EVOLUTION OF DETECTORS IN NEURAL NETWORK IMMUNE SYSTEM FOR PATTERN RECOGNITION

S. Bezobrazov, V. Golovko

Brest State Technical University
Brest, Belarus
E-mail: bescase@gmail.com, gva@bstu.by

In this paper we present the basic principles of the evolution of detectors in intelligent system for pattern recognition, such as malicious code detection. This system based on integration of both AI methods: artificial neural networks and artificial immune systems. The goal of the evolution is adaptation of detectors to new, unknown malicious code for increasing of quality of detection.

Key words: pattern recognition; artificial immune system; neural networks.

INTRODUCTION

The CIS company [1] reported that during 2009 – 2010 every second organization was attacked. There are four types of cyber attacks detected:

1. Attacks via Internet.
2. Local attacks.
3. Network attacks (detected by Intrusion Detection System).
4. E-mail attacks.

The today's malicious trend is characterized by intensive growth of the first two types of attacks. In 2010 the highest known level (more than 1,9 billion) of these attacks was fixed. Most famous from these attacks are: Mariposa, Bredolab, TDSS, Koobface, Sinowal, Black Energy etc. Each of them infected millions of computers all over the world.

The quality and complexity of malware are constantly enhancing. The striking example of such new cyber threat is Stuxnet [2]. It was discovered in July 2010 but experts are still analyzing the ability and hide functions up till now. Stuxnet was written especially for attacking Supervisory Control And Data Acquisition (SCADA) systems used to control and monitor industrial processes. Several security companies claim that Stuxnet is "a working and fearsome prototype of a cyber-weapon that will lead to the creation of a new arms race in the world" [3].

The users of social networks also undergo a cyber attacks. The main goal of such attacks is a stealing of confidential information and placement of links on infected web resources. Thus worm Koobface [4] is displayed the largest activity, attacked Twitter and sent messages with links to Trojans.

The currently applying methods of information security do not guarantee effective information protection level. The reactive defense [5] based on signature analysis is exact method but is able to detect only already known attacks. The modern proactive [5] defense based on different heuristic methods is characterized by low level of malicious code detection. We need new robust methods to defend against constantly evolvable cyber attacks.

In our previous works [6, 7, 8, 9] we presented and described self-organizing and self-adapting detection system based on integration of both artificial intelligence methods: artificial neural networks and artificial immune systems. Such system consists of population of detectors which scanned data for the purpose of malware detection. We had demonstrated that immune detectors with neural architecture can detect new, unknown attacks and provide better results in comparison with different methods of malware detection. The immune detectors are going through several stages during lifecycle: generation, learning, selection, cloning, mutation and transformation into the memory detectors.

In this paper we explore the adaptability of immune detectors to unknown threat. The adaptability of detector consists in modification of its structure for increasing of detection rate of unknown malware. The modification of the structure of detector is occurs if malicious code was detected. The process of reviewing of the features of new malware is occurs whereupon detectors changed self parameters.

The paper is organized in a following way: Section 2 explains the structure of developed detection system and gives the description of basic working principles. Section 3 presents the detailed mechanism of adaptation of immune detectors to the new detected malware. Results of experiments are discussed in Section 4. Finally, Section 5 concludes this paper.

THE ARCHITECTURE OF INTELLIGENT SECURITY SYSTEM FOR MALWARE DETECTION

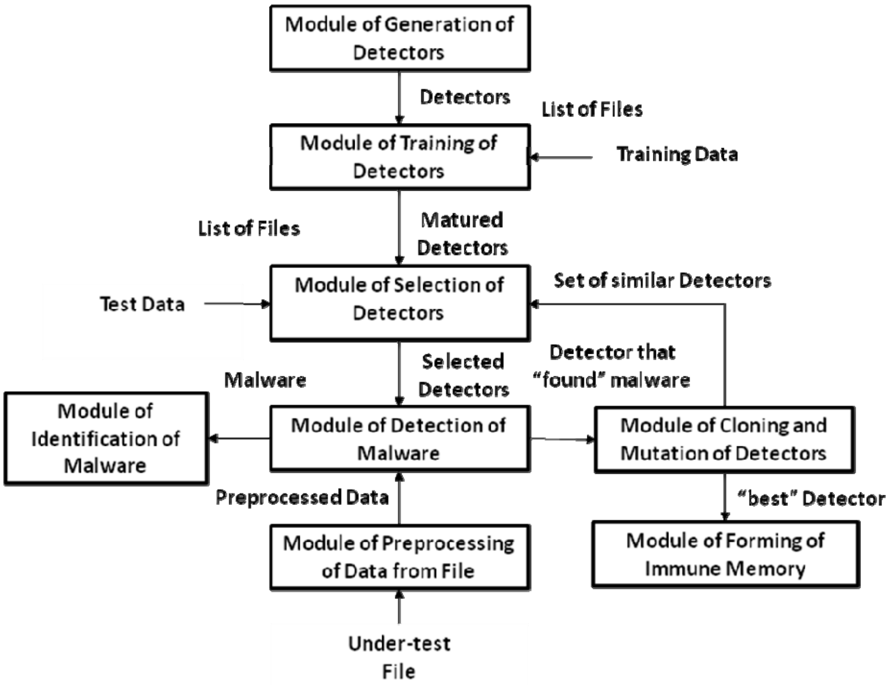
The developed detection system for malicious code detection based on mechanisms of artificial immune system [10] where immune detectors have neural structure [6]. Fig. 1 shows the architecture of the system.

