Review on the application of Artificial Intelligence in Antivirus Detection System[i]

Xiao-bin Wang Guang-yuan Yang Yi-chao Li Dan Liu
School of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu, China
xbwang@uestc.edu.cn to_yangguangyuan@163.com richardlyc@uestc.edu.cn liudan@uestc.edu.cn

***Abstract*—Artificial intelligence (AI) techniques have played increasingly important role in anti-virus detection. At present, some principal Artificial Intelligence Techniques applied in anti-virus detection are proposed, including Heuristic technique, data mining, Agent technique, artificial immune, and artificial neural network. It believes that it will improve the performance of anti-virus detection systems, and promote the production of new Artificial Intelligence Algorithm and the application in anti-virus detection to integrate anti-virus detection with Artificial Intelligence. This paper introduces the main artificial intelligence technologies, which have been applied in anti-virus system. Meanwhile, it also points out a fact that combining all kinds of Artificial Intelligence technologies will become the main development trend in the field of anti-virus.**

***Keywords*—Anti-virus, Artificial intelligence, Heuristic, Data mining, Neural network**

I. INTRODUCTION

As the application of computer and Internet is more popular, it provides a convenient way to share the information among different people, however it also gives chances to malware activities, such as propagating malicious programs, including computer viruses [1].Computer virus is a program, which invades your computer, as same as a biological virus injects an organism. It executes malicious functions to damage computer system.

Now, the main technologies of anti-virus include Virus Prevention Technology, Virus Detection Technology, Virus Elimination Technology, and Virus Immune Technology. However, the former three technologies have considerable progress compared to the last one for the lack of general solution.

Hence, the process of handling virus includes four steps: Virus Detection, Virus Elimination, Virus Prevention, and Virus Immunity. Figure 1 shows the process of handling virus.

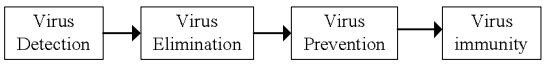

Figure 1. The Process of Handing Virus

Virus Detection is a process that takes the suspicious file (including the boot content, the memory content and others content) as its input, perform specific detecting arithmetic, and finally output the detection result that shows whether it includes virus or not[2]. There are two types of virus detection technology. One is Monitor Technology that detects Computer Virus Signatures including Virus Keyword, the content of Signature Program Segment, the method of infection, the changing of the file length and so on. Another is File-Self Detection Technology that detects the signature of the file-self instead of virus signature.

Virus Elimination eliminates virus from the program with virus to make it a benign program that can execute normally. Strictly, Virus Elimination is an extension of Virus Detection, because it only can eliminate virus after being detected.

Virus Prevention is that it will block the invasion or give warning while virus has not yet invasion or at the right time it starts to invade. Virus Prevention includes both the prevention of known virus and the prevention of unknown virus. Virus Prevention Technology includes Disk Boot District Prevention, Encryption the Executable Program, Read-Write Control Technology, System Monitor Technology and so on.

Virus Immune derives from Biological Immune Technology. It can defense all viruses in independent of Virus Database Update.

II. ARTIFICIAL INTELLIGENCE

The Dartmouth Summer Research Project on Artificial Intelligence created AI as a research discipline in 1956. AI is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. Artificial intelligence has been 50 years. Artificial intelligence is a developing comprehensive marginal subject and has become one of the world's three major high-techs, including Space Technology and Energy Technology since the 1970s.Further more, it has become one of the world's three major high-techs including Gene Engineering and Nano Science in the 21st century.

The main research of Artificial Intelligence is how to make the computer simulate some thinking process and intelligent behavior that belong to human, such as learning, reasoning, thinking, and planning, etc. Then intelligent computers similar to the human brain will be produced for making the computer to achieve more advanced application.

With the development of anti-virus systems and Artificial Intelligence, as a new anti-virus technology, Artificial Intelligence has been applied in anti-virus engines. The main research field of Artificial Intelligence, such as Heuristic

Technology, Artificial Neural Networks etc, has become the main technique of anti-virus system. With the development of Artificial Intelligence Technology, more advanced methods, such as Data Mining Technology, Artificial Immune Technology etc, have been applied in the new generation of anti-virus detection system and play a crucial role in improving the anti-virus's performance and veracity. Artificial Intelligence Technology is applied in the major anti-virus detection steps including virus detection, virus analysis and virus immune. As following, we will introduce the main technologies of Artificial Intelligence, which are applied in anti-virus detection system.

III. THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN ANTIVIRUS DETECTION SYSTEM

A. Heuristic Technology

At present, in the field of detecting computer virus, the first and most popularity application of Artificial Intelligence Technology is Heuristic Technology. Heuristic Technology means "the ability of self-discovery" or "the knowledge and skills that use some methods to determine", and intelligently analyze codes to detect the unknown virus by some rules while scanning [3]. It is very complicated while implementing in reality because it should identify and detect many suspicious code instruction sequences, such as format diskette operation, searching and locating various operations of executable programs, implementing resident memory operate and discovering unusual or unopened system function calls. In actual, virus detection software using Heuristic Scanning Technology is a dynamic higher or an anti-compiler by decompiling some relatable instruction sequences to understand and confirm the real motivation of them.

Symantec Inc. introduces that what is Heuristic Technology as well as its application of anti-virus in detail. And the future application of Heuristic Technology in Antivirus Detected System is presented [4]. Authors of [3] presented a heuristic skill of computer virus analysis based on virtual machine. Similar to the process of pattern identification, Virtual machine decodes virus and provides signature. And then Heuristic scanning method classifies those signatures. Authors of [5] presented a new method for anti-virus engine based on heuristic strategy. Detecting and killing polymorphic virus is the target of the model. In order to improve the effect of anti-virus engine, it provides matter for heuristic search by analyzing the behavior characteristic of virus and extracting the key behavior characteristic.

In a word, although heuristic analysis has shortcoming and deficiency, it represents the future development trend of anti-virus technique. And it is a new developing and improving technology. Comparing to other scanning analysis technique, it can almost provide enough supplementary information to determine whether the detected target contains virus or not.

Simultaneously, it also cause false positive and false negative. The best way to avoid this is integrating heuristic method with other traditional scanning methods. Further, heuristic analysis has paved the way for future research of general anti-virus technique.

B. Data mining Technique

With the rapid development of Information Technology, the rapid growth of data has exceeded the ability of the manual processing of data. So how to help people to extract the general knowledge from the mass of data has become more and more important. In order to implement it, data mining technique is put forward and soon becomes an active research direction.

Data mining analyzes the observed sets to discover the unknown relation and sum up the results of data analysis to make the owner of data to understand. Data Mining Algorithm that is from Statistics, Pattern Identification, Machine Learning, and Database and so on, has developed comprehensive.

Authors of [6] presented a data-mining framework that detects new, previously unseen malicious executables accurately and automatically. The data-mining framework automatically found patterns in our data set and used these patterns to detect a set of new malicious binaries. Comparing our detection methods with a traditional signature based method; our method more than doubles the current detection rates for new malicious executables. Yufeng Yang presented a network virus precaution system based on data mining shown in Figure 2.It can detect the abnormal connecting behavior of network in real-time to discover the trace of worm virus, especially the precaution action to the new worm virus to make administrator to adopt corresponding measure to avoid tremendous loss[7]. Authors of [8] proposed an innovative technique for detecting the presence of an unknown worm, not necessarily by recognizing specific instances of the worm, but rather based on the computer measurements. The concept of detecting unknown computer worms is base on a host behavior, using Data Mining algorithms.

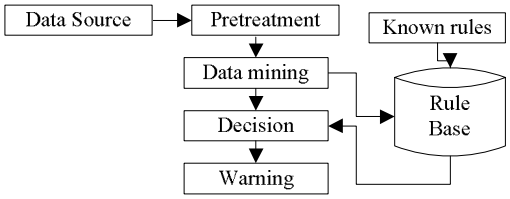

Figure 2. The structure of warning system

In conclusion, as the development of data mining technique, more and more industry will utilize it to extract the undiscovered data they really want to need. Simultaneity, the new development of data mining technique will be brought. Of course, data mining technique will be applied in anti-virus in depth with the high-speed development of virus technique.

