An innovative application of virus taxonomy and nomenclatures in the Information Technology

Prof. Dr. Samreen Amir
Chairperson
Department of Electronic Engineering

Outline

- Abstract
- Motivation
- Introduction
- Proposed Methodologies
- Implementation of Method – I
- Implementation of Method – II
- Conclusion

Abstract

- The basic objective is to propose the design to develop a Virus- the organism based communication system for intelligence information transfer; A secure data communication through V-technology. The taxonomy tree and the anatomical structure of virus is used as an encrypted data packet.
- The basic simulations are performed on Proteus platform. This idea shall extend the application of bio technology and biomedical engineering in communication systems.

Motivation

- In the communication science and technology the term biotechnology is found to be very rare or may be not at all.
- Most of the things that we use as a result of brilliant engineering design and implementation are analogues to some natural object or phenomenon.
- Similarly when the anatomical structure of virus was observed it was very much resembling to the data packet we transmit or receive over a network.
- A data packet encapsulate the actual data bits in so many layers. The open system interface (OSI) model gives seven layers scheme.
- On the other hand the virus taxonomy schemes define every virus as hierarchical tree of several layers commonly 7-8 in number [2].
- The network security is always been under attacks and threats these days. Every communication technology has risk of hacking or unauthorized taping.
- The evolution of CDMA technology got rapid acceptance due to its inherited property of encryption.
- Similarly in the proposed system virus would be used as an encrypted data packet.

Introduction

- Several taxonomy schemes for viruses have been presented from 1961to date by ICTV (International Committee on Taxonomy of Virus) [1].

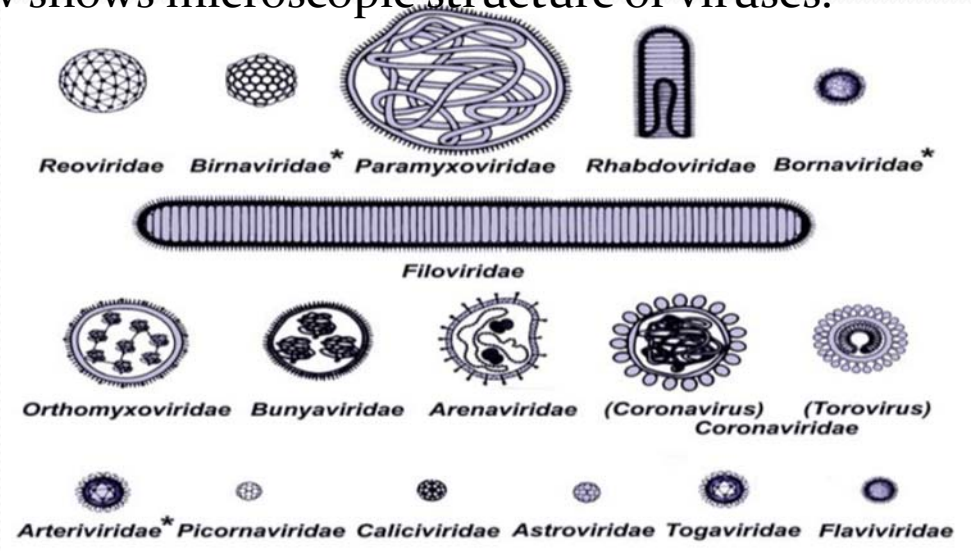
- The basic considerations in defining distinguished virus order, families and genera are genome, replication strategy and structure.

- Usually the advance biotechnology is been applied for health sciences and agricultural developments. Discovering new diseases and their causes & cures have been the main focus of this particular area of science.

Introduction – Virus Classification

- The classification of all viruses consist of phylum - class - order - family - subfamily - genus - species - strain/type

- The fig. below shows microscopic structure of viruses.



