RESEARCH THE BEHAVIOR OF ELEMENTS IN ARTIFICIAL IMMUNE SYSTEM FOR INTRUSION DETECTION SYSTEMS IN INFORMATION NETWORKS

M.E. Burlakov, M.N. Osipov

Samara National Research University, Samara, Russia

Abstract. This paper proposes for watching the artificial immune system. The definitions the basic element and the element with the memory like the part of artificial immune system in intrusion detection systems are described. The metric between elements with the limit measure is set. This metric is called affinity and the limit measure is called affinity threshold. The definitions of clone and mutation operations are set. Besides, the behavior between basic elements and elements with the memory on using clone and mutation operations in artificial immune system is researched.

Keywords: artificial immune system, information network, intrusion detection system, clone operation, mutation operation

Citation: Burlakov ME, Osipov MN. Research the behavior of elements in artificial immune system for intrusion detection systems in information networks. CEUR Workshop Proceedings, 2016; 1638: 895-901. DOI: 10.18287/1613-0073-2016-1638-895-901

Introduction

Today there is important task of classifying the data and messages transmitted from the sender to the recipient through different systems in modern information infrastructure (*mail*, *web*, *irq* и т.д.). This problem solved by using either non-adaptive (methods of attack graphs scenarios, methods of analysis of the state systems, expert systems, methods on specifications, signature-based methods, etc.) or adaptive (artificial neural networks, artificial immune algorithms, genetic algorithms, etc.) methods [1-15].

As a part of information system the problem of classification the messages is reduced to task of classifying incoming data to definite class (for example, by relevance, by sender, by content and volume).

Today, in analyzing and classifying messages problem the most actual solution is using the classification data by it content. If we have no any information about data source it's possible to determine the class which this data can be applied with a cer-

tain degree of probability. Further, this data is sent to recipient or blocked in information system.
One of the most important problem of classifying the data blocks and emails is it's distribution to appropriate classes of reliability [16-18]:

1. Reliable (actual, legitimate, and so on) class of data information;
2. Non-reliable (irrelevant, illegitimate, etc.) class of data information.

The **reliable (legitimate) information** is a set of data which doesn't include any threat in terms of availability, integrity and confidentiality for information system. Otherwise, the information is called non-reliable (illegitimate). Any anti-spam system or software and hardware anti-virus protection is classic example of such classification, because the information is classified by its content for two classes: reliable and non-reliable.
As said before, there are a large number of both adaptive and non-adaptive algorithms can classify the data blocks (emails) by its content. In [19] there is proposition which solves the problem by using two-classification artificial immune system.

Two-classification artificial immune system

Two-classification artificial immune system (2CAIS) is an adaptive algorithm with the teacher, which allows to classify the data blocks (emails) for two classes: reliable class of messages and the non-reliable class of messages.
The algorithm of two-classification artificial immune system was produced as an analogue of the biological immune system. The β-element (base element) and a β^m-element (element with memory) are the basic definitions of 2CAIS. This elements are analogs of *B*-lymphocyte and *B*-lymphocyte with the memory in biological systems.
B-lymphocyte and *B*-lymphocyte with memory, from a biological point of view, are "security flag". If this flag is changed the signal about it is created and the lymphocytes are generated by the body. The lymphocytes are deal with the (antigens). If the lymphocyte destroys a particular threat successfully or responds it with the higher probability, immune system transfers this *B*-lymphocyte to *B*-lymphocyte with the memory. Immune system transforms *B*-lymphocytes to *B*-lymphocyte with the memory using the presence of a weighting parameter. The more effective lymphocyte makes its functions the higher its "weight" and vice versa. If the lymphocyte does not provide effective kind of threat detection the immune system reduces the value of weight parameter until it reaches the point where it can be removed from the system finally.
Similarly, *B*-lymphocyte is a β-element and *B*-lymphocyte with the memory is β^m-element in two-classification artificial immune system.
Also, similarly, the *age* (*power*, *weight*) is the main parameter of β-element. The age is a non-zero value, which characterizes the element weight. It fulfills the "death" condition of elements. If its value becomes equal zero the element is removed from the system. Another words, the age parameter shows the efficiency of elements. Also there are **mutation and cloning operations** in 2CAIS.
The mutation operation is a process of changing the β-element structure randomly. The age of the β-element is finite and the process of mutation affects only a certain

number of entities. The measure of mutation is set on initializing stage 2CAIS algorithm. The mutating operation (*Mutating*) is equivalent to write:

$$Mutating(\beta) = \sum \beta_i \bigcap \alpha_j, \alpha = \sum \alpha_j, \ |\alpha| >> |\beta|, i = \lfloor m \cdot j \rfloor, \quad m \in [0,1] \tag{1}$$

where crossing sign is equivalent to the replacement operation;

α – set of entities from the β-element;

m – mutation rate.

