Malicious Threats Detection of Executable File

Manoj D. Shelar, S. Srinivasa Rao

Abstract: *Malware is a general problems faced in the present day. Malware is a file that may be on the client machine. Malware can root an uncorrectable risk to the safety and protection of personal workstation clients as an expansion in the spiteful threats. In this paper explain a malware threats detection using data mining and machine learning. Malware detection algorithms with machine learning approach and data file. Also explained break executable files, create instruction set and take a look at different machine learning and data mining algorithm for feature extraction, reduction for detection of malware. In the system precisely distinguishes both new and known malware occurrences even though the double distinction among malware and real software is ordinarily little. There is a demand to present a skeleton which can come across latest, malicious executable files.*

Keywords: Machine Learning, Malware Detection, Opcode Sequence, Support Vector Machine, SVM

I. INTRODUCTION

Malware means malicious software. Malware is a term commonly used to denote all dissimilar types of unwanted software programs. Malware is any small piece of software that was calm to do damage to information, gadgets, or individuals. Malware is any product used to disturb computer tasks, assemble touchy data, or access private computer systems. Malicious software includes spyware, trojans , bots, rootkits , viruses, worms, and so on. Malware is any product used to disturb computer tasks, collect delicate facts, also referred to as a malicious software program, scripts, energetic material, etc. Malware seeks to infect, damage, or disable computers, pc systems, networks, tablets, and cellular devices, regularly via taking partial control over a device's operations

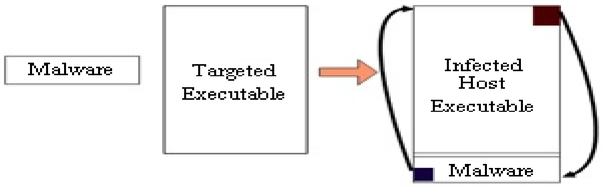


Fig. 1 Infected Executable File

Revised Manuscript Received on January 30, 2020.
***** Correspondence Author

Manoj D. Shelar*, Research Scholar Department of Computer Science and Engineering , Koneru Lakshmaiah Education Foundation Vaddeswaram, India Email: manojshelar58@gmail.com

Dr. S. Srinivasa Rao, Associate Professor, Department of Computer Science and Engineering , Koneru Lakshmaiah Education Foundation Vaddeswaram, India Email: srinumtech05@gmail.com

Fig. 1 shows malware is focused on the executable file, and an infected executable file is created.

Coming up next is a rundown of regular kinds of malware

A. Virus They can unfold wildly, harming a system's center usefulness and erasing or debasing files. They usually display up as an executable file.

B. Worms Worms contaminate whole networks of devices, either community or over the web, by making use of network interfaces. It utilizes every sequential tainted equipment to infect more.

C. Spyware Spyware is a malware planned to hold a watch on you. It conceals a way out and takes notes on what you do online, including your passwords, Visa numbers, surfing inclinations, and the sky is the restriction from there.

D. Trojans This kind of malicious application covers itself as certified software or is included in actual software that has been messed with. It will, in widespread act one by one, what is more; make indirect accesses in your protection to permit other dangerous programs.

E. Ransomware Ransomware likewise called as scareware, this kind of malware can secure your computers and threaten to delete everything besides if the charge is paid to its.

F. Adware It is a one type of malware. Some websites pop-up display is an adware, commercial implanted in program, get introduced through the permission of a client.

It once brought begins catching client's computer data, for example, individual data, firewall settings other browsing data.

The major tools identify malware by static or dynamic analysis such as

a. Signature-based: The specific features are obtain from files, which are used for the recognition of malware. It incorporates the vast majority of the antivirus tools that are utilized for recognition.

b. Heuristic approaches: The human specialists define guidelines for identify the patterns for malware recognition. These techniques have an inadequacy in distinguishing obscure or latest examples and can be avoided in different manners

II. MOTIVATION

There is much software in the market that detects viruses and worms. However, only a few software's for the detection of adware. The wide variety of unfortunate casualties is expanding who misplaced there cash also, perusing statistics in light of adware found in their device. Scarcely any enemy of adware programming is out there in a showcase; however, they're not effective as they make use of signature and heuristic methodologies for the popularity of malware.

