A Novel Virus Detection and Active Defense Algorithm Based on SVM Optimized by Differential Evolution Algorithm

Jieqing Ai[1,2, a], Zhenyue Long[1,2, b], Shang Gao[1,2, c]

[1]Guangdong Power Grid Corporation Information Center, Guangdong,Guang Zhou,510000, China

[2] Guangdong Power Grid Software Testing Lab, Guangdong, Guangzhou,510000, China

[a]email:aijieqing@gdxx.csg.cn[b]email:longzhenyue@gdxx.csg.cn[c]email:gaoshang@gdxx.csg.cn

Keywords: Differential Evolution; SVM; Active Defense; Virus Detection; Information Security.

Abstract. This paper proposes a novel active defense strategy focuses on users' behavior patterns which to classify the behaviors accurately by SVM for virus detecting. Differential evolution was introduced to improve the precision of SVM and turns it into an optimization problem which object is the classification precision. And the parameters are regarded as the variables to be optimized. The experimental results show that the proposed model has a higher precision than the compared methods, such as BPNN, SVM, GA-SVM, etc. In addition, the method is more efficient so that, it can be quickly updated and applied.

Introduction

Existing network security technologies rely on firewalls, intrusion detection and anti-virus software, etc., these are static and passive security defense system[1].These models focus on defending attack and make responding when it was detected. The responding time is too long to avoid serious losses cause the lacking of active defense ability and the ability to predict an attack. In order to guarantee the security and robustness, the world's governments, academia and business networks are looking for new defense technologies.

Active real-time protection mode includes situational awareness, risk assessment, safety testing and other measurements [2][3]. This is asafety system that judges the current network security situation and implements the active defense according to the judgments. In the field of information security, how to against malicious programs by using the active defense system is a hotspot [4].

Active Defense Strategy

We can know that common patterns exist in the same family of malware execution behavior called the behavior patterns, based on long experience of the anti-virus industry. For example, all the variations of the "Allaple" worm[5]will on infected systems to acquire and lock a particular semaphore. Hence, a new technology has become a hot topic called analysis technique that can identify patterns of behavior similar process by analyzing the relevance of the program execution behaviors.

The current behavior analysis technologies are focusing on the malware that is means to discover and identify the behavior patterns and judgment whether these are belong to an unknown virus according to this model. However, the number of families of malicious programs is very large and the emergence of new family will occur probably with the development of computer technology. So, it is very inefficient if we enhance the protection of malicious programs by learning patterns of behavior. Accordingly, based on the SVM classifier [6][7],this paper proposes a new defense strategy: proactive defense policy for the user behavior patterns.

Support Vector Machine

Behavior analysis is the basis of the virus infected detection system whether it is focusing on the user operational model or malicious operation model. The processing that can be achieved by

pattern recognition techniques. The main idea of this technique is to collect a large number of known types training samples and extracted their characteristic. Then design a classifier to divide the samples into different clusters. Hence, the trained classifier can use to detect the samples of unknown type. So, the SVM-based classifier is introduced to build the virus detector.

Support vector machine is a new machine learning algorithm bases on statistical learning theory which is mainly based on the following idea: the input vector is mapped into a high-dimensional feature pre-selected non-linear mapping space and construct the optimal decision function in this space; the constructed processing of the decision function considering the risk minimization principle, and clever usingkernel function to replace the dot product in original high-dimensional feature space which make complex calculations can be simplified.

Therefore, SVM has stronger generalization ability and is anoutstanding classifier whichcommonlyused in a majority of fields. However, the processing of parameter selection in SVM can be regarded as a nonlinear optimization problem, so the DE algorithm was introduced to optimize SVM in following.

Differential Evolution Algorithm

DE(Differential Evolution) was proposed in 1996 [8]. It is a novel algorithm that based on the idea of Genetic algorithm and it also simulating the evolution processing of natural biologics. The core idea of DA is that generates a temporary individual on account of the different degree in the entire population and then restructures randomly. Many researchers demonstrate that it performance better than GA, PSO and other previous intelligent algorithms. Consider the SVM, how to select the optimal parameters is a difficult problem cause those distribution spaces are not serially. To address this, DE-SVM was proposed. The corresponding flow chart has been shown in Fig.1.