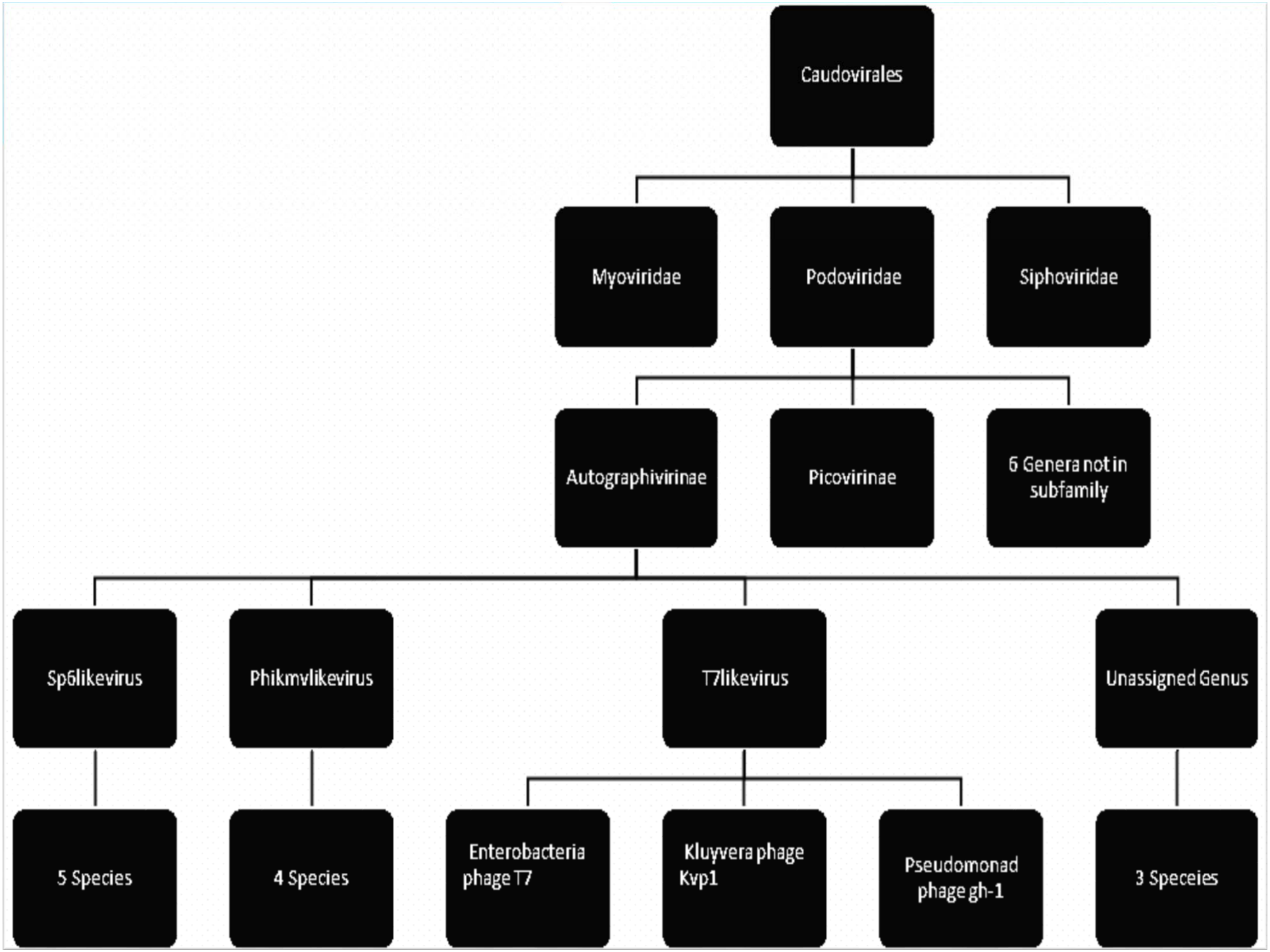
Proposed Methodologies

- The basic components of communication system are Transmitter- Medium- Receiver along with the data source and destination.
- In the light of this theory there can be two different ways of implementing the proposed method.
- The simplest of all is to use virus taxonomy as an encryption tool that is using the existing communication system with the proposed encryption technique.
- The mapping or encoding algorithm may be further divide in to different types depending upon the virus classification attribute adapted for encoding; for instance type1 may refer to Baltimore Classification and etc.
- The other method uses virus physically as a data carrier sent over suitable medium. In this case the receiving end should have all necessary equipment to identify the sent virus and decode the data bits. That means a different communication setup altogether.

