A Simple Message-Encryption Scheme based on Amino-acid Protein Sequence

Nirmala.V[1]
Student of Master of Technology,
Computer Science & Engineering Department,
Avanthi's St. Theressa Institute of Engineering & Technology,
Vizianagaram, Andhra Pradesh, INDIA
nirmala.vellaiswamy@gmail.com

Uppe.Nanaji[2]
Associate Professor,
Computer Science & Engineering Department,
Avanthi's St. Theressa Institute of Engineering & Technology,
Vizianagaram, Andhra Pradesh, INDIA
nanajistiet@gmail.com

***Abstract* -** Recently, biological techniques become more and more popular, as they are applied to many kinds of applications, authentication protocols, biochemistry, and cryptography. . Bioinformatics [2] plays a very important role on molecular datasets. Encrypting secret data in peptide sequence or amino-acid sequence becomes an important and interesting research topic. This paper presents a simple, secure and reversible encryption scheme that converts the message into an amino-acid protein sequence to provide security.

***Keywords:* Amino-acids, Cryptography, Encryption, Bioinformatics.**

I. Introduction

Today, network technologies have improved a lot so that more and more people access the remote facilities and send or receive various kinds of digital data over the Internet. However, the Internet is public but insecure channel to transmit data. Thus, important information must be converted into a non-readable form while delivered via the Internet such that only the authorized receiver can read it. Different methods of encryption and decryption techniques were used from ancient times. Nowadays biology techniques are proposed for encryption and decryption. As known Amino**-acids** are molecules containing an amine group, a carboxylic acid group and a side-chain that varies between different amino acids.

The key elements of an amino acid are carbon, hydrogen, oxygen, and nitrogen[5]. Amino acids are critical to life, and have many functions in metabolism. One particularly important function is to serve as the building blocks of proteins, which are linear chains of amino acids. Amino acids can be linked together in varying sequences to form a vast variety of proteins. Twenty-two amino acids are naturally incorporated into polypeptides[6] and are called proteinogenic or standard amino acids, those are shown in the table given below.

A = alanine	M = methionine
C = cysteine	N = asparagine
D = aspartic acid	P = proline
E = glutamic acid	Q = glutamine
F =phenylalanine	R = arginine
G = glycine	S = serine
H = histidine	T = threonine
I = soleucine	V = valine
K = lysine	W =tryptophan
L = leucine	Y = tyrosine

(In addition, there are two additional amino acids that are incorporated by overriding stop codons)

U = Selenocysteine
O = Pyrrolysine

From the above 22-aminoacid sequence a user can always randomly select one permutation among 22! Permutations .The present work discusses a symmetric key encryption[1] technology using the above amino-acid sequence. The sender selects a random permutation of the above sequence as a key and generates a dynamic look-up table for encryption and decryption. As this is a symmetric key algorithm, the sender transmits the key to the receiver as this secret key[3] that should be used for both encryption and decryption.

II. Key Generation

1) Alice divide the above sequence into two subsets randomly.

i){A,C,D,E,F,G,H,I,K,L,M,N,O,P,Q,R,S,T,W,Y,U } and ii) { V }.

2) As Alice wants to transmit secure information to BOB, he selects a random permutation of the first subset , say selecting a random number in the range { 1,21!} as secret key. Say he selects the permutation “HGFCADILNMYTSWFERPOQU”.

3) Alice also selects a random character as a second subset.

4) Alice sends the key { “HGFCADILNMYTSWFERPOQU”, “V” } to Bob.

III. Encryption:

Alice generates a look-up table that represents amino acid sequence[4] for standard ASCII character-set that range from 32 to 126.

ASCII CODE	Amino-acid character rep.	ASCII CODE	Amino-acid character rep.
0	H	18	O
1	G	19	Q
2	F	20	U
3	C	21	FG
4	A	22	FF
5	D	23	FC
6	I	24	FA
7	L	25	FD
8	N	26	FI
9	M	27	FL
10	Y	28	FN
11	T	29	FM
12	S	30	CH
13	W	31	CG.
14	F	32.	CF
15	E	.	.
16	R	.	.
17	P	94	MA

Alice now represents each character in the plaintext with not more than two characters to generate the ciphertext. Since he chooses {V} as the second subset, V represents a single character representation in the ciphertext. Alice generates the ciphertext of each character by subtracting 32 from the ASCII value of the character and gets the amino-acid sequence for the character from the above table.

Example:

Let us suppose Alice has to send the plaintext " This is Amino-acid Sequence. "

The way Alice generates the cipher text for the given plaintext is shown in the table below.

Character in the Plaintext	**ASCII code**	**ASCII code - 32**	**Amino-acid** **Representation for the value (ASCII-32)**
T	84	52	DF
H	104	72	LF
I	105	73	LC
S	115	83	NC
Space	32	0	H
I	105	73	LC
S	115	83	NC
Space	32	0	H
A	65	33	CC
M	109	77	LL
I	105	73	LC
N	110	78	LN
O	111	79	LM
-	45	13	W
A	97	65	ID
C	99	67	IL
I	105	73	LC
D	100	68	IN
Space	32	0	H
S	83	51	DG
E	101	69	IM
Q	113	81	NG
U	117	85	ND
E	101	69	IM
N	110	78	LN
C	99	67	IL
E	101	69	IM
.	46	14	F

Alice precedes every 2-character sequence with ‘U’ and every 1-character sequence with ‘V’. Now Alice sends the cipher text

“DFLFLCNCVHLCNCVHCCLLLCLNLM VWIDILLCINVHDGIMNGNDIMLNILIMVF” for the plaintext “This is Amino-acid sequence”.

IV. Decryption

The decryption process is just the reverse of the encryption process with the same key. After receiving the cipher text, Bob generates the same look-up table to decrypt the cipher. He decrypts the cipher by taking 2 characters at a time to get a plaintext character and whenever he encounters the character “V” he takes single character to get a plaintext character.

V. Security Analysis:

This Encryption Scheme is secured as along as the algorithm is kept secret or else this scheme is subjected to known cipher text and Brute-force attack as there 21! Permutations are to be experimented by the cryptanalyst to get the secret key.

VI. Performance:

As the proposed algorithm is simple and based on ASCII decimal codes of the characters, the encryption and decryption times are very less.

VII. Future Work:

- The algorithm can be also used to hide or embed a text in an amino-acid sequence.
- This Scheme can be made more complex to provide more security if we permit more permutations in the key-generation and encryption algorithms.

ACKNOWLEDGMENT

We are grateful to our computer science & engineering department for supporting us. This paper was written as a part of thesis work in master of technology, supervised by Uppe.Nanji.

REFERENCES:

[1] W. Stalling (2003), Cryptography and Network Security, Prentice Hall, New Jersey, USA, Third Edition, Chapter 10.
[2] T.K.Attwood et.al “Introduction to bioinformatics”, Pearson Education.
[3] “An overview of cryptography” by Gary C.Kessler ,21st March , 2011, http://www.garykessler.net/library/crypto.html.
[4] Introduction to Computational Molecular Biology by Carlos Setubal, Joao Meidanis
[5] Bioinformatics for Dummies by Jean-Michel Claverie, Cedric Notredame.
[6] Developing Bioinformatics Computer Skills by Cynthia Gibas, Per Jambeck