Hybrid model for Computer Viruses: an Approach towards Ideal Behavior

Ankur singh bist
G.B. Pant University of Agriculture &Technology,
Pantnagar, India

ABSTRACT

Computer viruses are big threat for our society .The expansion of various new viruses of varying forms make the prevention quite tuff .Here we proposed an ideal hybrid model to prevent from computer viruses. The proposed model try to produce the environment for static and nature changing viruses.

Keywords

Polymorphic, false negative ,false positive, hybrid.

1. INTRODUCTION

Computer virus detection has evolved into malware detection since Cohen first formalized the virus in 1983[1]. Since the beginning , there is a big contest between virus creators and experts and it is becoming more complicating everyday ,and will continue afterwards. This is a quite general that the major cause for this the new tactics get developed by virus designers every time. There are various issue that make the pure antiviral software development harder. In this paper we will take a look on the various detection methods and finally will propose our hybrid model for virus detection.

2. RELATED WORK

A look of work consisting has been done in this field . It can be summarized by defining in term of various generation. First generation scanner including optimizing techniques, it includes simple scanning, bookmarks and speed up techniques which includes further wildcards, mismatch, generic degree optimizing technique, speed up techniques are further classified into hashing, top and tail scanning ,entry point/ fixed point, second generation scanners include smart scanning ,skeleton detection, near exact identification and heuristic analysis . Virus specific detection includes general and optimizing techniques which further include filtering ,static decryptor detection and x ray scanning. At last code emulation which includes generic detection and dynamic decryptor method.

3. PARAMETER FOR DEFINING PERFECT VIRUS CHARACTERISTICS

1 promise perfect disinfectant

2 scanning speed must be good

3 detects virus family

4 new or unknown virus detection

5 encrypted/polymorphic virus detection

6 metamorphic and macro virus detection

7. false negative and false positive detection

4. RELATIVE ANALYSIS OF WEAK POINT OF ANTIVIRUS TECHNIQUE

The first generation scanners work good for known viruses but for zero day viruses or unknown viruses ,it does not hold good , the second generation scanners include technique like heuristics, work for new viruses but problem of false positive and false negative is big threat ,the virus specific detection and code emulation suffer from the same problem. To predict the type and nature of incoming unknown viruses fall under a kind of undecidability, but any kind of undecidability can not be allowed by us for destroying the synchronization created by us.

5. PROPOSED ARCHITECTURE

Our proposed architecture is a hybrid model consisting of various elements to create a environment such that it can recover from the unavoidable situation created by viruses. The various approaches like artificial intelligence approach using case based reasoning are trying to recover with known and unknown viruses. Data mining approaches like n gram, naïve bias classifier are being used to find the unknown viruses.

There are various phases in the proposed model. In the **phase1** virus detection is based on checking their signatures . It is the basic approach of identifying the viruses once the appropriate sequence of bytes also called signature found appropriately then it leads to proper identification of viruses .Some example of virus which are published in virus bulletin [2]are Accom.128, die.448.

While we see the special case of string matching it includes the wildcards, mismatches, generic degree .Use of bookmark is a simple way to ensure a more reliable detection and decrease the risk of false positive. Further speed up technology [3]algorithms are hashing[4] ,it makes the access of the searching data faster ,top and tail includes scanning the first and last part of the data .these type of approaches get used in the first phase the purpose of first phase is to find all the viruses of known signature efficiently.

Phase 2 , includes the three sub phases in it ,In the **first sub phase** it uses second generation approaches like smart scanning ,it refers to a defense optimizing method for the newer generation of viruses, which try to disguise their code within a sequence of worthless instructions such as no operation nop instructions[5]. Another approach of skeleton detection is specially effective in order of detection of macro viruses it does not utilize strings or checksums for detection purpose. Firstly in it parsing get performed an unusual entities are removed . Hence the skeleton of the code will be remained containing of only fundamental macro code which the scanner exploit it to detect the viruses[6]. Another is Nearly Exact Identification ,in this method more stronger methodology is used ,it involves two string for identification. The virus is located in if both string get occurred in file. Therefore, it creates an environment where process of virus identification is more suitable and increase the probability that virus get identified and classified properly . Combination with bookmarks increments utility of this technique. The exact identification technique utilizes constant bytes in the virus code as many as required to find a checksum of all bytes in the virus program, which contains non variable value. The mapping is done for every constant byte rather than variable bytes. This is the only method, which can promise an accurate detection of virus variants. It is generally implemented as combination with the techniques of the first generation virus detectors. Exact identification method can classify exactly among many types of a virus, as well. Even though it has various benefits, but implementation of this technique leads to make scanner work slowly, slightly. In addition, really it is hard to implement it for the oversized computer viruses. Another technique in it is heuristics Analysis, The heuristics analysis is a useful method for detection of new unknown malwares [7]. It is especially helpful for detection of macro viruses too. It can be so worthwhile for binary viruses, as well, but it may extremely produce false positive output that is a major drawback of scanners . Entry-Point and Fixed-Point scanning [8],These techniques also help the scanning engines to execute more rapidly. They use the concept of the program execution entry-point, which is achievable by the headers of executable files. Because viruses the usually seek entry-point of the file as a target, search can be started from this point. In order to keep the execution of a file in normal manner, the virus has to get the executing control from the original start point and pass it again to the original entry point of infected file after it finishes its code subroutine. Fixed-point scanning is widely used when there is not many helpful strings in the entry point. The scanner firstly specifies an initial location and works for finding it in every position which is determined by adding the initial position with some constant factor associated with it and thus it evolve itself.