Implementation of Method - I

- In this suggested method the existing communication system may be used.
- The mode of communication may be wired or wireless.
- The basic form of data shall be text or voice.
- The receiver in this case must be synchronized with the encryption scheme running at transmitter end.
- The specific encryption software must be installed on the receiving device whether it's a cell phone, computer or the any other electronic device.

- This scheme is based upon ICTV standard of virus classification.
- The current ICTV release shows that there are 71 virus families which do not contain registered “Order”.
- While 25 are divided in to seven different order. It gives one of the data bit. With order logic “1” without order “0”. It would be easy to understand taking a specific virus as an example.
- The figure 2 shows the taxonomy tree for one of the species of order named “**Caudovirales**”.
- The available virus orders can have maximum 5 families as per latest release. So 3-data bits are good to decode them.
- Likewise every family has maximum 4 sub families and minimum zero that is no sub family. 02 data bits are enough for this entry.
- Maximum 17 genera are found under any sub family so 5- bits can be assigned to this field. Up till now 5000 or more species are known but maximum 37 species found under one genus. So 8 bit assignment shall easily serve the purpose.
- The total data bits are 22 bits, shown in Table1.


Caudovirales
Myoviridae
Podoviridae
Siphoviridae
Autographivirinae
Picovirinae
6 Genera not in subfamily
Sp6likevirus
Phikmvlikevirus
T7likevirus
Unassigned Genus
5 Species
4 Species
Enterobacteria phage T7
Kluyvera phage Kvp1
Pseudomonad phage gh-1
3 Speceies

- The available virus orders can have maximum 5 families as per latest release. So 3-data bits are good to decode them.
- Likewise every family has maximum 4 sub families and minimum zero that is no sub family. 02 data bits are enough for this entry.
- Maximum 17 genera are found under any sub family so 5- bits can be assigned to this field. Up till now 5000 or more species are known but maximum 37 species found under one genus. So 8 bit assignment shall easily serve the purpose.
- The total data bits are 22 bits, shown in Table1.

Table 1: Total data bits and their assignment

order	Listed Order name	family	Sub family	genera	species
1-bit	3-bits	3-bits	2-bits	5-bits	8-bits

The mechanism of number of bits selection for an entity is as equation 1.

$$N = 2^n \qquad \text{Equation.1}$$

Where **N** is the total number of available options for a given entity and **n** is the number of bits.

What these bits mean?

As the English language rules a sentence must have subject, predicate, action and auxiliary words. As a working example the table 2 provides the encoding map.

Table 2: Encoding map

Order+name= numbers	Family=pronouns	Sub-family=auxiliary	Genera=action word	Species = Nouns
1000=0	000=I	00=Am/is/are	00000=Go	00000000=Park
1001=1	001=We	01=was/were	00001=Come	00000001= Hotel
1010=2	010=You	10=will/shall	00010=Leave	00000010=School
1011=3	011=Me	11=has/have/had	00011=Take	00000011=Money
1100=4	100=They		00100=Drive	00000100=Wine
1101=5	101=Them		00101=Drink	00000101=Water
1110=6	110=Their		00110=Sleep	00000110=President
1111=7	111=Us		00111=write	00000111=Washington
0XXX=not registered order			01000=Invite (up to 11111)	00001000= Children (up to 11111111)

If the receiver receives a text or voice message as "*gh-1*" then it must be decoded as "10000010000011000000010".