The module of generation of detectors produce pre-detector which should go through several stages before it acquires the ability of correct classifications of files. Every immune detector has the limited lifetime in which it can function in the system. At the end of this period detector is and replaced by new different detector. The lifetime mechanism saves from inefficient detectors and provides the system of new structurally various detectors.

The module of learning performs the function of detectors learning for correct and robust classification of objects and malware detection. The developed learning algorithm is described in detail in [6].

As experiments show, not all randomly generated detectors are able to learn and perform a correct classification. Most of them produce false alarm or can't learn at all. This occurs because of stochastic principle of generating of detectors and creating of learning samples for each detector [6, 8]. The module of selection of detectors performs the task of eliminations those detectors which generate false alarm. For this purpose the test data is formed and every learned "mature" detector undergo is tasked by correctness of presented data classification. As a result only those detectors survive that correctly classifying test data. The module of selection allows decreasing false alarm and increasing defense level.

"Mature" and "selected" detectors "live" in system and scan program objects (files, active processes, streams etc.) for the purpose of malware detection. The set of worked immune detectors forms multi-agent system. Each detector is intelligent agent with own list of task. It selects the target of scanning, makes clones and evolves. If immune detector finds malicious code it emits a signal of malware detection and initiates mechanisms of exploring and adaptation to detected malware and its eliminations.



Module of Generation of Detectors
Detectors
List of Files
Module of Training of Detectors
Training Data
List of Files
Matured Detectors
Module of Selection of Detectors
Test Data
Set of similar Detectors
Selected Detectors
Detector that "found" malware
Malware
Module of Identification of Malware
Module of Detection of Malware
Module of Cloning and Mutation of Detectors
Preprocessed Data
"best" Detector
Module of Preprocessing of Data from File
Module of Forming of Immune Memory
Under-test File

Fig.1. The architecture of the intelligent detection system

The module of cloning and mutation produces some copies of detector which made alarm. These clones are useful in the case when malware has ability to infect executable files and if it gets into the computer it can have enough time to infect some objects. Several immune detectors with identical structure and capable to react on found malware and are able to check all objects in short time and to detect all infected files.

Moreover the module of cloning and mutation allows to security system to explore the detected malware and to adapt to new threat. This is achieved via mutation process. In our system mutation consists in additional learning of detectors when detected malicious code is includes into learning sample. Thus every new clone learns on malware which was detected, as a result the ability of detection of new active malware is increasing. Thus our system is in permanent evolution that allow it adapt to changes of environment and provide all-time high defense level without interference from experts as antivirus base and program code update.

The module of immune memory performs the actions on quick reaction on repeated infection of computer by already known malware. Actually it transforms the "best" clone for detected malware into the memory detector which "live" in the system much longer then "usual" detector. The set of immune detectors form the immune memory and safely defense computer against repeated infections.

EVOLUTION OF IMMUNE DETECTORS

The evolution of detectors is important part of intelligent security system, because it allows to expose new regularities and features of unknown malware and to adapt to them. Thereby the system is evolving and increasing its defenses abilities.

If *i*-th detector detects malware in file, then it activates alarm. In this case cloning and mutation of given detector is performed. As a result the set of clones are generated and each clone is trained by using detected infected file (mutation).

The algorithm of evolution can be divided into following steps:

- Creation of copies (clones) of detector that found malicious code.
- Creation of learning sample based on data from found malware.
- Training of clones.
- Calculation of fitness of clones.