These things influenced to construct malware recognition using a training series era with gadget gaining knowledge of and the mining of a record algorithm, which is most competent than existing anti-malware.

III. LITERATURE REVIEW

1) Alireza Khalilian [1] In this paper offer a model of G3MD is to use graph mining at the opcode graphs of a metamorphic family of malware to take out the common sub-graphs[1]. In light on those sub-graphs, a classifier is prepared to differentiate among a good file and malware file. Conducted experiments on four groups of malware not unusual in preceding studies, particularly Next Generation Virus Generation Kit (NGVGK), Second Generation Virus Generator (SGVG), and Mass-Produced Code Generation Kit viruses (MPCGK) and worms.

2) Shiva Darshan S.L. [2] In this paper clarify malware identity framework using highlight choice procedure. This paper offers the execution exam of 4 picked filters based FSTs and their touch with deference to the classifier conclusion [2]. FSTs, for example, Mutual Information (MI),Distinguishing Feature Selector (DFS), Categorical Proportional Difference (CPD), and Darmstadt Indexing Approach (DIA) were applied on this work, what's more, their productiveness has been assessed making use of specific datasets, different full length and classifiers, This paper clarifies the method utilizing 1) the Training stage and 2) the Prediction stage. The tutoring degree is applied to extend an instruction document, which is relied upon to set up the classifier. The expectation level estimates the detection capacity of the prepared classifier. The primary goal of the paper method is to expose the execution examination of the Filter, primarily depends on FSTs.

3)Zhihua Cui, Zhihua Cui, Yang Cao, Gai-ge Wang [3] In this paper used a technique that use deep getting to know to get better the recognition of malware. Research author used in-depth gaining knowledge of showed performance in image recognition correctly. In implantation, translate the malicious data into grayscale images. Then use a convolution neural network that would extract the functions of malware images. In the proposed approach creator worked on grayscale images. This technique does not provide paintings on coloration images.

4) Prapulla S B, Sharad J Bhat [4In this paper used approach for malware detecting using feature extraction from opcodes in the executable files. Paper authors used devices gaining knowledge of algorithms on the extracted feature to test the report is malware or benign [4]. This paper method no longer possible to hit upon new instances of malware. In this paper used the IDA pro disassembly tool, Exe information PE Win32 home identifier. In this paper, the concept now not paintings on a real set of malwares [4].

5) Rushabh Vyas [5] In this paper, explain malicious executable files detected on the community by the usage of system gaining knowledge of algorithms. This paper describes 28 features extracted from DLLs and features of 4 distinct forms of PE Files for malware detection. Also, the authors labored on static functions based on malware detection by the use of the extraordinary supervised set of rules and attention on PE Files.[5] Malware detection classified documents: malware or benign. Hence, in this paper, the writer deployed four supervised studying techniques (class models) for the undertaking of malware detection. Specifically, they have been k-Nearest Neighbours' (kNN), choice tree, help vector machines, and random forest.[5]

6) Farnoush Manavi, Ali Hamzeh In this paper, proposed an approach of malware detection of sensitive documents using a photograph processing technique. The projected technique indicates capable results. In this method producing a graph of opcodes from executables[6].SVM system studying a set of rules used in the paper.

7) H. Johnson R. K. Shahzad, [7] In this approach, use executable file and check file is malicious or benign. The projected method used to identify known and novel adware correctly. In upcoming work is supplanting the GIST with an suitable deep learning to extract features from images and finding a superior demonstration of opcodes to create the images [7].