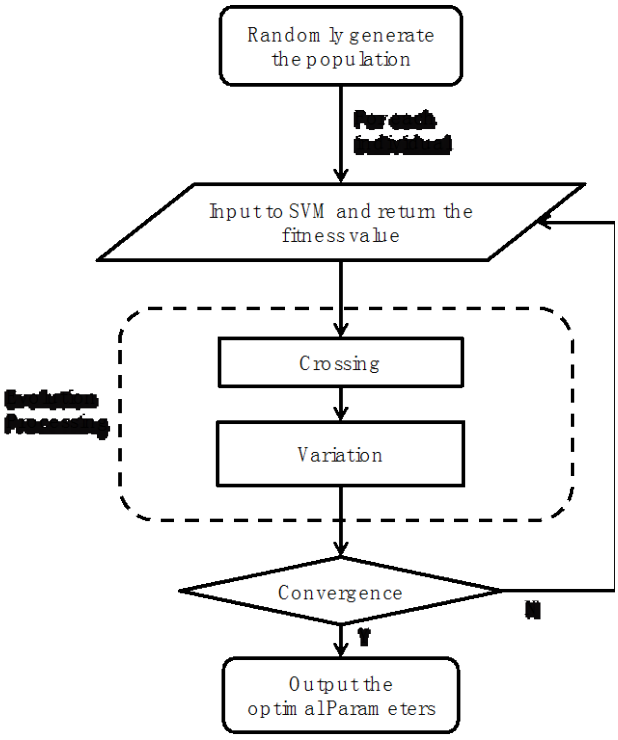


Fig.1. The flowchart of DE-SVM

Experimental Results

In order to better exhibit the proposed algorithm, the data were divided into ten parts and perform the experiments respectively. As we knew about that, BPNN was commonly utilized into pattern recognize. However, cause the lacking of physical theory. BPNN always feedbacks an unstable result. Compare to it, SVM has a more robust generalization. Such as showed in Fig.2, the

comparison results showed the SVM was superior to BPNN when both of them were applied to the virus detection problem.

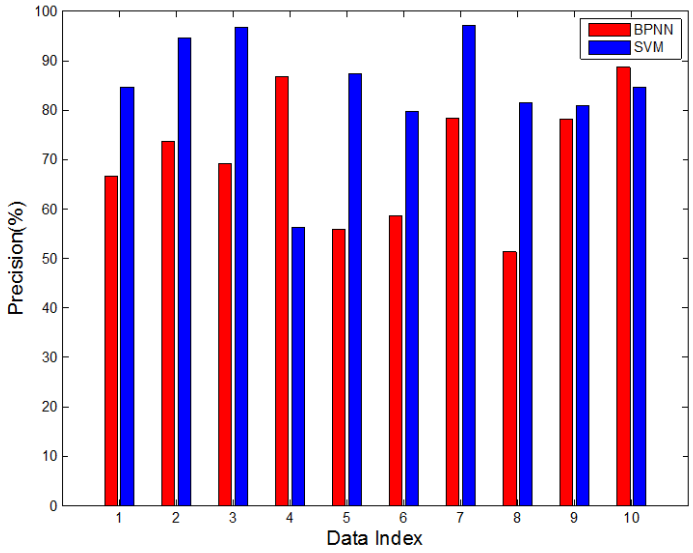
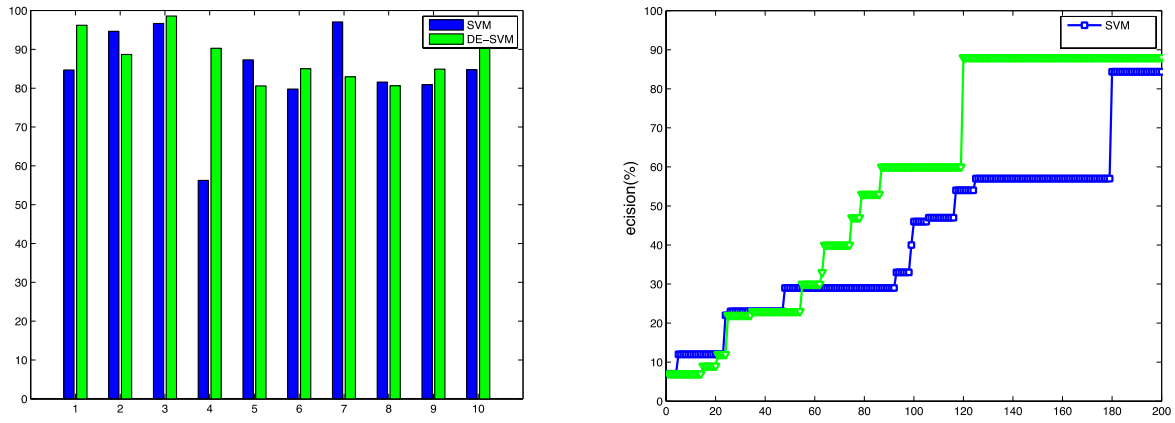


Fig.2. Comparison results between BPNN and SVM, each bar denotes the detection precision of the corresponding group. Obviously, SVM has a higher average precision than BPNN.