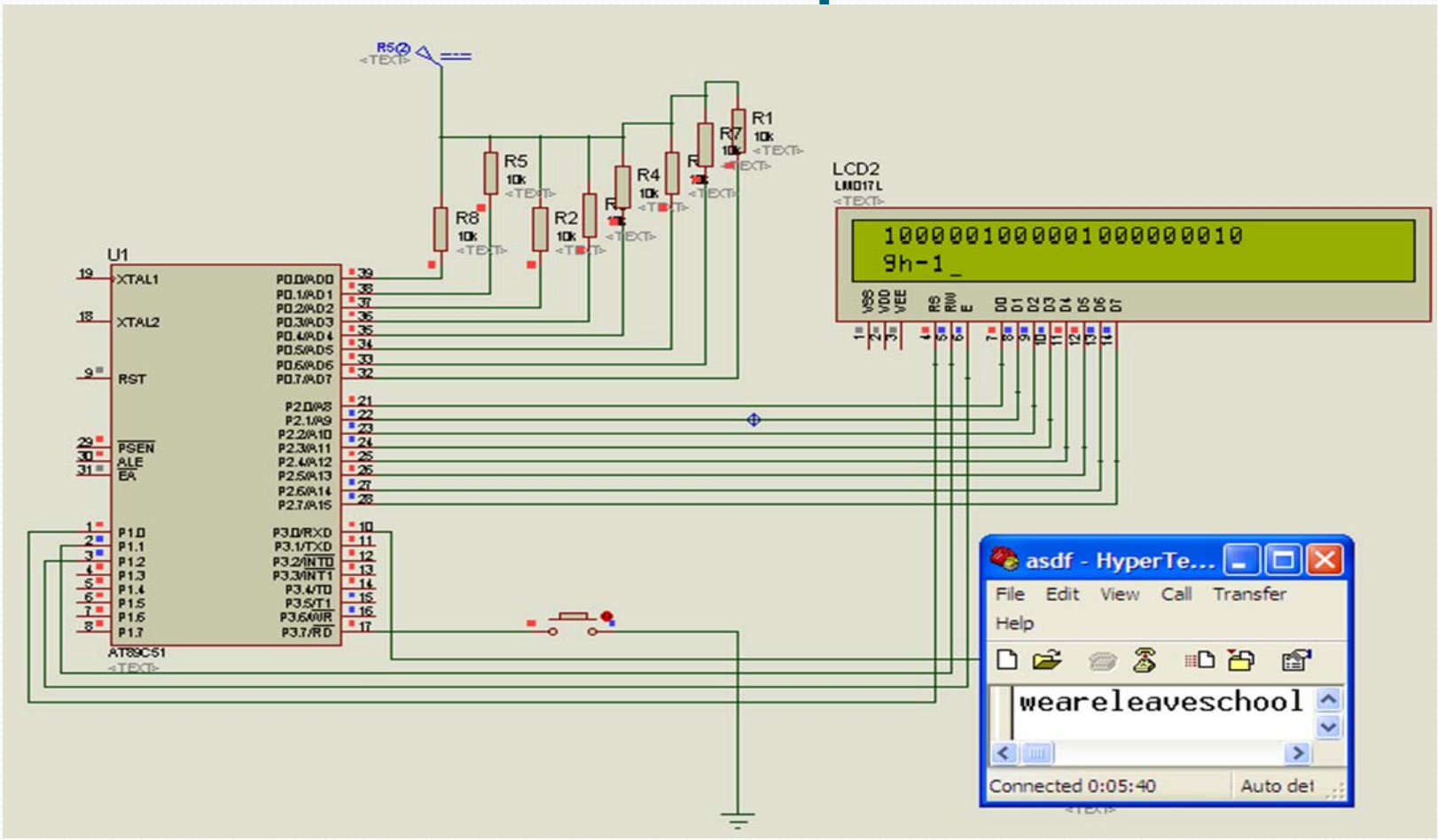
The said or sent virus name is listed as first order entry that is "Caudovirales" in ICTV latest release.

It's the third specie of third genera of first sub-family of second family of order Caudovirales.