8) M. G. Schultz [8] It use an information mining structure that identify latest and unnoticed malicious executables correctly. The author uses updated MacAfee's virus scanner and labeled our programs, either malicious or benign executables. This paper proposed technique is carried out as a network mail filter and paper to get pernicious executables before clients get them thru their mail [8]. In this paper provided feature scope has applied the device on a community of computers to evaluate overall performance in real-world environments.

9) Abubakr Sirageldin, Malware recognition using the call graph became explained. This paper describes a malware recognition technique dependent on the investigation of graphs introduced from instructions of the executable items also explains the assessment among MCS and SVM in the paper. The accuracy of type and clustering technique is challenged within the proposed method [9].

10) Boris Rozenberg, Ehud Gudes, Yuval Elovici, [10] In this paper present technique for detecting malicious executables, Following Steps i) Offiine training phase, finding a fixed of system call sequences which are traits of executable archive when malicious files are achieved and storing in a database. ii) Real-time recognition phase, for every going for walks executable, continually observing its gave system calls and evaluating with the stored sequences of system calls in the database. The objective of this approach is to give a method that could locate new malicious executables, whose signatures are unknown yet.

11)TE-EN WEI [11] In this paper, projected technique base on CSS to reconstruct the dumped file, The aggregate of CSS and the rebuild technique is known as RePEc File, which can be used to routinely reverse the packed PE report immaterial of walking on Windows or Linux platform. It offers a system to opposite a PE file via packing, unpacking, and rebuilding [11]. RePEc not only automatically reverses the packed PE file but also can support dissimilar platforms like Windows and Linux. Feature scope of this work to extend studies in this vicinity to address this encryption shell problem.

12) K. Ramamoorthy [12] It gives a fixed recognition approach to the usage of breaking of malware and create signature set. The diagnosed malware may be scrutinized to extort the mark. The method uses disassembled code. It offers an exam on how the strategy can be stretched out to distinguish spyware is too introduced.

13) S. Gordon [13] In this paper given the genuine significance of spyware region. An NCSA Study, as of late discovered that 80 percent of checked PCs had some sort of adware or spyware present.

14) R. K. Shahzad, [14], DM based malignant code indicators that are acknowledged to feature admirably for figuring out infections and comparative programming, this sort of identifier has now not been explored as some distance as to how well it may identify spyware. There is extraction of parallel highlights, referred to as n-grams ,from both spyware and right programming and follow five numerous managed to study calculations to put together classifiers that can group obscure pairs by breaking down separated n-grams.

15) Martin Boldt, Andreas Jacobsson, Niklas Lavesson, Paul Davidsson [15] This paper researches the principle that it is doable to identify from the End Client License Agreement (EULA) irrespective of whether its related programming has adware or not. There is an age of an informational index utilizing accumulating 100 packages with EULAs and ordering each EULA as either high-quality or negative.

16) Hashem Hashemi, Zahra Bazrafshan [16] There are three techniques used for malware recognition: Signature based and Heuristic ones. The slicing area heuristic malware recognition techniques and quick review unique highlights utilized in these strategies, for example, API Calls, Opcodes, N-Grams, and so on. Furthermore, examine their favorable circumstances and hindrance.

17) A. Sulaiman, K. Ramamoorthy, Member, TEEE, and A. H. Sung In this paper, [17] it presents a powerful low-level computing construct signature primarily based malware recognition procedure. There is an accentuation on recognizing polymorphic malware and changed (or transformative) malware.

18) R. K. Shahzad [18] Here introduces a scareware detection technique that relies upon the use of AI calculations to study designs in extricated variable period opcode preparations were given from guidance preparations of double documents [18]. The examples are then used to arrange to software as genuine or scareware yet they may likewise discover interpretable behavior this is considered one of a type to either type of programming.