C. Agent Technology

Intelligent anti-virus system always is the goal of anti-virus field. With appearance of the new unknown malicious virus, new methods of anti-virus generate very slowly. Thus let new unknown malicious virus spread quickly without holdback. So there is growing hope that the computer can automatically and efficiently respond to the virus. In the background, the appearance of Agent in the field of Artificial Intelligence resolves this issue satisfactorily.

Authors of [9] presented a method named VICEd that is a system for generic virus detection over the Internet. VICEd is based on a virus detection methodology which is a combination of software emulation and knowledge base. It detects viruses using their behavior instead of pattern matching. It is thus more effective against unknown or mutated viruses than scanners. This methodology is interesting in its own right. T. Okamoto and Y. Ishida proposed a distributed approach against computer virus using the computer network that allows distributed and agent-based approach. Their anti-virus system consists of several heterogeneous agents similarly to the immune system. Among these agents, antibody agents use the information of "self" (files of host computer) rather than the information of "non-self" (computer viruses). After detection and neutralization of computer viruses, the anti-virus system tries to recover original files that are distributed over the uninfected hosts connected by LAN. This recovery is also done by several heterogeneous agents. As a whole, the proposed anti-virus system can be regarded a backup system with computer network [10]. Authors of [11] introduce what is agent that includes the operating environment of agent and its cooperation mechanism in detail (Figure 3 shows them), the application of Agent technique, and the future development of agent. And the particular process of using Agent to deal with virus has been given in Figure 4. Authors of [12] proposed a Multi-agent system for Worm Detection and Containment in MAN (MWDCM) to provide a first-class automatic reaction mechanism that automatically applies containment strategies to block the propagation of the worms and to protect MAN against worm scan that wastes a lot of network bandwidth and crashes the routers.

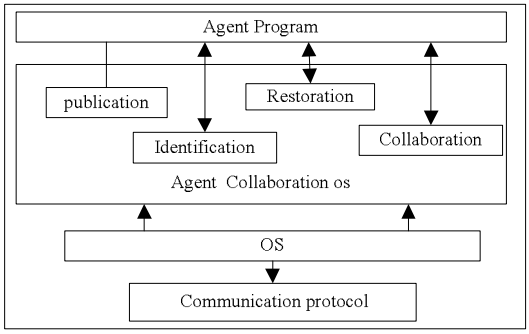

Figure 3. Agent Collaboration Mechanism

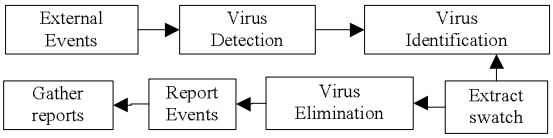

Figure 4. The process of automatically handling virus

D. *Artificial Immune Technology*

The traditional methods of anti-virus including character-code analysis method, check-sum analysis method, behavior-inspect analysis method and so on have played an important role in the history of virus detection. But the traditional methods of anti-virus can not detect successful the new unknown virus with the popularity of the Internet and advanced technique of virus increasing. Faced with this situation, it is important to seek a new method of virus detection can be fast, accurate, and effective for detection. So the appearance of Computer Virus immune technologies that derive from the principle of Biological immune system that can withstand and destroy Biological Virus the unknown implements the new requirement of virus detection.

IBM has developed a prototype computer immune system in 1996[13].Thus, the anti-virus filed is starting to research and apply Computer Immune Technology deeply. Authors of [21] presented a distributed architecture for an adaptive computer virus immune system. It used new technologies (such as evolutionary algorithms and intelligent agents) to detect unknown virus. It also provided an effective and efficient CVIS (computer virus immune systems). Authors of [14] proposed a self-adaptive distributed agent-based defense immune system based on biological strategies and develop within a hierarchical layered architecture. And the system design integrates the power, flexibility, adaption, and capabilities of the BIS into architecture realizable in the information system domain. In order to upgrade the existing algorithms which only aim to detecting unknown viruses, Authors of [15] presented a new, immure-associative-memory-based virus detection algorithm by incorporating the non-self idea of immune first reaction and the associative memory idea of secondary reaction. The algorithm combines the recognition ability of immune non-self selection for unknown viruses and the memory ability of immune associative memory for known viruses. It detects unknown viruses and variation viruses effectively, as well as those known viruses with the similar behaviors.

