Mathematical Model

on

“Secure Data Storage On Multi-Cloud Using DNA Based Cryptography”

Submitted by :

Sagar Vithal Dhuri

Prakash Kondibhau Naikade

Nilesh Balakrishna Gade

Abhijeet Bhagwan Teke

Under the Guidance of:

Mrs. D. S. Zingade

Bachelor of Computer Engineering At

सत्याला मरण नाही

All India Shri Shivaji Memorial Society’s

Institute Of Information Technology

Shivaji Nagar, Pune - 01

Academic Year 2014 - 2015

Affiliated to

Savitribai Phule Pune University

Chapter 1

Mathematical Model

Laboratory Assignments on Project Analysis of Algorithmic Design

1.1 Embedding Data

In order to explain embedding phase, separating the phases into some successive and vivid sub-phases, is the best way of proposing current method. In below, sub-phases have been shown, respectively.

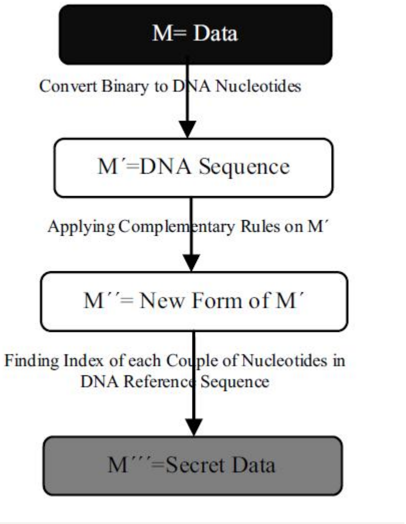
M= Data
Convert Binary to DNA Nucleotides
M´=DNA Sequence
Applying Complementary Rules on M´
M´´= New Form of M´
Finding Index of each Couple of Nucleotides in DNA Reference Sequence
M´´´=Secret Data

Figure 1.1: Embedding secret data

The first sub-phase is, converting by DNA base pairing rules. The product is M. M contains nucleotides sequences. **The next (second) sub-phase is**, applying complementary rules. Increasing the complexity is the real and exact purpose of this step. By applying the complementary rules, the new form of the M which is M emerges.

When all the indexes have been extracted, M has been made, properly. M is precisely the secret data with some changes through the embedding phase.

- DNA Reference Sequence:
  AT$_1CG_2AA_3TT_4CG_5CG_6CT_7GA_8GT_9CA_{10}CA_{11}AT_{12}TC_{13}GC_{14}GC_{15}TG_{16}AG_{17}TG_{18}AA_{19}CC_{20}$

- Let The Message Be M.
  M = 100111000011

- Sub-phase1$_{(A=00,T=01,C=10,G=11)}$:
  $M' = CTGAAG$
  $M' = \Sigma_{start}^{eof}(encode1(d));$
  where eof → $End\ Of\ File$
  $encode1 \rightarrow function\ returning\ DNA\ nucleotide\ for\ set\ d$
  $d = \ (d1, d2) \mid d1\ and\ d2\ are\ two\ consecutive\ digits\ in\ the\ file$
  $M'\ is\ phase\ 1encryption\ message.$

- Sub-phashe2$_{((AC)(CG)(GT)(TA))}$:
  $M''\ =\ GATCCT$
  $M''\ =\ \Sigma_{start}^{eof}(encode2(x));$
  $encode2\ \rightarrow\ function\ returning\ DNA\ nucleotide\ compliment\ for\ set\ 'x'$
  $x\ =\ \ \{A, T, C, G\}$
  $M',\ is\ phase\ 2\ encryption\ message.$

- Sub-phase3 (Indexes):
  M”’ = 8137
  M”’ = $\Sigma_{start}^{eof}(encode3(g));$
  $encode3\ \rightarrow\ function\ returning\ indexes\ for\ DNA\ pair$
  $g\ =\ (g1, g2) \mid DNA\ Pair\ from\ M''$
  $M'''\ is\ phase\ 3\ encryption\ message.$

Now, embedding phase is finally completed. Then, sender sends 8,13,7 to the cloud. In the next section, the client 2 will apply the extracting phase for extracting the original data by using three consecutive phases.

1.2 Extracting Original data

the secret data in form of some numbers. For the purpose of extracting the original data from DNA reference sequence, phase two with its sub phases will extract the original data, correctly.

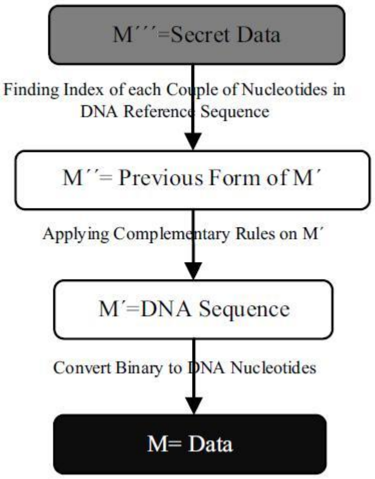
M´´´=Secret Data
Finding Index of each Couple of Nucleotides in DNA Reference Sequence
M´´= Previous Form of M´
Applying Complementary Rules on M´
M´=DNA Sequence
Convert Binary to DNA Nucleotides
M= Data

Figure 1.2: Extracting original data

  - DNA Reference Sequence:
    AT$_1 CG_2AA_3TT_4CG_5CG_6CT_7GA_8GT_9CA_{10}CA_{11}AT_{12}TC_{13}GC_{14}GC_{15}TG_{16}AG_{17}TG_{18}AA_{19}CC_{20}$

- Sub-phase 1 (Indexes):
  M = 8137
  M” $= \Sigma_{start}^{eof}(decode3(g));$
  $decode3 \ \rightarrow \ function\ returning\ DNA\ pair\ for\ indexes$
  $g \ = \ index\ from\ M'''$
  $M''\ is\ phase\ 31\ extracted\ message.$

- Sub-phashe $2_{((AC)(CG)(GT)(TA))}$:
  $M = GATCCT$
  $M'' \ = \ \Sigma_{start}^{eof}(decode2(x));$
  $decode2 \ \rightarrow \ function\ returning\ DNA\ nucleotide\ compliment\ for\ set\ 'x'$
  $x \ = \ \{A, T, C, G\}$
  $M',\ is\ phase\ 2\ extracted\ message.Sub-phase3_{(A=00,T=01,C=10,G=11)}$:
  $M = CTGAAG$

$M' = \Sigma_{start}^{eof}(decode1(d));$

$decode1 \rightarrow function\ returning\ binar\ digits\ for\ DNA\ nucleotided =\ DNA\ Nucleotide\ From\ M''M'\ is\ original\ message$

- M = 100111000011

1.3 Throughput

$\mu = (\Sigma_{i=0}^{n} D_n)/T$

where ,

$\mu\ \rightarrow\ Throughput.$

$D_n\ is\ data\ upto\ n\ bits.$

$T\ is\ Average\ time\ consumed.$

1.4 Time Complexity Conclusion (NP Hard/Complete, P)

The Mathematical model obtained gives the solvable output for proposed system. But the time depends on the positions of the base pairs in the reference string. So we conclude that our system lies in NP-complete.