of this study. Discussion on the background of cryptography is required for readers to understand the context behind the HE scheme.

Honey encryption connects cryptography with diverse computer science domains, such as decoy systems, natural language processing, cognitive sciences and others. By interfacing the intuition of decoys to the machinery of cryptography, HE translates the creative defensive tactic of decoys, traditionally the concept of system security, into cryptographic theory and practice. Thus, addressing eavesdropping, interception, intractable attacks such as the exhaustive key search where the adversary is computationally unbounded. Honey encryption is a budding encryption paradigm in its infancy and as such, related works to the scope of this research are few. However, this study covers all substantial research that has been proposed since the inception of the HE scheme and some studies on related decoy-based schemes. A generic preview of the road map of this chapter is depicted in Figure 2.1.

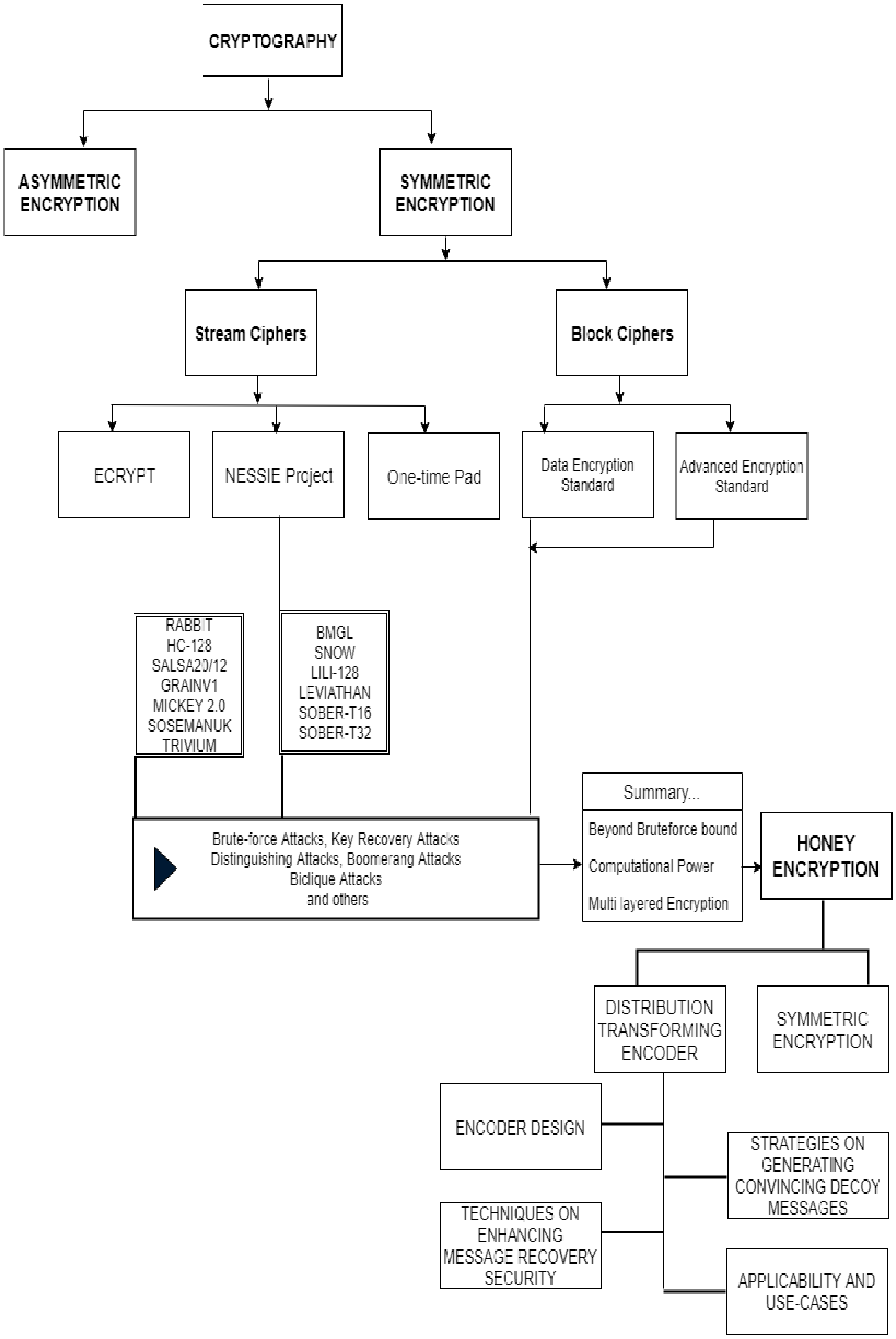


Figure 2.1 Roadmap for Literature Review

2.2 Introduction

The history of cryptology can be traced to ancient Egypt (as far back as 4500 years ago) as an art used by the military and government for obfuscating secret message in order to protect the message from being read by an adversary that intercepts the message carrier during its transmission. Before World War I, cryptography was traditionally deployed by the government, military circles, and secret agents for message confidentiality. An extensive treatment of the early history of cryptography can be found in the monograph of Kahn (1967). Furthermore, a comprehensive and exhaustive reference elucidating its evolution from traditional to modern cryptography can be found in (Kerckhoffs, 1883; Shannon 1948, Shannon 1949; Kahn, 1967; Menezes et al., 1996; Lindell & Katz, 2014).

The art was later transformed into science in 1948 by Shannon in his seminal work on a mathematical model for cryptography (Shannon, 1948). This transition was apparently the advent of modern cryptology, in which checking the **integrity** of the messages and verifying (**authentication**) the identities of the communicating parties have progressed toward becoming as vital as guaranteeing message secrecy (**confidentiality**).