E. *Artificial Neural Networks*

Now, the popular methods of virus detection all can not automatically extract virus signatures. And the whole virus detection system lack of associative memory and the capacity of real-time calculation also can not automatically detect and learn, and can not engage in large-scale parallel processing [22]. For solving the difficulty of anti-virus field, Antivirus experts present Artificial Neural Networks.

Artificial Neural Networks simulating behavior features of Animal Neural Networks that parallel processes distributed information is a mathematical model algorithm. Artificial Neural Network with large-scale parallel and distributed storage and processing, self-organizing, adaptive and self-learning ability, in particular applies to handle the problem that must consider many factors including conditionality, imprecision and fuzziness and so on at the same time. In the early 1990s, Guinier presented a virus identifying way based on Artificial Neural Networks because of its virtues. And now virus detection technique based on Artificial Neural Network has been applied in real.

Authors of [16] has introduced how to use Single layer neural classifier to detection boot viruses, and the generic virus detector was incorporated into IBM Antivirus in May, 1994.Its structure has been shown in Figure 5. William Arnold and Gerald Tesauro constructed multiple neural network classifiers which can detect unknown Win32 viruses by combining the individual classifier outputs using a voting procedure, following a technique described in previous work (Kephart et al, 1995) on boot virus heuristics [17]. And the system has

achieved effectively. Authors of [18] presented a new rule generation method from neural networks formed using a genetic algorithm (GA) with virus infection and deterministic mutation. This method can extract rules (regularities) for a pattern classification and a chaotic system identification by using the same system. Authors of [19] put forward a new method of computer virus detection on back-propagation neural networks by researching the main virus detection methods as well as their fault. This new detection is more effective in analyzing system information including file system, and diagnosing which kind of computer virus are infected by compared with the traditional methods.

Authors of [20] proposed a method of automatically detecting malicious code using the n-gram analysis by researching the standard signature-based technique. The BP neural network is used in the process of building and testing the proposed multi-classifiers system after selecting features based on information gain. And the system based on this method is a generic virus detection system. Meanwhile, the experiment results produced by the proposed detection engine shows improvement of accuracy and generalization compared to the classification results of the individual classifier.

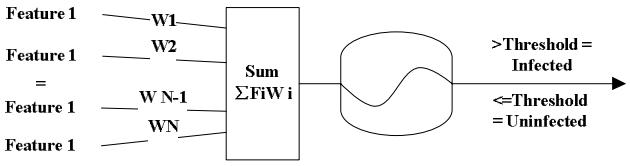

Figure 5. Single layer neural classifier

IV. CONCLUSION

Now, anti-virus technique has become the important prevention technique. Using these technologies, the system can detect virus invasion in real-time, and enlarge security management capacity of system administrators to enhance the integrity of the infrastructure of information security. However with the large appearance of computer virus, especially unknown and polymorphic, mutational virus, anti-virus system must accomplish more difficult things than before. So the traditional anti-virus technology has been unable to achieve the purpose of anti-virus detection. Fortunately, the continuous development of Artificial Intelligence technology has provided new methods and ideas for anti-virus detection system. Intergraded anti-virus detection with artificial intelligence will greatly improve the performance of the existing anti-virus detection system, promote more effective artificial intelligence algorithms to be proposed, and be applied in the popular anti-virus detection field. Moreover, combining all kinds of Artificial Intelligence technologies will become the main development trend in the field of anti-virus.

REFERENCES

[1] Xi Zhang, D. Saha, and Chen Hsiao-Hwa, “Analysis of Virus and Anti-Virus Spreading Dynamics,” IEEE GLOBECOM, 2005.

[2] Wentao Jiang, Jinfeng Liu, and Yifei He, “The application of Pattern Classification Technology in Computer Virus Detction,” http://www.xinxijishu.org/article/pc/jingyan/200607/2576.html, 2006.

[3] Xianwei Zeng, Zhijun Zhang, and Zhi Zhang, “Heuristic skill of computer virus analysis based on virtual machine,” Computer Applications and Software, Vol. 22(9), 2005, pp. 125-126.

[4] Symantec Corporation, Understanding Heuristics: Symantec’s Bloodhound Technology,” Symantec White Paper Series, 1997.