Second sub phase :- Virus-specific Detection ,Sometimes the general methods and algorithmic procedures do not work properly for specific viruses. In such conditions, a virus specific detection algorithm must be created and properly implemented to carry out detection procedure. Actually, this kind of detection is not used generally, but it denotes any special technique that is specifically modeled for a given particular virus. This approach is also called algorithmic scanning, but because it can be misleading [9], we use virus-specific detection term instead of algorithmic scanning. This technique may arise vital problems such as portability of the scanner on various platforms and stability of the code. To overcome these problems, virus-scanning languages have been developed that in their plainest form, seeking and reading operations in scanned objects are allowed.

Subphaze3: Code Emulation ,it is one of the best detection techniques. It simulates the storage resources, main memory, computer central processor and some essential functions of operating system by a virtual machine to execute the malware virtually and identify its behavior and performance. The harmful code does not execute on real machine and it is controlled by the virtual machine precisely, therefore there is no risk for unintentionally replication of malware. The emulator resembles instructions of the machine by simulating CPU registers and flags, virtually. It appears to be like the execution of programs and detection process analyzes all instructions, individually. For polymorphic viruses or other types of encrypted codes [10], after a given quantity of iterations or after a pre-defined stop situation, the scanner checks the contents of memory of the virtual machine. After sufficient iterations, polymorphic virus will decrypt its encrypted body and the real code will be revealed in the virtual memory. Scanner may use the following methods to choose when it breaks off the emulation loop: Stopping with break points, Tracking of decryptor using profiles, and Tracking of active instructions. When the emulation terminates, the virus will be checked by using string pattern matching or other scanning techniques [11]. Veldman in called more generally this method as Generic Detection, in the case of any kind of encrypted malwares. He describes it as a way to decrypt an encrypted virus. Essentially, a generic detection consists of four parts: processor emulator, memory emulator, system emulator, and decision mechanism [11].Some malwares have been designed by their designers that their behavior is very unpredictable and while these type of malicious codes finds the presence of emulator then they tries to conceal their behavior. Beyond this dynamic decryptor detection includes the combination of code emulation and static decryptor detection technique, and due to this appropriate formation this technique provides better results.