Encryption is a process of transforming a ***plaintext*** message into an ‘unintelligible’ ***ciphertext*** form and decryption is the process of recovering the message by transforming the ciphertext into the plaintext. The sender (given the placeholder name Alice) transforms the plaintext into a ciphertext using a key and an algorithm (referred to as cipher) and sends it to the intended recipient (given the placeholder name Bob) across an insecure communication channel that may be controlled by an adversary (given the placeholder name Eve).

Cryptographic systems can be broadly classified into asymmetric and symmetric encryption schemes. The next subsections present some brief descriptions on both systems before progressing into the specific area of study; honey encryption.

2.2.1 Asymmetric Cryptography

In 1976, Diffie and Hellman established the Public-key cryptography (PKC) also known as asymmetric cryptography (Diffie and Hellman, 1976). In asymmetric cryptography, each communicating party has a pair of cryptographic keys – a public encryption key and a private decryption key. The keys are essentially large numbers that have been paired together yet are not identical (distinguishable). One key in the pair can be shared with everyone; which is referred to as the public key. The other pair of the key is kept secret and it is referred to as the private key. Any person can encrypt a message using the recipient's public key but the message can be recovered by only the person who has the corresponding private key. The basic concept of the PKC cryptosystem is depicted in Figure 2.2.

Bob has a pair of key, the public key $P_{k,}$ (which is publicly disseminated) and a private key S_k (secret). For Alice (or others) to send a message M to Bob, she uses Bob's public key to encrypt the message M as input in an encryption function, enc() and generates a ciphertext C to Bob,

$$C \leftarrow enc(P_k, M)$$

Bob on his part will recover the messages M from the ciphertext C by applying an inverse transformation function on the ciphertext C dec() using his private key, S_k as input to decrypt and acquire the plaintext.

