Technium Vol. 5, pp.21-28 (2023)
ISSN: 2668-778X
www.techniumscience.com

A New Proposed Public Key Cryptography Based on Bio Strands

Auday H. AL-Wattar*

University of Mosul
ahsa.alwattar@uomosul.edu.iq

Abstract. The recent rapid advancement of technology has increased the capability of attackers. The main challenge to information security is the requirement for using unconventional philosophies and alternative means and focusing on new aspects to achieve security. This article proposes a new method that uses the collected genetic information on GenBank and its characteristics. The statistics that were calculated for the data that were hidden using this method prove that they meet the security standards. This paper employs unique elements for achieving information hiding based on that information.

Keywords. Steganography, hiding information, Public key, and GenBank.

1. Introduction

Computer security is a broad term that refers to actions, techniques, procedures, and technologies to preserve, safeguard, and defend computer systems' information and data by restricting unauthorized access to systems.

A secure connection is required for every entity to exchange data reliably. The internet has served as the foundation for all e-business and finance activities. The rise of the Internet of Things has introduced substantial security concerns centered on identifying acceptable approaches to achieve security, mainly because the IoT demands a unique environment, and specific requirements must be considered. Many strategies and systems have been created within traditional steganographic methods to meet these security criteria, specifically in the theoretical area of cryptographic protocols. Most research is concerned with Hiding in providing security for the Internet of Things. Since rare studies have been involved with the use of steganography in providing IoT security

This paper proposes a novel process that uses GenBank DNA data as steganography.

Due to its indispensable nature in modern society, data security has perpetually occupied a preeminent position on the list of top priorities. As computers have become increasingly prevalent in everyday life, so too has this fascination. The term "data security" is used to describe a wide range of activities, strategies, and resources that aim to keep intruders out of computer systems and their data and information. If two parties are going to be able to exchange information with complete confidence, they must use a secure connection. All online financial and retail transactions depend on the reliability of the internet. Particularly given that the IoT necessitates a distinct environment and specialized circumstances must be considered, the advent of the Internet of Things has introduced important security issues, focused on identifying acceptable approaches to accomplish security. In the mathematical realm of data hiding and extraction, various approaches and systems have been created within traditional steganographic methods to fulfill these security needs. These strategies are defeated using new steganography algorithms. Scholars have tried biotech steganography. This article proposes GenBank DNA data as novel public-key steganography.

1.1. Steganography

Steganography is the art that comprises communicating undisclosed information in a suitable transporter; that is to say, it is the method of embedding data (message) inside another file [1]. Steganography has several valuable tenders. Undisclosed communications where private data could be sent without concern of drawing attention to the threat from possible invaders[2]. Conventionally, steganography is recognized as a technique allowing two or more parties to create a secret message over an insecure channel that is observable to snooping. A significant area of security goals is achieved by tools of steganographic techniques [3].

1.2. GenBank

In [4], "GenBank® is a comprehensive public database that provides publicly available nucleotide sequences to enable bibliographic and biological notations." NCBI's GenBank is part of the International Nucleotide Sequence Database Collaboration (ENA), which comprises GenBank at NCBI and the DNA Data Bank of Japan (DDBJ).

The National Center for Biotechnology Information (NCBI), a division of the National Library of Medicine (NLM), is responsible for its development and dissemination on the Bethesda, Maryland, campus of the National Institutes of Health (NIH) [5].

Genome shotgun (WGS) data and other high-throughput sequencing data from sequencing centers are the primary sources of NCBI's GenBank. Additionally, the US Patent and Trademark Office makes patent sequences available. GenBank collaborates with the EMBL-EBI European Nucleotide Archives (ENA) and the DNA Data Bank of Japan (DDBJ) as part of the International Nucleotide Sequence Database Collaboration (INSDC) [6]. There are monthly meetings amongst the INSDC partners to keep the global sequence of information collections consistent and comprehensive. GenBank data can be accessed for free via the internet, FTP, and a wide range of web-based tools for analysis and recovery [[7, 8] from the NCBI .

Steganography techniques that use segmental arithmetic or are based on biological workroom research aren't fit for digital computing environments, for example. A new, secure steganography approach is proposed, and its performance is evaluated. To conceal information, this steganography makes use of GenBank data segments.