19) Dai Wei, Ding Yuxin, [19] In this paper, show that the opcode practices extricated through the method can entirely speak to the conduct attributes of an executable. With the recognition approach based at the opcode appropriations, the proposed technique has better generally precision and a lower false excellent rate[19]

Main Contribution	Drawbacks
It gives an information of heuristic malware detection techniques[16]	A high False positive ratio is the maximum disadvantage of heuristic Malware Detection.
This paper affords a static detection technique for the use of the disassembly of a malware.[12]	This paper algorithm does not work on Binary codes. Tools used for specific machine code of particular Operating System
a rebuild strategy depends on CSS to remake the dumped file. (RePEF) File , it is used to automatically turn around the packed PE file.[11]	Not work on encryption shell problem
This system for real-time malware detection based on the projected technique.[10]	Not work on polymorphic malware.
Here Use malware recognition algorithm using call graphs. Malware recognition method based on the analysis of graphs presented from instructions of the executable file.[9]	Not Good accuracy of classification & clustering technique
It presents a skeleton that detects unnoticed malicious executable files.[8]	Algorithms not more good in time and space
This paper indicates an Adware discovery method based on the consumption of facts mining on dismantled code.[7]	This paper techniques not performed on the Larger set of Adware and Benign Files
Detect unfamiliar malware based on their opcode sequence.[6]	Not use deep learning network to extract features from images, and finding a superior portrayal of opcodes to build the images.
Examine Lat how vindictive flexible executable files (PE) can be recognized on the machine by the usage of AI calculations.[5]	Examined 28 features extracted from metadata Not detect portable executable files for different OS platforms

This paper detects the malware where function extraction is based on opcodes inside the executable report of the malware. Opcodes of the malware might be extracted into a text File using IDA pro.[4]	This paper does not work on the Real set of malware.
Deep studying demonstrated splendid overall performance in image recognition. Conversion into grayscale images from malicious code. Achieved accuracy and speed.[3]	Conversion into color image not possible from malicious code
The fundamental undertaking of this work was to research the usefulness of filter-primarily based FSTs, which include the DFS, MI, CPD, and DIA in classifying the PEPE files as good or malware.[2]	
This paper supplied G3MD, a novel method for the detection of metamorphic malware, specifically for viruses & worms Conducted experiments on three families of viruses specifically NGVCK, G2, and MPCGEN.[1]	I)Not practice deep learning strategies to construct a computational model ii)Not use inference-based professional structures for common sub-graph-based malware detection

IV. PROPOSED SYSTEM ARCHITECTURE

This segment provides research directions. Focus on the malware detection of executable files and propose a primary malware detection module

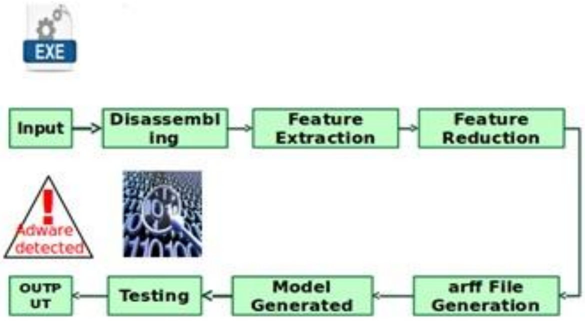


Fig. 2 Malware Detection System

In this paper, we focus on exploiting machine learning methods, malicious software Detection.

A. Input Executable file starts downloading when the user clicks on ads, the system starts working on the downloaded executable file. Data set generation block. The data set is created using two types of .exe files as input: i) Adware Files ii) Normal Files

B. Dissembling and Opcode Extraction block The composed programs are disassembled to get instruction set data using the N-Disassembler.

C. Feature Extraction block It is performed by using the TF algorithm for the calculation of frequently occurring words within a document. Also, the algorithm is used for the reduction of features.

D. Feature Reduction .arff file is generated using the algorithm and WEKA tool.

E. Testing block The generated internal representations are then tested with the existing data set, and then testing is performed.