Then, the comparison experiments were constructed on the same data as in the Fig.2. It has been seen in Fig.3, the performance of SVM was refined by DE algorithm. In addition, convergence time has been shorted.



(a) detection precision(b) iteration curves

Fig.3. The comparison results from both SVM and DE-SVM.

Another thing to be worth mentioned is that the DE algorithm has superior convergence speed to the other methods as showed in Tab.1.

Detection Models	Running Time
SVM	1.934s
BPNN	3.991s
PSO-SVM [9]	5.241s
GA-SVM [10]	8.487s
DE-SVM	1.354s

Tab.1. Therunning times of different algorithms

Obviously, DE-SVM has the fastest convergence speed. This is very favorable for practice application.For more impressions, we also detailed the detection precisions according to the above models in Fig.4.

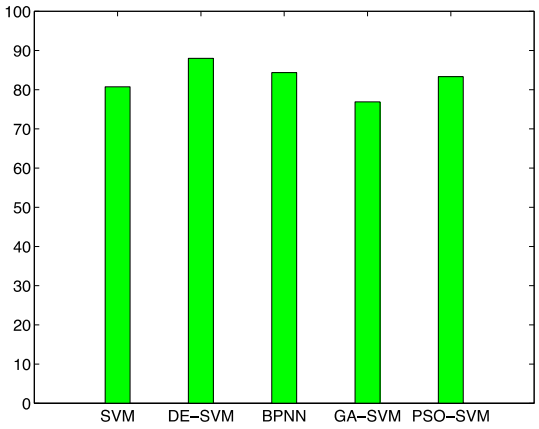


Fig.4. Average precisions of different models performed on the same data.

Conclusion

We had established a novel active defense model by utilizing SVM, and refine the virus detecting method by DE algorithm.By enhancing the SVMutilizing DE, the average precision and coverage speed also have significant improvements. The proposed model can be widely applied in the existing information security models.

References

[1] Manshaei M H, Zhu Q, Alpcan T, et al. Game theory meets network security and privacy[J]. ACM Computing Surveys (CSUR), 2013, 45(3): 25.

[2] Herrmann A, Drube R, Lunt T, et al. Real-time protection of in-vessel components in ASDEX Upgrade[J]. Fusion Engineering and Design, 2011, 86(6): 530-534.

[3] Orjuela J, Deless E F T, Kolade O, et al. A recessive resistance to Rice yellow mottle virus is associated with a rice homolog of the cpr5 gene, a regulator of active defense mechanisms[J]. Molecular Plant-Microbe Interactions, 2013, 26(12): 1455-1463.

[4]Kadam A R. Protecting persistent secondary platform storage against attack from malicious or unauthorized programs: U.S. Patent 8,413,253[P]. 2013-4-2.

[5] Song J, Takakura H, Kwon Y. A generalized feature extraction scheme to detect 0-day attacks via ids alerts[C]. Applications and the Internet, 2008. SAINT 2008. International Symposium on. IEEE, 2008: 55-61.

[6] Metzler S, Kalinina O V. Detection of atypical genes in virus families using a one-class SVM[J]. BMC genomics, 2014, 15(1): 913.

[7] Zhao M, Ge F, Zhang T, et al. AntiMalDroid: an efficient SVM-based malware detection framework for Android[M]. Information Computing and Applications. Springer Berlin Heidelberg, 2011: 158-166.

[8] Price K V. Differential evolution: a fast and simple numerical optimizer[C]. Fuzzy Information Processing Society, 1996. NAFIPS., 1996 Biennial Conference of the North American. IEEE, 1996: 524-527.

[9] Tu C J, Chuang L Y, Chang J Y, et al. Feature selection using PSO-SVM[J]. IAENG International journal of computer science, 2007, 33(1): 111-116.

[10] Wang H, Yu Y, Liu Z. SVM classifier incorporating feature selection using GA for spam detection[M]. Embedded and Ubiquitous Computing–EUC 2005. Springer Berlin Heidelberg, 2005: 1147-1154.