1.3. Steganography Public key principle

Two or more parties who have never met or shared a secret may use the public-key steganography protocol informally to communicate secret messages over a public channel without the adversary being able to discern their existence[9, 10].:

Party Alice can hide a message using Bob's public key if he wants to interact with Party Bob confidentially. Only Bob can unembed such communication as only Bob had access to the respective private key. This can be demonstrated in Figure 1,

As follows:

Technium Vol. 5, pp.21-28 (2023)
ISSN: 2668-778X
www.techniumscience.com

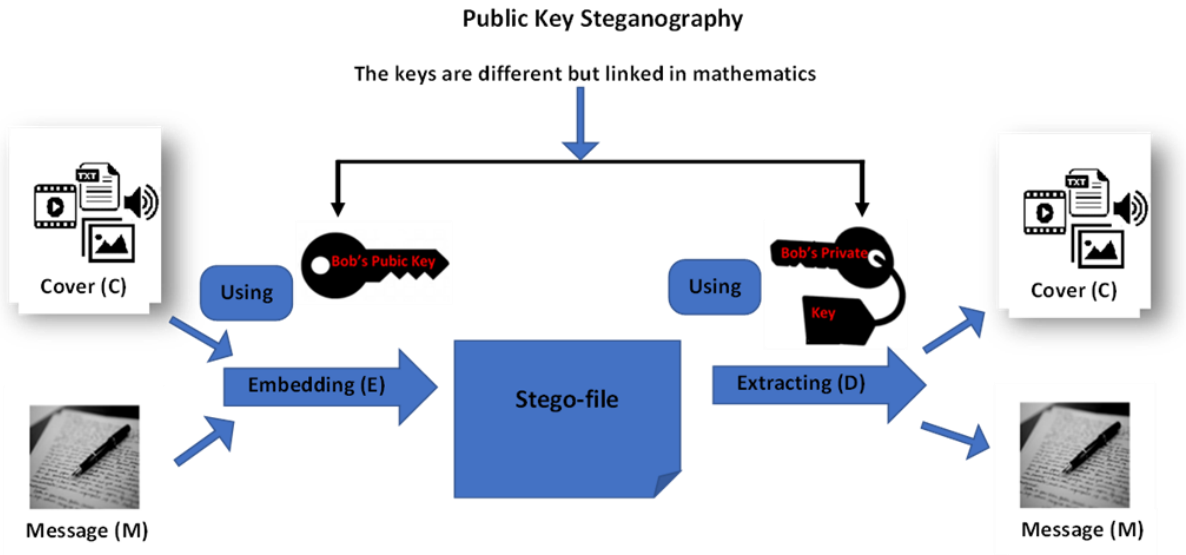

Figure 1. Public key Steganography

Formally:
f(x) is a one-way function from a set X set Y so that f(x) is easily computed by all x ϵ X, but it is "computationally ineffective," to locate any x ϵ X, such as f(x) = y for "fundamentally all" elements y ϵ Y.

1.4. Man in the middle attack (MIMT)

A MITM attack occurs when a hostile third-party intercepts data travelling from a transmitter to a receiver and then maliciously modifies the data before sending it on to the receiver. The consequences of this MITM attack, which involves transmitting false information via a network, are severe[11, 12].

1.5. Black hole attack

In most cases, the black hole assault is a denial-of-service attack, often known as a DoS attack, and is one of the most apparent types of attacks. The Black Node appeared during the process of determining the best route to take; initially, the sender was unaware of the most direct route to the receiver. A malicious node used its routing protocol to announce that the node obtains the shortest way to the target node, although there is no route to the receiver of the black hole node. This caused the target node to be unable to receive information from the black hole node. The black hole node is present along the data channel in this particular scenario; the black hole node is current along the data channel. If the routes have been established, the transmitter will deliver the packets to the black hole node, where it will then begin dropping the packets without sending them to the target node [13, 14].