Finally we can get the set of clones D_i, which are aimed to detect given virus

$$D_i = (D_{i1}, D_{i2}, \dots, D_{in}). \qquad (1)$$

The algorithm of creation of training data for cloned detectors differs from the algorithm for "usual" detectors. In the first case, we randomly select some legitimate files and malware and take pieces of its code for creation of training data for detector [6, 7]. The presence of different files of both types, legitimate and malicious, allows detectors during training phase to acquire ability to find differences between classes and to detect unknown malicious code. The goal of mutations is to explore new found malware, to find samples of new malicious techniques and to elaborate robust detectors. Therefore we apply relearning process of clones. The training data for clones consists of data pieces only from found malware. Thus, clones adapt to new threat and provide effective defense from active cyber attack.

The fitness function is used to determine detection quality. As fitness the mean-squared error of between input and output vectors for immune detector can be used [6]:

$$E_i = \frac{1}{2}\sum_{k=1}^{L}\sum_{j=1}^{2}(Z_{ij}^k - l_{ij}^k)^2, \qquad (2)$$

where Z_{ij}^k is *j*-th output unit of *i*-th clone for *k*-th pattern; l_{ij}^k is reference output value for *i*-th clone.

Detector-clone with minimal value of mean-squared error is transformed into memory detector:

$$D_i = M_k, if\ E_i < E_j, \qquad (3)$$

where M_i is memory detector.

EXPERIMENTAL RESULTS

In our previous works [6, 7, 8, 9] we demonstrated that immune detectors allow detecting new unknown malicious code.

For exploring adaptation ability of immune detectors we Next as an example of chosen one detector that detects several malware: *Trojan.INS.gi, Virus.Bee, Virus.Neshta.a and Virus.VB.d* [9]. After detection of *Trojan.INS.gi* detector D3 is acts as material for cloning. For every clone the learning sample is formed which consists of data extracted from the detected malware. As a result the clones are tune in to detected new malware and improve the detection ability.

After cloning clones detects *Trojan.INS.gi* with higher rate than detector *D3*. In addition the clones demonstrate the ability to detect such malware (*Trojan.Bagle.f, Tro-*

jan.INS.bl, Trojan.Ladder.a, Trojan.Small.da) that are stayed undetectable in the case of *D3* scanning.

New learning sample composed of data from *Trojan.INS.gi* allowed clones to improve the ability detect Trojan programs. But on the other part new clones practically lost the ability to detect malware from other classes.

Even so, detected new malware added to learning sample for training new neural network immune detectors that allowed generate new detectors with different structure. These two mechanism cloning and learning sample update allow to increase the detection quality and to adapt to new unknown malware.

CONCLUSION

The ability of immune detectors to evolve by exploring of new malicious material allow to intelligent detection system adapt to new malware and provide effective defense against known and unknown cyber attacks by oneself. Adapted detectors acquire the ability to detect some new malware with higher quality. And new detected malware is adding to learning sample that increase the difference in immune detectors which detecting more unknown malware.

REFERENCES

1. http://cisecurity.org/en-us/?route=default
2. http://www.securelist.com/en/descriptions/15071647/Rootkit.Win32.Stuxnet.a
3. http://www.bbc.co.uk/news/technology-12056594
4. http://www.symantec.com/security_response/writeup.jsp?docid=2008-080315-0217-99
5. http://www.virusbtn.com/vb100/latest_comparative/index
6. *Bezobrazov, S.* Neuronet artificial immune systems for malicious code detection / *S.* Bezobrazov, V. Golovko ICNNAI'2010, P. 147–153, June 2010.
7. *Golovko, V.* Neural Network and Artificial Immune Systems for the detection of Malware and Network Intrusions / V. Golovko, S. Bezobrazov, P. Kachurko, L. Vaitsekhovich. Springer Berlin, Vol. 263. P. 485–513. Januar 2010.
8. *Bezobrazov, S.* Artificial immune systems approach for the malware detection: neural networks applying for immune detectors construction / S. Bezobrazov, V. Golovko. Inernational journal of «Computing», 2008.Vol. 7. P. 44–50,
9. *Bezobrazov, S.* Neural Networks for Artificial Immune Systems: LVQ for Detectors Construction / S. Bezobrazov, V. Golovko. IDAACS'2007. P. 180–184. September 2007.
10. *De Castro, L.* Artificial Immune Systems: A New Computational Intelligence Approach / L. De Castro, J. Timmis. Springer, London, 2002.