A SOLUTION TOWARDS OVERLAPPING CLUSTERS OF MALWARES USING MKM-PSO ALGORITHM
Ankur Singh Bist*
*CSE Department, KIET, Ghaziabad

ABSTRACT
Complexity of network applications has enhanced the problem of computer worms. At the same time the advancement in technology has given rise to evolution of portable mobile devices. It has become the danger point of day today activities like mails, mobile banking etc. Our purpose is to design an algorithm that can answer the problem of overlapping malware clusters.

KEYWORDS: Classifiers, Classification Methods, Computer Virus

I. INTRODUCTION
Internet has become target of malicious codes due to its increasing use. Malicious codes are executable code and have the capability to replicate. It makes their survival strong. Viruses design and evolution attached with the area of programming. Similar to other computer programs viruses carry functions that are intelligent for providing protection in such a manner that detection remains not easy for virus scanner [1].

Viruses have to take various procedures of intellect for continued existence. That is why they may have complex encrypting and decrypting engines. These are the most frequent methods used by computer viruses in current scenario. They make use of these techniques to mask the antivirus and to adopt the certain environment for their expansion [2].

Polymorphic viruses try to hide the decrypting module. More complex methods were developed enabling the virus designers to change the code of one virus file and make multiple morphed copies while maintaining its functionalities. These are the type of viruses which have the ability to mutate itself with the code changed but without changing its functionalities. Metamorphic virus can become a serious threat considering the fact that there can be thousands of variants of one virus file with their signature being totally different.

Metamorphic viruses transform its code in a specific manner very frequently and require to be prohibited. Their analysis will lead to evolve a framework where the overall process of detection will be bounded in specific outcomes of continuing evolving results. It is essential to make a distinction between replicating programs and its similar forms. Reproducing programs will not necessarily damage your system [3] [4] [5]. There is big fight between designers of virus and antivirus. The enhanced knowledge about the certain patterns, specifications can be designed. Various malicious codes can be evolved and incremented in well precise and efficient manner. For
perfect identification of a metamorphic virus, identification routines must be written that can generate the essential instruction set of the virus code from the actual occurrence of the infection [6-12].

Code obfuscation is one of the important properties adopted by metamorphic viruses. The mutating behavior of metamorphic viruses is due to code obfuscation techniques. There are various code obfuscation techniques.

a. Dead code insertion
b. Variable renaming
c. Break and join transformation
d. Expressing reshaping
e. Statement reordering

II. PROPOSED ALGORITHM (MODIFIED KMEANS-PSO) (MKM-PSO)
Malware samples are generated from NGVCK, MWOR kit and half of the samples are taken from different web links. Normal files are taken from windows. The score is generated with the help of pair-wise alignment algorithm.

III. K-MEANS ALGORITHM
Step 1: Input data set, clustering variable and maximum number of clusters
Step 2: Initialize cluster centroid
Step 3: Calculate Euclidean distance
Step 4: Move on to next observation and calculate Euclidean Distance
Step 5: Calculate Euclidean Distance for the next observation, assign next observation based on minimum euclidean distance and update the cluster centroids.

IV. PARTICLE SWARM OPTIMIZATION ALGORITHM
Generate random population of N solutions (particles);
For each individual i∈N: Calculate fitness(i);
Initialize the value of weight factor, w;
For each particle:

Set pBest as the best position of particle i;
If fitness(i) is better than pBest;
pBest(i)=fitness(i);
End;
Set gBest as the best fitness of all particles;
For each particle;
    Calculate particle velocity;
    Update particle position;
End;

Update the value of weight factor, w;
Check if termination = true;
End;

[In this scenario PSO is used to optimize cluster center]

**Code 1**

Step 1: $Z_i$=Objective data points, $C_i$, $C_j$ are two clusters.
    After clustering find($Z_i$)
    Finds whether $Z_i$ belongs to $C_i$ or $C_j$ for all $i$
    And calculate freq($Z_i$)
    Finds frequency of data points getting overlapped.
    Severe Function= freq($Z_i$)/((point($C_i$)+point($C_j$))

Step 2: Move cluster $C_i$ and $C_j$ in such a manner.
    Move center of $C_i$ to (-£) distance and relative move all data points strictly belongs to $C_i$ and
    $C_j$ to (£) until find($Z_i$)==Null

Step 3: while ($Z_i$==0)
    If(Hamm_distance($Z_i$)toCenter($C_i$)>Hamm_distance($Z_i$)toCenter($C_j$))
        $C_i$←$Z_i$
        $Z_i$ - -
    else
        $C_j$←$Z_i$
        $Z_i$ - -
    End if
End while

Following parameters are calculated:-

1. Accuracy=$\frac{TP+TN}{TP+TN+FP+FN}$
2. Precision=$\frac{TP}{TP+FP}$
3. Recall=$\frac{TP}{TP+FN}$
4. F-measure=$\frac{2 \cdot precision \cdot recall}{precision + recall}$
5. ROC- Receiver Operating Characteristic (ROC) curve defines how a identification rate change as the internal threshold varies to produce more or fewer false alarm.

**V. CONCLUSIONS**
Large number of research papers has been written in the field of overlapping clusters. Our purpose in this paper is to address the problem of overlapping malware clusters. After experimentation it is observed that this process improved the classification accuracy. In future we will test the same algorithm for different data sets.

**VI. ACKNOWLEDGEMENTS**
This work was supported by Dept. of Computer Science and Engineering, Krishna Institute of Engineering and Technology (K.I.E.T) Ghaziabad, India.
VII. REFERENCES

7. Das, Purushottam, and Ankur Singh Bist. "INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY FEATURE SELECTION."

CITE AN ARTICLE