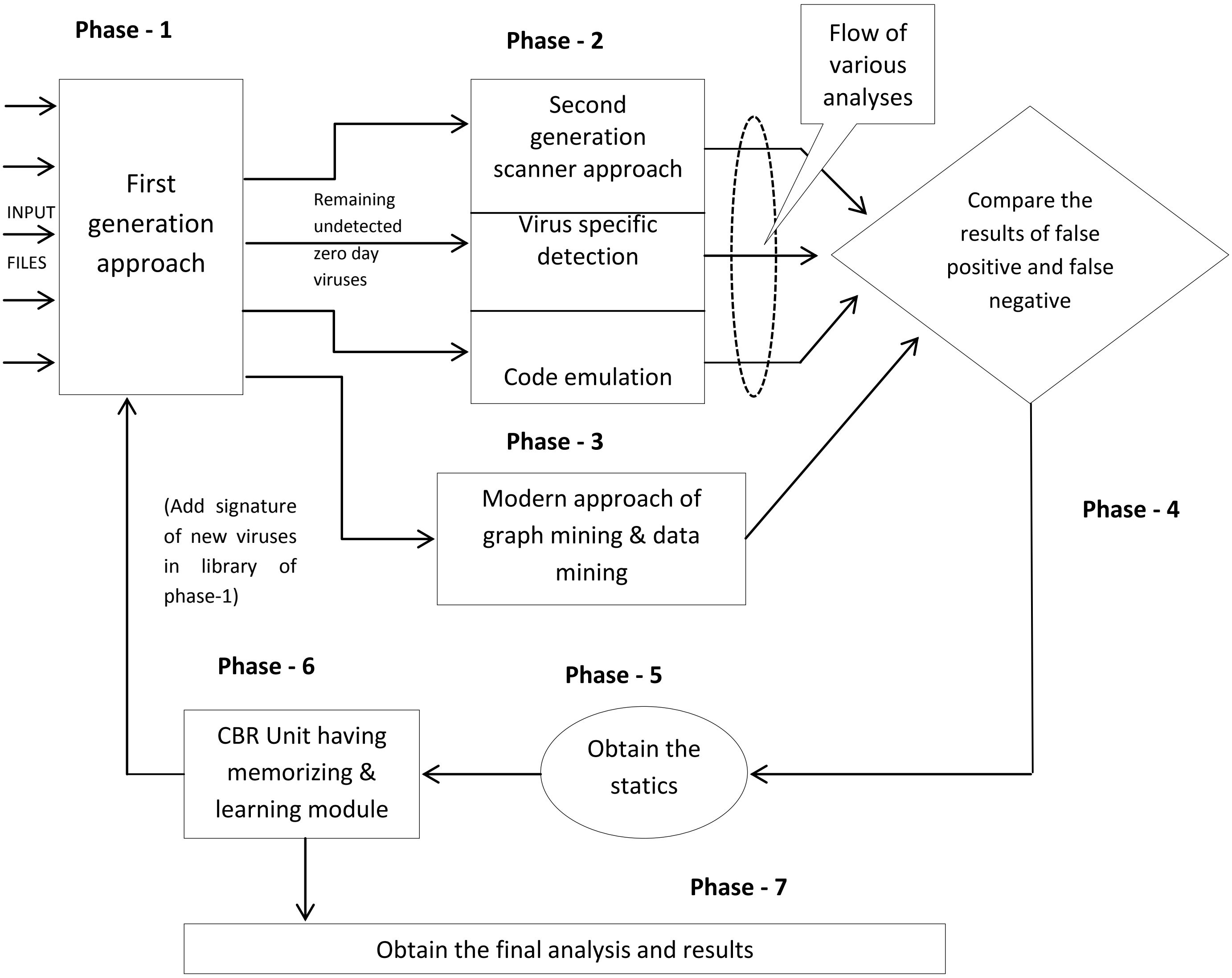
Phase - 1
Phase - 2
Flow of various analyses
INPUT
FILES
First generation approach
Remaining undetected zero day viruses
Second generation scanner approach
Virus specific detection
Code emulation
Compare the results of false positive and false negative
Phase - 3
Modern approach of graph mining & data mining
(Add signature of new viruses in library of phase-1)
Phase - 4
Phase - 6
Phase - 5
CBR Unit having memorizing & learning module
Obtain the statics
Phase - 7
Obtain the final analysis and results

Fig 1: hybrid antivirus scanner architecture

Phase3:- Using the latest approaches used in graph mining we can perform the detection in efficient manner it includes the identifying viruses extract behavior , graph preprocessing and frequent graph mining to find out certain behavior of viruses., modeling program semantic[12] and system details includes [12,13] the following parameters:- file access ,system information ,networking, registry access, process, system information are another entities in graph mining. Latest data includes application of n- gram and naïve base classifier . data mining approach includes malware analysis ,file size analysis ,disassembly, parsing, feature extraction, feature selection, independent test. Random forest, bagging, decision tree are different classifier used provide increment in overall accuracy. **Phase 5**,This phase is used for collecting the resulting comparing outcomes for further processing , the processing **phase 4** is very important in such away that it tries to break the negative aspect of false negative and false positive .It compares the results of various analysis then produce optimal results ,the various results are compared and comparatively categorized then selective criteria get performed on the basis of probability of occurrence.

phase 6 , it is the another important entity of our model using cbr unit (CASE BASED REASONING) [14], involving the methodology of artificial intelligence. In this method new virus That does not exist in the database can be detected , the updating of database can be done without connecting to database used by our application. this presents major advantage.cbr has various phases named as development, memorizing adaption[15] , revision, learning . The learning and memorizing part of this unit is very important because it adds a new virus in the database. **Phase7**, In this phase it is connected to phase 1 ,it is done to analyze the newly detected virus to phase 1 ,hence it involve artificial intelligence unit[16], it involves the undecidability for new viruses that create undecidability for us in term of identification so due to which we can not remove viruses totally.phase7 connected to phase 6 , it is connected to obtain final analysis and results.