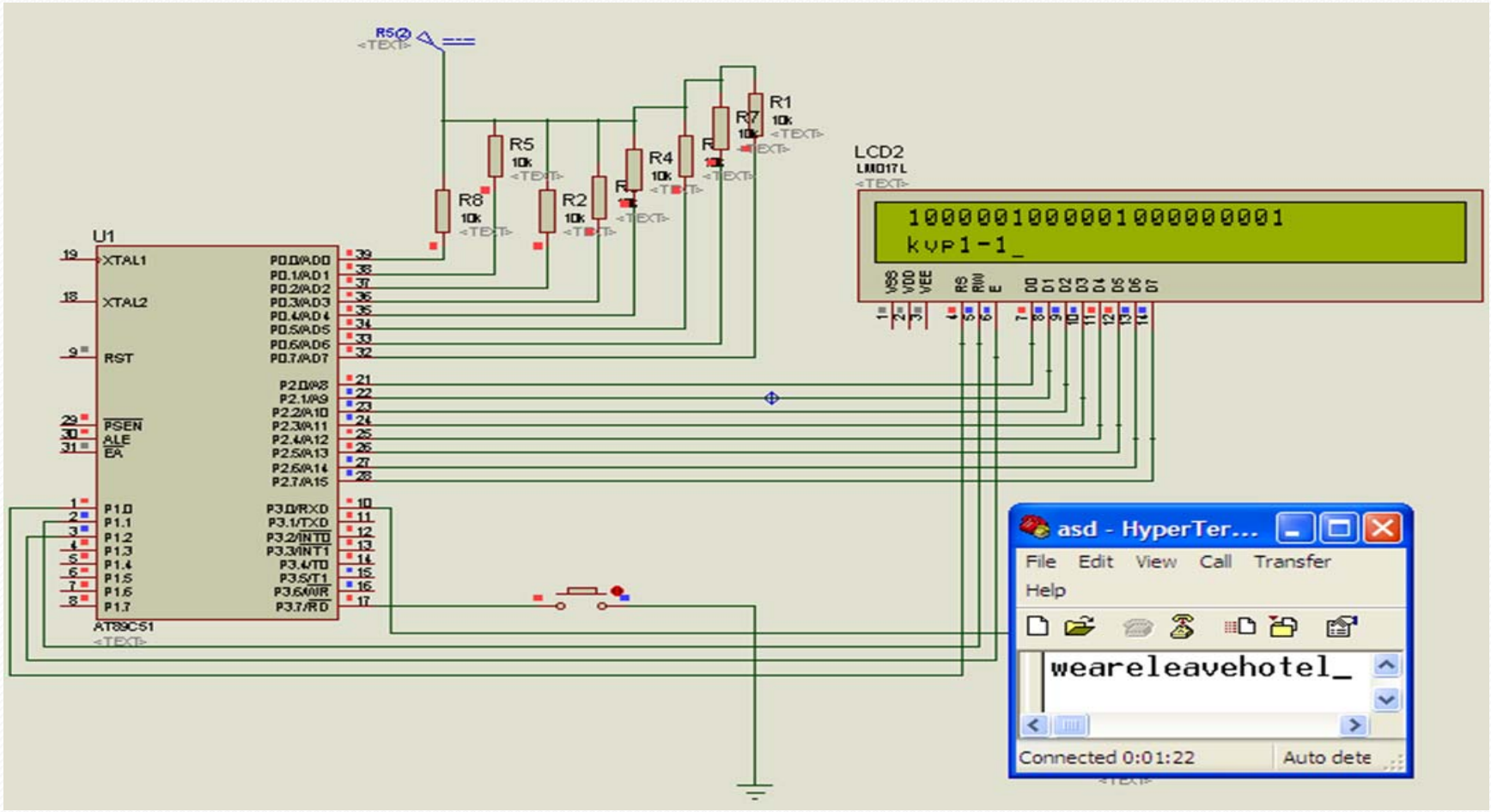
Order + name	family	Sub family	genera	species
1000	001	00	00011	00000010

The information extracted is "We are leaving School", shown in next figure.

Results of Technique -I



Implementation of strategy 1 for virus “gh-1”



Implementation of strategy 1 for virus “kvp1-1”

Platform

- The implementation is done on Proteus simulation platform.
- The AT89C51 microcontroller was used.
- The purpose is to illustrate the mechanism hence kept simple.
- The number assignment may be used for several purposes like sequences or the sender code word etc.
- If the message to be sent is “o I am Going Park” then all we have to write is *O18P*; the name of virus equivalent to “oooooooooooooooooooo”.
- Remember it’s not for normal data communication it’s for Military or defense related applications.
- The list of commonly used terminologies in such type of conversation can be easily encoded with in these limits.
- Although the communication medium is same as for general purpose communications but the encryption technique is entirely different as that of routine Digital Encryption Standards (DES) or Advanced Encryption Standards (AES).