2. The Proposed Method

Figure 2 illustrates a protocol between two entities (Alice and Bob) that can be used to describe the proposed technique. The following are the components that make up the suggested procedure: -

- (Alice) - Transmitter.
- (Bob)- Recipient.
- Message (m): Represents the data to be hidden.
- The cover (c): The file in which the message is hidden.
- Stego-File (St): The file after hiding the message in it.
- GenBank's: DNA banks.
- Pi: The segment with the index which is included in GenBank.

- Sec_Bob: Recipient's secret key.
- Pub_Bob: Recipient's public key.

DNA Databases can be found within GenBank. There is a specific position and value associated with each segment (several bases with a specific length). Within the framework of the proposed approach, the private and public keys of the receiver will be derived from the locations of the values of the segments (Bob). In the next parts, the scenario of the job will be discussed in greater depth.

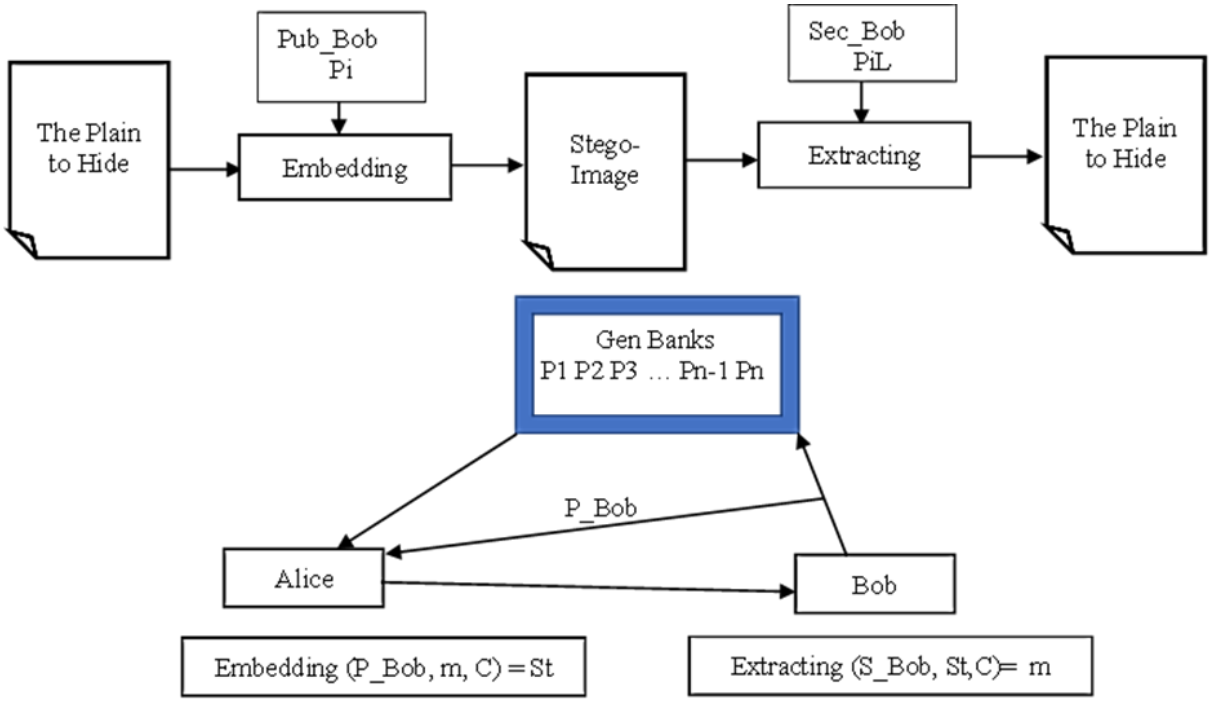

Figure 2. The scenario of the proposed method

The position of the selected DNA segment will be denoted by the public key that is assigned to the receiver (P_Bob).while Bob's secret key (Sec_Bob) may be one or more DNA segments (P1, P2,..., Pn) within the GenBank. This would mean that the value of the selected DNA segment would have a certain number of DNA bases and a certain length. While the position of the selected DNA segment will be denoted by the public key that is assigned to the receiver (P_Bob). The following is an overview of the possible combinations for the pairs of keys: -

- The segment's position inside GenBank provides Bob's public key.
- The slice's value reflects Bob's secret key (Sec_Bob).
- Only Bob could determine Sec Bob using DNA segments.
- Sec_Bob is now a GenBank DNA segment with a value and length that can be any sequence of DNA bases.