F. Output Displays the warning message if the executable software is malicious.

V. RESULT AND DISCUSSION

A. Login Window

In Fig. 3 login page one login as a administrator login and second login is user login.

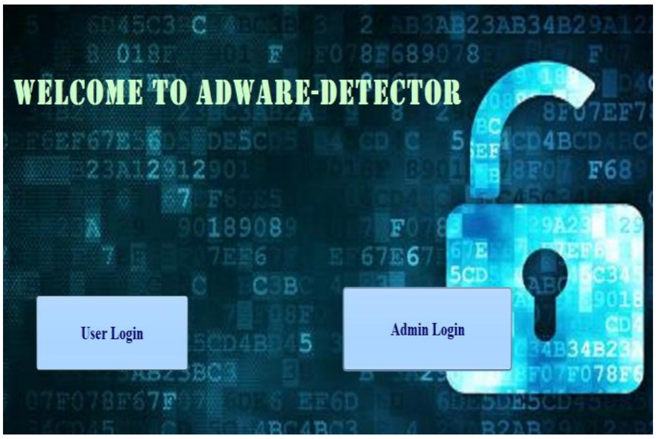


Fig. 3 Login window

B. Browsing File and First Phase Scanning

In Fig. 4 Browsing executable file and perform first phase scanning on executable file.

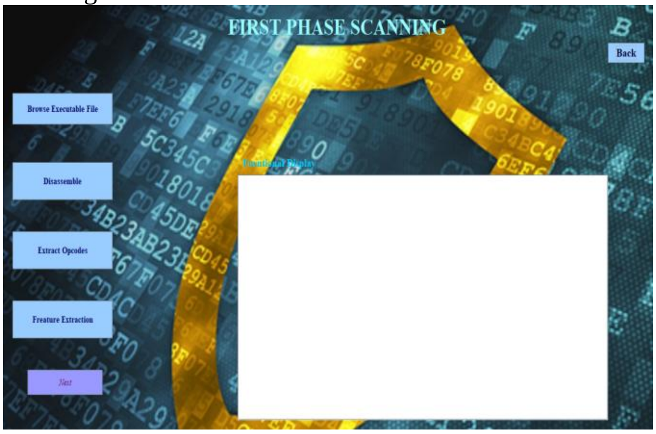
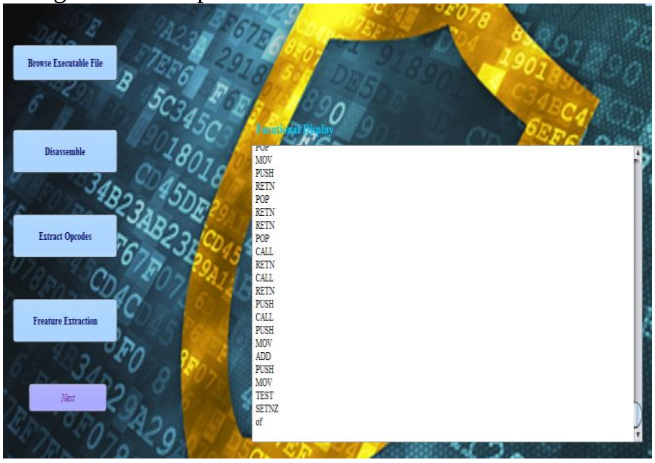


Fig. 4 Browsing File and First phase scanning

C. Extract opcodes

In Fig. 5 Extract opcodes from executable file.



```
POP
MOV
PUSH
RETN
POP
RETN
RETN
POP
CALL
RETN
CALL
RETN
PUSH
CALL
PUSH
MOV
ADD
PUSH
MOV
TEST
SETNZ
of
```

Fig. 5 Extract opcodes from executable file

D. Second Phase Scanning

Fig. 6 performs Second phase scanning operations.