$$M \leftarrow dec(S_k, C)$$

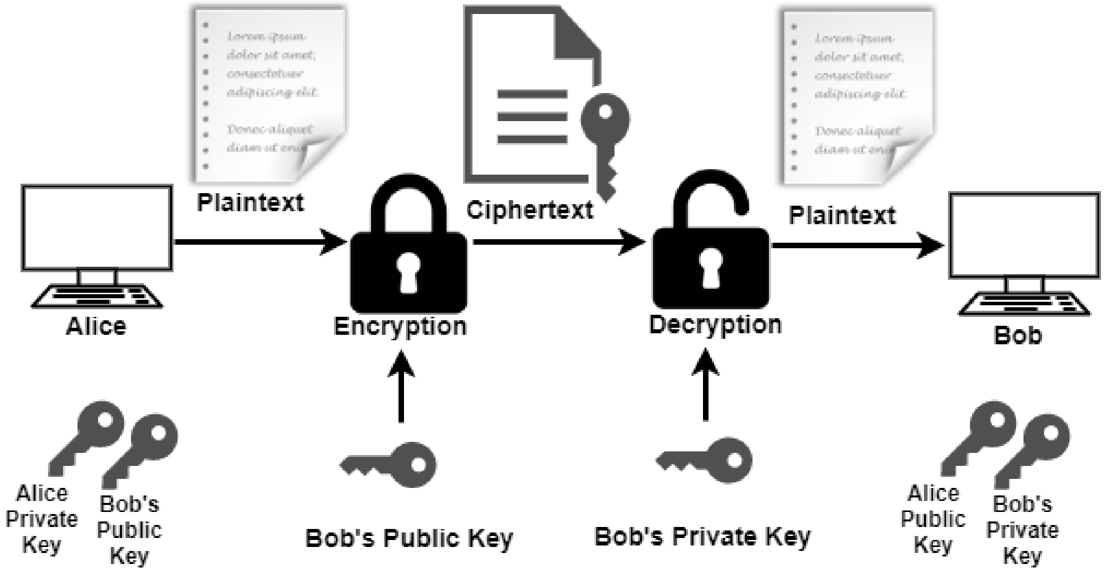

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Donec aliquet diam ut eni
Plaintext
Ciphertext
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Donec aliquet diam ut eni
Plaintext
Alice
Encryption
Decryption
Bob
Alice Private Key
Bob's Public Key
Bob's Public Key
Bob's Private Key
Alice Public Key
Bob's Private Key

Figure 2.2 A Simplified Model of Asymmetric Cryptosystem

A simple analogy to use in describing a PKC cryptosystem is a mailbox. In this mailbox, every person can place a letter in the box, but only the person who holds the key can open/unlock the mailbox and recover the letters. Popular variants of PKC cryptosystem often employed in real-world setting includes; Rivest Shamir Adleman (RSA) cryptosystem (Rivest et al., 1978), the elliptic-curve cryptosystem (Koblitz, 1987), ElGamal cryptosystem (Elgamal, 1985) and others.

The security of public-key cryptosystem depends on the assumption that it is computationally impossible to compute the private key from the public key. Such security postulation is guaranteed by computational hard mathematical problems, such as the integer factorization problem, in the case of RSA, elliptic-curve discrete logarithm problem, in the case of the Elliptic Curve Discrete Logarithm Problem (ECDLP) cryptosystem, discrete logarithm problem as in the case of ElGamal cryptosystems and several hard scientific problems defined over lattices in Lattice-based cryptosystems. The PKC encryption schemes provide a seamless way of encrypting as the sending and receiving parties need not share a priori key (as in the case of symmetric encryption schemes).

However, most PKC cryptosystems are computationally costly and are very slow hence, they are mostly used to agree on a key. For instance, the RSA cryptosystem provides a conceptually simple encryption scheme, however, its major drawbacks are its large key size and slow operation. On the other hand, elliptic-curve cryptosystem requires much smaller key size but its computational cost is high.

Public-key cryptography continues to attract a plethora of research activity in the cryptographic scene. This is because they represent the essential security ingredient in the design and implementation of cryptosystems in applications. Several Internet standards, such as the Transport Layer Security (TLS), Gnu Privacy Guard (GPG), Secure/Multipurpose Internet Mail Extensions (S/MIME) and Pretty Good Privacy (PGP) depends on PKC for security and functionality (Blake et al., 2006; Lucas, 2006; Du et al., 2009; Kurniawan et al., 2011). The current implementation of PKC cryptosystems under the present-day classical computer is considered secure when the key size is sufficiently large and critical countermeasures against side-channel and fault attacks are taken into account. Nevertheless, this condition falters in the domain of quantum computing. In 1994, Shor designed an algorithm on a quantum computer that solves the integer factorization and discrete logarithm problem in a few seconds which renders the PKC cryptosystems insecure in the advent of quantum computers (Shor, 1994). Although, there is no known ground-breaking quantum computer in the world today, yet several high-tech organizations, such as Google, Microsoft (Pradeep, 2019) are highly involved with building quantum computers due to its potential features.