Figure 3 shows the proposed keys (Pub_Bob) and (Sec_Bob).

As Sec_Bob PiL.

Technium Vol. 5, pp.21-28 (2023)
ISSN: 2668-778X
www.techniumscience.com

Where:
Pi is the GenBank DNA slice P, and L is its length.
The selected DNA segment's length determines Sec_Bob. The recipient can obtain this key using many approaches. Bob may use an accurate mathematical way to retrieve a particular segment or a chaotic way to retrieve a particular part or portion of the segment. This might allow random access to segment bases. The receiver can produce or utilize the key using any method. Bob knows the key, but Alice doesn't.
Alice knows the receiver's public key, which is the DNA segment's GenBank location as HMD (Pi).
Where HMD (Pi) Pi's GenBank location
Message m may be extracted using the paired public and private keys (the DNA segment's GenBank location (Pub_Bob) and its bases) (Sec_Bob). DNA segment size varies by the party (sender and receiver).

	Locations of DNA segments within GenBank HMD(Pi) Pub_Bob							
	P_1	P_2	P_3	P_4	. P_5.P_6.P7...P_{n-1}			P_n
Value of DNA Segment (PiL) Sec_Bob	atattcgtcttatc	tggctcgttatatc	actaactactccc	gcgcttatgtgca	ccggc	.....	aaatg	tagaatattgtggc
	tgacgtgccgttc	tggccctcgtgca	agcatagtcatcc	ttacacggcatgt	ggagt	.....	ctcct	gtatggtcagattt
	atatgctgagcgc	tggacgtctcatc	tccgcacgagtct	acggacgcaagc	cgctg	.....	gggca	cgaacccaaaaca
	agctggcatgaca	accgccgtgctgg	acttaactagtga	taccttcattgcgt	acggc	.....	ctgct	ttgtgggggcaca
	tgtaatagatatc	ttatcgaccaacg	aacacgcaggca	ggacatgccggcc	ttgac	.....	gggcc	gacgctgaggtac
	taaaaataggtat	gttttttattcaag	tatttcggcgagg	tgcgttagacacc	ttgct	.....	ctatc	ggcatgcatgaat
	agacgttagactt	ggagcttttctcgt	tcacagtcgccgt	gccctcaaaaag	cgcct	.....	tgcgc	cgtcgaagtttgct
	cacggcattctat	tctcgtggttgcgc	catcacactggac	cagacacatacg	tgcac	.....	cctgt	agcttgccaggag
	cccaggctgaatt	gttcgtcgtctacg	gcttgagtatgtat	aagtccagccccg	atctc	.....	cctga	attggcgaaccat
	gcttatcagagcg	atcgagttcttcg	ggacgtcaatag	tatcggtcgaagg	caagg	.....	gatgg	gcaacaggactgg
	ggagaggagccc	aaggtcaaaatct	gggcaagagcag	gccggattttagt	cccct	.....	ctaca	aattggcgtgttat
	ctctctgttatgcc	tctttgttgctgag	tgcgggagctacg	actactcccccatt	tagtg	.....	aaggt	cagcagctttgcta
	ctagaccaggaa	cggaaatgtcaa	gaatagctttcgg	agaccagggagt	cggtc	.....	aaggt	agccgtggctgttt
	aggaatccaattg	gatttacgagcgc	agtatttagagga	ttaacaacaggg	aagct	.....	taccc	tcgaaccaagcaa
	gaacgtcttgcca	gaatccgccatgg	tgggggtcttgca	atgtccaacagta	ccttg	.....	tttca	gtaaacacgcgcg
	gcccggcggcgg	aacaatggtacac	atataagttctca	tctttatgccagg	gtcta	.....	caaat	gcacccctcatcgt
	actcatgcaagaa	tttcagacccgta	actcgctcggtgtt	gtattcctcgctca	gagag	.....	acata	acttcgttcgtggg
	ttacatagcggat	ggtccgcattact	gaaggatatttga	catcacatttccct	tatgg	.....	taatg	aaagtttaggaca
	gtccaatctgtgg	ggtacttcgtaat	atccgtgcatcgc	ggccctggtagtc	tcggc	.....	gctca	cgtgcccggagaa
					atact	.....	ccgta	
					ctccc	.....	taatc	
					tgtcc	.....	ggaca	
					gattc	.....	ttcat	
					attcg	.....	aaaca	
					ctgtt	.....	aatta	

Figure 3. Pub_Bob and Sec_Bob using the DNA segments in GenBank

Figure 4 displays the Hiding and Extraction operations as an algorithm for both Alice and Bob.