[5] Zhenhai Wang and Haifeng Wang, ”Study on Anti-Virus Engine Based on Heuristic Search of Polymorphic Virus Behavior,” Research and Exploration in Laboratory, Vol. 25(9), 2006, pp. 1089-1108.

[6] Matthew G. Schultz, Eleazar Eskin, Erez Zadok, and Salvatore I. Stolfo, “Data Mining Methods for Detection of New Malicious Executables,” The 2001 IEEE Symposium on Security and Privacy, Oakland. CA, 2001, pp.38-49.

[7] Yufeng Yang, “The Network Virus Precaution System Based on Data Mining,” Journal of Shaoguan University, Vol. 26(12), 2005, pp. 31-33.

[8] Robert Moskovitch, Ido Gus, Shay Pluderman, Dima Stopel, Chanan Glezer et al, “Detection of Unknown Computer Worms Activity Based on Computer Behavior using Data Mining,” Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2007), 2007, pp. 169-177.

[9] Lee J.S., Hsiang J., Tsang P.H., “A Generic Virus Detection Agent on the Internet',” Proceedings of the Thirtieth Hawaii International Conference on System Sciences, Wailca. HI, 1997, pp. 210-219.

[10] T. Okamoto and Y. Ishida, “A Distributed Approach to Computer Virus Detection and Neutralization by Autonomous Heterogeneous Agents,” Proc. of International Symposium on Autonomous Decentralized Systems (ISADS'99), 1999, pp. 328-331.

[11] Maoguang Wang, Zhaolin Yin, Helong Ding, and Xianzhong Zhang,“The Application of Agent in Virus Detection,” Network & Computer Security, 2002, pp. 55-57.

[12] Xiantai Gou, Weidong Jin, and Duo Zhao, "MULTI-AGENT SYSTEM FOR WORM DETECTION AND CONTAINMENT IN METROPOLITAN AREA NETWORKS," JOURNAL OF ELECTRONICS (CHINA), Vol. 23(2), 2006, pp. 259-265.

[13] IBM, "Combating computer viruses: IBM's new computer immune system," IEEE Parallel and Distributed Technology , 1996.

[14] Harmer P., Williams P., Gunsch G., and Lamont G.B., "An Artificial Immune System Architecture for Computer Security Applications," IEEE Transactions on Evolutionary Computation, Vol. 6(3), 2002, pp. 252-280.

[15] Zhen Yu; Jianhui Ma; Xianbin Cao, and Xufa Wang, “A Virus Detection Algorithm Based on Immune Associative Memory,” Vol. 34(2), 2004, pp. 246-252.

[16] Kephart, J.O., "Biologically inspired defenses against computer viruses," Proceedings of International Joint Conference on Artificial Intelligence, 1995, pp. 985-96.

[17] William Arnold and Gerald Tesauro, "Automatically generated Win32 heuristic virus detection," Proceedings of the 2000 International Virus Bulletin Conference, 2000, pp. 51-60.

[18] Fukumi M., Mitsukuwa Y., and Akamatsu N., "A new rule generation method from neural networks formed using a genetic algorithm with virus infection," Proceedings of the International Joint Conference on Neural Networks, Como. Italy, Vol. 3, 2000, pp. 413-418.

[19] Chen Guo; Jiarong Liang; and Meilian Liang, “Method of Virus Detection Based on BP Neural Networks,” Vol. 31(2), 2005, pp. 152-156.

[20] Boyun Zhang and Jianping Yin, Dingxing Zhang, Jingbo Hao, Shulin Wang, “Computer viruses detection based on ensemble neural network,” Computer Engineering and Applications, Vol. 43(13), 2007, pp. 26-29.

[21] Marmelstein R.E., Van Veldhuizen D.A., and Lamont G.B., "A Distributed Architecture for an Adaptive Computer Virus Immune System," IEEE International Conference on Systems, Man, and Cybernetics, Vol.4, 1998, pp. 3838-3843.

[22] White R., "Open Problem in Computer Virus Research," Virus Bulletin Conference, Munich Germany, 1998

---

i This research is partially supported by the National Information Security 242 Program of China under Grant No. 2007B30