Encryption Strategy -II

- In this type of encoding the criterion for the classification may be treated as a data bit.
- For instance the morphology of the virus is one of the classification tools and it has three attributes shape, size and ultra-structure.
- The two common shapes found in viruses are spherical and isometric. We may assign logic 1 for spherical and 0 for Isometric.
- Similarly for size the logic assignment may be as if diameter ranges between 20-160 nm its 1 and for 160 to 300nm its 0 since the size of viruses is found between 20-300nm.
- Likewise for ultra-structure bit; for enveloped assign bit "1" and un-enveloped/complex assign bit "0". It can give eight possible combinations (3-bits code). So if the morphology represents the information about countries in G8 then the name of country may be decoded using predefined schemes with these data bits.
- The table 3 provides the basic morphological information of two viruses. These two examples can be helpful in understanding the proposed mechanism.

TABLE 3: MORPHOLOGICAL INFORMATION OF TWO VIRUSES

GROUP	ENVELOPE	SHAPE	SIZE (NM)
BUNYAVIRIDAE	YES	SPHERICAL	80-120
CALICIVIRIDA	NO	ISOMETRIC	35-39

As per the aforementioned protocol the first entry shall be decoded as “111” and second entry will be “001

Some more virus details used in defining the class and phylum of organism are given in table 4.

TABLE 4: VIRUS CLASSIFICATION CRITERION

Property	Parameters
Nucleic acid	• DNA • RNA
Strand style	• Single-stranded • Double-stranded • Double-stranded with regions of single-strands
Sense	• Positive sense (+) • Negative sense (–) • Ambisense (+/–)

Each entry may be assigned the number of bits sufficient to provide number of unique possible combinations required to represent the available options as described in equation 1.

Another possibility to generate the encoding states are considering virus groups; for instance in Group A the further sub division is performed on the basis of head formation of the virus.

For A1 its small isometric, for A2 it's prolate and for A3 it's elongated.

Another attribute to help the cause is tail of the virus. If the virus is tailed then the structure of tail may have three possibilities namely short non contractile, long non contractile and long contractile.

- If we consider the same specie as that for strategy 1 then it would be easy to understand the difference between the two.
- *PhiO18P* is spherical, 50nm diameter, enveloped virus. Its double stranded DNA genome and gram negative.
- The head is icosahedral and elongated and tail is thick and short contractile.
- So the encoder's decision tree can use this information and assign the respective code.
- In the previous case the virus was treated on the basis of hierarchy while in this scheme the encoding criteria is the physical structure.

Method-II

- This method would use biotechnology labs as the front end transmitter and receiver.
- Data will be virus itself and medium could be any material like plastic, plant, water/ liquid, paper etc that can hold that virus.
- The same encryption algorithms as described previously in method1 as strategy 1 or 2 may be used to process the extracted information of virus.
- Among thousands of virus species it was found that there are not just disease viruses but some plant virus and human friendly viruses too.
- The adeno-associated virus type 2 (AAV2) is one of the non-disease viruses
- Viruses are present around our environment all of the time but reproduce only when host cell come along.
- Regardless of the type of host cell, usually viruses follow the lytic cycle; that involves search for host, absorption, entry, genome replication, maturation and release.

- This process may be utilized for data propagation after appropriate mapping.
- The proper host cell is recognized or felt by a type of protein on the outside coat or envelope.
- This property can help in the development of the point to point communication sessions.
- The new copies of genetic instructions are inside the new protein coats hence producing new viruses.
- A virus is merely a set of genetic instructions surrounded by a protein coat and it does not carry out any biochemical reactions without host cell.
- Viruses can live for years or longer outside a host cell. That means a data packet on a route with long “Time to live” (TTL) over a network.

Conclusion

- The initial outcome will find applications in military or defense institutions and intelligence agencies.
- The transportation media in Method2 are also very unusual which is very unlikely to be attacked for hacking or tapping purpose.
- The data transmitted through V-technology Method2 can only be extracted in a specified laboratory.
- It is also important to consider the human and environment safety while developing such system.
- The basic required physical resources are already there in terms of virology department labs and qualified personnel.
- The development of proposed mapping/ encoding algorithms using these existing facilities shall result in a highly encrypted data communication system.
- Since major facilities are already there so this proposed system doesn't demand high investment.


Questions