Technium Vol. 5, pp.21-28 (2023)
ISSN: 2668-778X
www.techniumscience.com

Alice	*Bob*
1. get the message to be hidden → get(m)	*1. get the ciphertext (St)*
2. get the cover used to hold m → get(c)	*2. get Pub_Bob and Sec_Bob*
3. embed m within c using Pub_Bob HMD(P_i) *St=Em(m, c,HMD(Pi))* *Where HMD → location(s) of the segments within GenBank*	*3. extract m from St using Pub_Bob and Sec_Bob* *a. Pub_Bob HMD(P_i),* *b. Sec_Bob (P_iL)* *P_{iL} = The value of the DNA segment and its length.*
4. Send St	*4. Get m*

Figure 4. The steps of Hiding and extraction methods on both sides (Alice and Bob)

The suggested Hiding and Extraction techniques leverage public keys without using any mathematical calculations (modules or elliptic curves). GenBank's vast DNA data can be used. Alice knows the segment's position as a key; while Bob knows its value and length. Using DNA as a private key also involves Bob alone.

3. Discussion

By eliminating the requirement to send the secret key via a public channel, the suggested technique makes data embedding and extracting more secure. Depending on the terms of the agreement, the public key may be shared with a wide variety of recipients. This key will be used to embed the message m by the sender. Any meaningless code of digits or letters can serve as the key to a GenBank site, and each of those locations can store a billion different pieces of DNA. GenBank is useful for our suggested embedding approach since, even if the adversary has the key, he cannot examine it. When considering the attacker's computational resources and time, it's also difficult to examine all sites in GenBank. The secret or private key is known only to the recipient, with the sender having no access to it under any circumstances.

The recipient extracts the stego-file using his public and private keys. He uses the public key to find the segment in GenBank and the segment's value to get his private key for extraction. This private key can have any form, depending on how it's obtained. Regardless of the strategy, it will provide robust security as only the receiver knows the DNA key.

GenBank is a public resource of 15.3 trillion base pairs from 2.5 billion nucleotide bases which be used as a private key and are known only by the receiver himself sequences for 504 000 species, according to [4]. So, calculating the likelihood of finding the location of one DNA segment used as a public key from these segments is tough. The search must attempt all feasible sites, which is time-consuming. Even if the attacker finds the exact position, i.e., the public key, he will obtain a fuzzy meaningless number. If we know that the amount of bases that are stored in GenBank has typically doubled every 18 months according to [5]. Also, the secret key, which may be any collection of DNA bases, is challenging or impossible if each base is two bits and the four bases compose one byte. As 00-A, 01-C, 10-G, and 11-T. So, the private key will be a group of DNA bases and a chosen technique done on them by the receiver. The number and the locations of DNA.

15.3 trillion bases, each of which can be C, A, T, or G. So, the probability of getting the value of the DNA chosen DNA segment will be 4^(17.3 trillion) if the segment length consists only of 4 DNA bases. This probability will increase if we use the binary coding for the DNA bases as every base can be represented by two bits with different coding as in Table 1
Table 1 lists the possible DNA base coding

Table 1. DNA Base Coding.

Coding	Bases			
	A	C	G	T
Code1	00	01	10	11
Code 2	01	00	11	10
Code 3	10	11	00	01
Code 4	11	10	01	00

The NIH genetic sequence database is freely available online, allowing access anytime, anywhere.