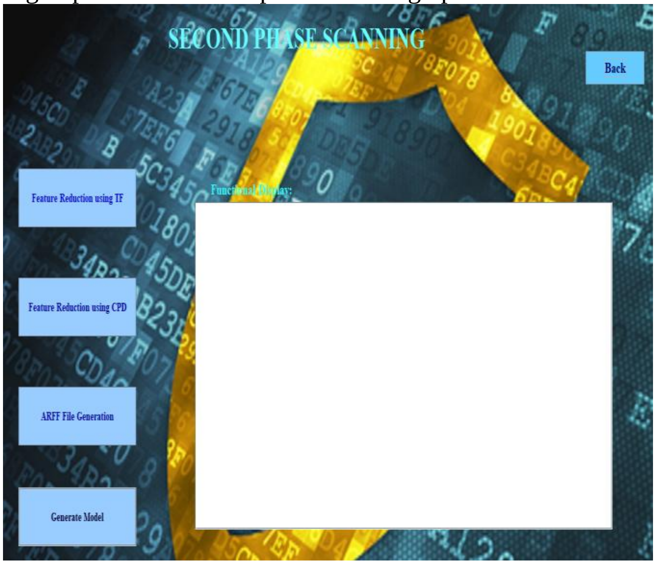
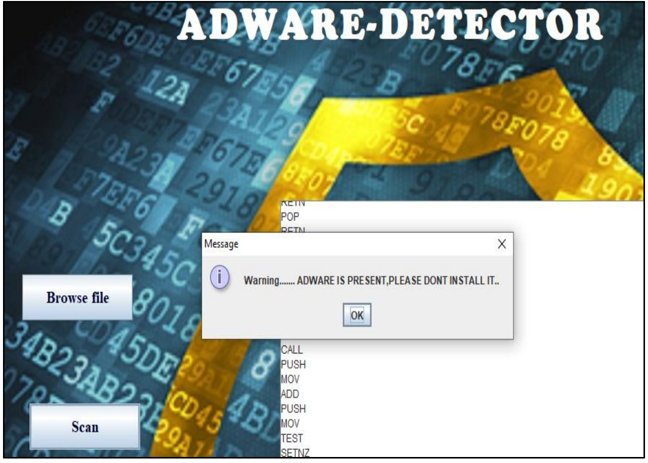


Fig. 6 Second phase scanning

E. Display Warning Message

Fig. 7 display message executable file is malicious or not (adware or malware)



Warning....... ADWARE IS PRESENT,PLEASE DONT INSTALL IT..

```
RETN
POP
RETN
CALL
PUSH
MOV
ADD
PUSH
MOV
TEST
SETNZ
```

Fig. 7 Display warning message

VI. CONCLUSION

This paper examines data mining and machine learning techniques for the detection of malicious software and how to show caution to the user. Also efficient way to detect known and unknown malware instances. Every one software program is in executable structure, so enter to the system is an executable file. Which can be malware files or benign files. Then testing is carried out to produce a warning if software consists of malware in it. Here check only one file in future design use own large dataset.