6. EFFICIENCY OF PROPOSED MODEL

If we talk about the ideal behavior of the proposed Model , it depends on the behavior of each phases it consist of. The whole journey of viruses in these phases will bring out various features of viruses so that the varying and suspected viruses can be controlled. But in this hybrid model architecture , it can be stated that if the case based reasoning become so strong that the learning and other supporting modules of it take optimal decision and phase 1 and phase 2 works ideally ii means in such a manner that it performs its defined work with full efficiency which means pure disinfection then it will lead to an antivirus engine having tremendous potential of self updation and bringing new revealing feature .

7. COMPARATIVE ANALYSIS OF EXISTING TECHNIQUES

Different techniques are trying to control and prevent the negative aspect of viruses .Conventional methods are evolving with the assessment of new techniques. Existing techniques like random forest ,bagging, decision tree shows the overall accuracy 96%,93.8%and 90% respectively [17]. Several immunity based model for computer viruses detection are developed producing good results but the constraints associated with gene library still to be developed[18].Further the graph mining techniques gives 86% detection rate on new and unknown malwares[19]. These different data shows the various evolving defending mode against computer viruses but the stability and fertility against new viruses could not achieve till now ,in our proposed model our hybrid approach arranges the different techniques in well synchronized manner when the certain problem cannot be solved under one frame of reference then hybrid solution produced as given by our model become efficient also our model comparative zones at different phases fight against biggest obstacle of false negative and false positive so it makes our model behave and producing results in different orientations.

8. CONCLUSION

Most of detection method are not powerful against evolutionary new viruses ,and need of antivirus updation is necessary to work it correctly we arrange the composition of various scanner and techniques together to find new viruses and comparative analysis to overcome the false positive and false negative rate of scanners and the self updation criteria can be enhanced.

9. FUTURE RECOMMENDATION

Scanning process usually take considerable amount of time to search system for patterns we are using several modules together to lead to an big approach for detection. In the future work is to be done for cbr module improvement and the time factor reduction of whole system then it will become of pure disinfectant which provides results very frequently.

10. REFERENCES

[1] F. Cohen. Viruses . PhD. Thesis, university of southern California 1985.

[2] Vb," ibm viruses(update)",virus vol. no.1999,pp.5-6.

[3] Szor,p.,the art of computer viruses research and defense ,adison –Wesley ,professional,2005.

[4] Erdogan,o. and p.cao, "hash-av:fastvirus signature scanning by cache –resident filters", in ieee global telecommunication conference.

[5] Szor, P. and P. Ferrie, "Hunting for Metamorphic", in 11th

[6] Virus Bulletin International Conference, 2001, pp. 123-144.

[7] Catalin,b. and A.vi oiu,"optimiozation of antivirus"

[8] Bidgali,h., handbook of information security ,wiley 2006.

[9] Jordan, M., "Dealing with Metamorphism", Virus Bulletin, October 2002, pp. 4-6

[10] Aycock , j.computer viruses and malware, ab,Canada:springer,2006.

[11] Veldman, F., "Generic Decryptors Emulators of the future", in IVPC conference, 1998.

[12] S.forest,S. hofmeyr,A.Somayaji, and t.longstaff,"A sense of self for Unix procesess".in IEEE symposium on security andprivacy.1996.

[13] D. Wanger ,and d. dean .intrusion detection via static analysis". In IEEE symposium on security and privacy.2001.

[14] Madihah mohd Saudi shaharudin Ismail "an efficient control of virus propagation"

[15] Christina Carrick and Qiang Yang(Simon Fraser University Burnaby, BC, Canada, V5A 1S6) , Irene Abi-Zeid and Luc Lamontagne(Defense Research Establishment Valcartier Decision Support Technology 2459, boul. Pie XI, nord Val Belair, Quebec, Canada, G3J 1X5),'' Activating CBR Systems through Autonomous'', Springer-Verlag Berlin Heidelberg 1999 .

[16] Magda liliana RUIZ ORDONEZ (Girona university),'' Multivariate statistical process control and case-based reasoning for situation assessment of sequencing barch reactors'',ISBN:978-84-691-6833-2,Dipositlegal:GI-1299-2008.

[17] M. siddiqui,M.C Wang and J.Lee data mining methods for malware detection using intrusion techniques in proceedings of artificial intelligence,2008.

[18] J.O Kephart ,"a biologically inspired immune for computers", proceedings on conference synthesis and simulation of living system,1994,pp. 985-996.

[19] M.fredrikson and s. jha,R.sailer,X.yan."Synthesis near optimal malware specification from suspicious behavior".2010,pp.45-60.