The attacker must check every GenBank to estimate public and private keys. He must also divide the DNA segment's worth by billions, an impossible task. The recipient alone knows the attacker's approach. If the secret key has 4 DNA bases, the attacker must try 4^(〖17.3 trillion〗^16!)In this scenario, the DNA bases were handled as one unit consisting of 4 bases. However, the potential increases considerably if these 4 bases were gathered based on a particular sequence as a key. The Hiding approach requires no arithmetic or computations. It might give robust security utilizing biological concerns like GenBank DNA data and DNA sequence features.

3.1. Steganalysis

Information connections may be attacked in numerous ways, the man in the middle attack is one of the most prominent. This attack is described as the person in the middle breaks off the data handed on by the sender and sends it.

If an MITM or black hole attacker tries to access data, it must be extracted. The attacker can't see the sender's data in transit. Increased DNA data volume leads to data storage and privacy difficulties, yet the attacker has no knowledge of the receiver's private key or sent data. The attacker also obtains ciphertext. DNA coding is changed. The attacker can't access the hidden plaintext and ciphertext information.

The proposed method will be very effective when the message to be hidden is converted, as well as the carrier file is converted to the same encoding as the DNA format. Here, the attacker's task will become very difficult, if not completely impossible, as the public key will be specific sites for the DNA Segments, while the secret key will be either the number of bases within these sites or it will be calculated in a chaotic manner that also depends on these bases.

4. CONCLUSIONS

This work uses public and private keys for hiding and extraction. The sender encrypts the public critical public key to embed, while the receiver uses public and private keys to extract. The sender, receiver, and private key are unknown. DNA Banks and segments are used. This approach provides good security with fewer arithmetic operations. The simple method is presented. Due to limited power and storage, the recommended solution can be employed in IoT security.

References

[1] M. Bishop, "Introduction to computer security," 2005.
[2] B. A. Forouzan and D. Mukhopadhyay, *Cryptography and network security* vol. 12: Mc Graw Hill Education (India) Private Limited New York, NY, USA:, 2015.
[3] N. Hamid, A. Yahya, R. B. Ahmad, and O. M. Al-Qershi, "Image steganography techniques: an overview," *International Journal of Computer Science and Security (IJCSS),* vol. 6, pp. 168-187, 2012.
[4] D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell, K. D. Pruitt*, et al.*, "GenBank," *Nucleic acids research,* vol. 46, pp. D41-D47, 2018.
[5] E. W. Sayers, M. Cavanaugh, K. Clark, K. D. Pruitt, C. L. Schoch, S. T. Sherry*, et al.*, "GenBank," *Nucleic acids research,* vol. 49, pp. D92-D96, 2021.
[6] M. Y. Galperin and X. M. Fernández-Suarez, "The 2012 nucleic acids research database issue and the online molecular biology database collection," *Nucleic acids research,* vol. 40, pp. D1-D8, 2012.
[7] E. W. Sayers, J. Beck, E. E. Bolton, D. Bourexis, J. R. Brister, K. Canese*, et al.*, "Database resources of the national center for biotechnology information," *Nucleic acids research,* vol. 49, p. D10, 2021.
[8] N. R. Coordinators, "Database resources of the national center for biotechnology information," *Nucleic acids research,* vol. 46, p. D8, 2018.
[9] I. Hussain and N. Pandey, "Carrier data security using public key steganography in ZigBee," in *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, 2016, pp. 213-216.
[10] Z. K. Al-Ani, A. Zaidan, B. Zaidan, and H. Alanazi, "Overview: Main fundamentals for steganography," *arXiv preprint arXiv:1003.4086,* 2010.
[11] M. Conti, N. Dragoni, and V. Lesyk, "A survey of man in the middle attacks," *IEEE communications surveys & tutorials,* vol. 18, pp. 2027-2051, 2016.
[12] V. Annapurna, S. N. Rao, and M. Giriprasad, "A Survey of different video steganography approaches against man-in-the middle attacks," in *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 2021, pp. 1601-1607.
[13] K. J. Sarma, R. Sharma, and R. Das, "A survey of black hole attack detection in manet," in *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 2014, pp. 202-205.
[14] G. M. Keerthi, M. Lalli, and V. Palanisamy, "Secured Solution and Detection against Black Hole Attack in MANET by finding the Optimum Path in AODV protocol and high secured data transmission using Steganography."