REFERENCES

1. Alireza Khaliliana, Amir Nourazar," G3MD Mining frequent opcode sub-graphs for metamorphic malware detection of existing families" 0957-4174/ Elsevier 2018.
2. Shiva Darshan S.L. and Jaidhar C.D." Performance Eval uation of Filter-based Feature Selection Techniques in Classifying Portable Executable Files" 6th International Conference on Smart Computing and Communications, ICSCC 2017,Kurukshetra, India, 7-8 December 2017
3. Zhihua Cui, Zhihua Cui, " Detection of Malicious Code Variants Based on Deep Learning"IEEE TRANSACTIONS ON INDUSTRIAL INFORMAT-ICS, VOL. 14, NO. 7, JULY 2018.
4. Prapulla S B, Sharad J Bhat "Framework for Detecting Metamorphic Malware based ob opcode Feature Extraction " 2nd IEEE International Conference on Computational Systems and Informational Technology for Sustainable Solution 2017.
5. Rushabh Vyas, Xiao Luo, Nichole McFarland, Connie Justice" Investigation of malicious Portable Executable File Detection on the Network using Supervised Learning Techniques "2nd International Workshop on Analytics for Network and Service Management 2017.
6. Farnoush Manavi, "A New Method for Malware Detection Using Opcode Visualization" IEEE Artificial Intelligence and Signal Processing Conference (AISP) 2017.
7. Lavesson,R. K. Shahzad "Accurate Adware Detection using Opcode Sequence Extraction" Sixth International Conference on Availability, Reliability and Security IEEE 2011.
8. M. G. Schultz, E. Eskin,"Data Mining Methods for Detection of New Malicious Executables" International Conference on Availability, Reliability, and Security 2010.
9. Abubakr Sirageldin, Baharum Baharudin, Low Tang Jung" Detecting Malicious Executable File Via Graph Comparison Using Support Vector Machine"International Conference on a Computer and Information Sciencem,IEEE 2012.
10. Boris Rozenberg, Ehud Gudes,"A Method for Detecting Unknown Malicious Executables"international Joint Conference of IEEE TrustCom11/IEEE ICESS- 11/FCST-11 2011.
11. TE-EN WEI,"REPEF – A SYSTEM FOR RESTORING PACKED EXECUTABLE FILE FOR MALWARE ANALYSIS"International Conference on Machine Learning and Cybernetics, Guilin, 10-13 July 2011.
12. K. Ramamoorthy, " Disassembled Code Analyzer for Malware " in Knowledge and Information Systems, IEEE, 2005.
13. S. Gordon, " Fighting Spyware And Adware" in The Journal, 2005.
14. R. K. Shahzad, "Detection of Spyware by Mining Executable Files" Fifth International Conference on Availability, Reliability and Security,2010
15. X. Chenglong, " Malicious Code Detection using Opcode Running Tree Representation." International Conference on P2P,Parallel, Grid, Cloud and Internet Computing.2012.
16. H. Hashemi,"A Survey on Heuristic Malware Detection Techniques "5th Conference on Information and Knowledge Technology ,2013.
17. A. Sulaiman, K. Ramamoorthy,"Malware Examiner, using Disassembled Code (MEDIC)" Workshop on Information Assurance and Security United States Military Academy, West Point, NY,IEEE 2005.
18. R. K. Shahzad,"Detecting Scareware by Mining Variable Length Instruction Sequences" International Conference on Availability, Reliability, and Security,IEEE,2010
19. Milan Jain1, Punam Bajaj2," Techniques in Detection and Analyzing Malware Executables: A Review," Milan Jainet al, International Journal of Computer Science and Mobile Computing, May 2014."

AUTHORS PROFILE

Manoj D. Shelar, completed the B.E degree in Information Technology Engineering from Vidya Pratishthan's college of engineering Baramati (Pune University, India) in 2008 and the M.E (IT) degree in Information Technology Engineering from Sinhgad College of Engineering Vadgaon Pune (Pune University, India) in 2011. He is Research Scholar in the Department of Computer Science and Engineering at Koneru Lakshmaiah Education Foundation Vaddeswaram, India. He is currently working as Assistant Professor in Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering & Technology (Department of Computer Engineering), VPKBIET, Baramati, Maharashtra, India. He is the Life member of CSI and ISTE.

Dr. S. Srinivasa Rao, completed Ph.D in computer science and engineering from Acharya Nagarjuna University (Guntur, India) and the M.Tech (cse) KLCE (Acharya Nagarjuna University, Guntur, India). He is currently working as Associate Dean (P&D) and Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, Andhra Pradesh, India. He received “Best teacher Award” and “Gold medal” twice in the year 2009-2010 and 2010-2011 from CSE dept, KL University. He is the Life member of CSI and Member of ACM, Member of CSTA.