IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Pattern Matching Algorithms For Computer Virus Detection

Ankur Singh Bist
Govind Ballabh Pant University Of Agriculture And Technology, Pant, India
ankur1990bist@gmail.com

Abstract

Computer viruses are serious threat for society . There are various procedures that are used for solving this problem . Our purpose here to make a collective look on various pattern matching approaches that are used in the domain of computer virus detection and mitigation .

Keywords: Pattern Matching , Algorithms .

Introduction

There are various applications that are handled now a day's using the approach of pattern matching this method can be used to solve various type of problems like the problem of classification and other . There are number of algorithms that are used to the task of pattern matching and are used in various other related activities .Computer viruses problem regarding there are various approaches , mostly used primary approach is signature scanning , it includes the process of pattern matching in which the antivirus maintain a database of signatures of most frequent known viruses and during the scanning process the signatures of files that passes through scanning , their signature get extracted and get matched with signature present in database thus detection is made .Daily viruses are increasing thus the database of antivirus is increasing day by day , it makes the scanning process long .Secondly the long process of scanning irritates the user so to overcome these drawbacks it is require that these pattern matching techniques when used in the domain of viruses then there should be some optimization process that can lead to some modified algorithms that has less time complexity so that finally scanning duration can be reduced.

PATTERN MATCHING ALGORITHMS

There are various pattern matching algorithms that are used in the domain of computer virus detection .These algorithms can be classified as[1]---------

Single Keyword pattern matching –

a) Brute force algorithm
b) Karp-Rabin algorithm
c) Boyer Moore Algorithm
d) Backward oracle matching algorithm
e) Knuth Morris Pratt algorithm

Multiple Keyword Pattern Matching –

a) Aho Corasick algorithm
b) Commentz walter algorithm
c) Wu-Manber Algorithm
d) Fan-Su algorithm

The Aho-Corasick algorithm needs a tri-like DFA and a failure function. The construction of these is described below.
Sandeep kumar in his paper A Generic Virus Scanner in C++ defines pattern matching algorithm in respect of virus scanner as developed in paper-------
This algorithm is similar to the Aho --Corasick algorithm but…..
it has been extended for wildcard characters[3].
? match any nibble in the input stream
%n skip 0-n nibbles in the input stream
*n skip exactly n nibbles
** skip an arbitrary number, including 0

```
for(each nibble in the input stream)
if(traverse(inputNibbleStream, TreeRoot))
{
/* virus detected */;
}
node *traverse(ifNibbleStream& i, node& n)
{
if(there are virus signatures associated with this node)
return &n;
ch = next digit from the input stream;
pos = current nibble position of i;
if(there is a link on ch from n)
if(ret = traverse(i, the node found by following the
link from n
on ch))
return ret;
for(all the %d, *d, ? & ** links of node n)
```

```
{
if(link is of type %d)
{
int k = value of d;
for(int l = 0; l <= k; l++)
{
restore file position to pos;
skip exactly l nibbles;
if(ret = traverse(i, the node obtained by following the
link from n on %d))
return ret;
}
}
else if(link is of type *d)
// all ? are converted to *1
{
restore file position to pos;
skip exactly d nibbles;
if(ret = traverse(i, the node obtained by following
the link from n on *d))return ret;
}
else if(link is of type **)
{
for (int l = 0; not end of file; l++)
{
restore file position to pos;
skip exactly l nibbles;
if(ret = traverse(i, the node obtained by following
the link from n on **))return ret;
}
restore file position to pos;
}
}
return 0;
}
```

Pattern matchimg algorithm used in generic scanner in c++[3]

This approach is taken by author to design their scanner having the thought process of certain modification in future for better detection. Another approach is taken by another algorithm that make use of Boyer Moore Horspool algorithm for pattern match .

Boyer Moore Horspool algorithm[2]--

```
while i < n do // a window is defined
j=m-1
k=i
while j >= 0 and T[k] = P[j] do
j - -
k - -
end while
if j < 0 then
report the occurrence of the pattern
end if
i = i + D[ti] //shift the window on the right
  End while
```

To determine the shift distance ***d***, the pattern is preprocessed: for each character ***a*** ***A***, and a distance ***ds*** is computed. Where ***A*** is the alphabet size. All the preprocessing steps can be written as[2]:

```
For a A do
D [a]=m
End for
For i=0 to m-2 do
D [Pi]=m- (i+1)
End for
```

Database Size (No. of Patterns)	Sequential Algorithm (Sec)	TBM (Sec)	BM (Sec)	BMH (Sec)
20	8.6	6.8	6.3	5.3
40	10.7	9.4	8.2	7.2
60	11.4	10.2	9.2	8.5
80	15.6	14.8	11.5	9.6
100	18.5	18.1	13.9	11.8

Performance according to size of database[2]

Finally a comparative analysis is done on the basis of database sizes and a reduction in the time is measured . There are various other pattern matching algorithm that we discussed that can be modified as shown by other authors . Other a new algorithm can be developed for fast searching but it require a deep analysis of existing algorithms in detail.

Conclusion

In this paper we make a review on various pattern matching algorithms that are used for computer virus detection and show the results obtained by authors to show the efficiency of their modified algorithms . In future there are lot of work that can be done in this field . We only discussed two pattern matching algorithms not all in the computer virus domain because our purpose here to make a review study that gives a initial idea that how pattern matching algorithm are modelled for solving virus domain problems.

References

[1] www.wikipedia.com.

[2] Sunita Kanaujiya, Dr. S.P.Tripathi , N.C.Sharma ," Improving Speed of the Signature Scanner using BMH Algorithm" *International Journal of Computer Applications.*

[3] Sandeep Kumar Eugene H. Spafford ," **A Generic Virus Scanner in C**++."Department of Computer Sciences Purdue University