The cloning operation (*Cloning* (β)) is used for creating new β-elements in artificial immune system. This operation is applied when the mutation operation is finished. The cloning copies new mutate β-elements to current β-element set. The cloning coefficient is an important parameter of cloning operation. It shows how much elements will be created relative to the original.

Thus, the mutation operation creates new β-elements which potential to identify new threats. The cloning operation clones these elements to system.

These operations provide the variability between β^m-elements and a simple β-elements. The variability is the process which considers β^m-element like the element with the memory. The β^m-element is the most well-established "cast" the threat that it is able to effectively detect. Simple β-element is produced by a combination of a random selection of words from the emails or datasets with the applying of cloning and mutation operations. The β^m-element is derived from iterative measure of effectiveness β-element working in a set of emails.

The distance (metrics) between the β-elements is calculated by using the **affinity** definition.

In 2CAIS the **affinity of the elements** defined as the ratio between the number of common entities, which are composed of these elements to norm. The norm of two elements is the minimum number of entities that make up each of the elements. Therefore, the affinity (*Affinity* or *α*) may be expressed as the following equation:

$$Affinity(\beta^1, \beta^2) = \alpha(\beta^1, \beta^2) = \frac{Count(\sum \beta_i^1 \bigcap \sum \beta_j^2)}{\min(|\beta_i^1|, |\beta_j^2|)}, i \neq j \tag{2}$$

where *i, j* – the order of the entities in the element;

Count – a function of the number of intersections of the elements;

$|\beta_j|$ – element capacity (the number of entities in the elements).

If the value of the affinity between the antigen (threat or potential message from the class of non-reliable information) and any β-element is above some threshold value (**threshold affinity**), then it means that the β-element recognizes the threat. After that, the system mark this data block (email) as a non-reliable, otherwise the data block is marked as reliable.

On the one hand, the affinity allows to detect the threat and thus carry out the process of the primary classification. On the other hand, there is a question to research the relationship between the processes of creating of β-elements and β^m-elements. In other words, there is a task that requires to explore the relationship between basic (simple) β-elements and β^m-elements with the memory which have the affinity metrics cloning

and mutation operations through the problem of classification data block (emails) into two classes: reliable information class and non-reliable information class of messages. To solve the task of finding the relationship between β-elements and β^m-elements prove Proposition 1.

Proposition 1.

For each $\varepsilon > 0$ exists $\lambda \in [0,1]$ such that $|\beta| = \lambda|\beta^m|$.

Proof

Let introduce some notation.

Consider the initial set of messages (emails, data blocks) *S.* β-elements initialized from *S*, $|S| = s$, $s>0$ and $s \in N$. Let k – the age if β-element, and n – the threshold when $\beta \to \beta^m$ (affinity threshold), and $r = n-k$, $r>0$, the difference between them. There are $\bigcup_{i=1}^{s} \beta(k)$ – elements (words in email) when the system is initialized.

Let α – the value of the affinity function *Affinity()* for two elements. P_α – probability of appearing two β-elements which have the affinity distance greater than or equal to α, p_i – its corresponding value. Let l the cloning coefficient in function *Cloning(β)* (hereinafter briefly denoted by *C(β)* function). Let m the mutation coefficient which corresponds the probability characteristics of new β-elements in *Mutating(β)* (hereinafter briefly denoted as a function of *M*(β)).

General algorithm of producing β-element and β^m-elements consists of several steps:

Step 1. Initialization β-element set from the *S* set;

Step 2. Application the mutation and cloning functions to β-element set;

Step 3. Calculate the probability P_α for getting β-element set;

Step 4. Go to Step 2.

Consider the process iteratively. In Steps 1 and 2 the set of β-elements applied the cloning and mutation operation:

$$C(M(\bigcup_{i=1}^{s} \beta(k))) \tag{3}$$

In Step 3, for each β-element from β-element calculated the probability P_α. After that, there is new set of β-elements with the different age due to affinity threshold. Next, go to Step 2. Finally, the process consists of a finite set of iterations.

Iteration 1.

$$M_1 = C(M(\bigcup_{i=1}^{s} \beta(k))) \mid P_\alpha = \bigcup_{i_1=1}^{b_{11} \le s_1} \beta_{i_1}(k-1) \bigcup_{i_2=1}^{b_{12} \le s_1 - b_{11}} \beta_{i_2}(k+1) \tag{4}$$

In the first iteration the set M_1 is obtained. M_1 consists of two β-element subsets with the age equals to k-1 (if the distance is less affine α) and the age equals k+1 (if affine distance greater than or equal to α), where $s_1 \ge s$ due to cloning operation. The dimension of M_1 is finite and equals (due to the finite of *S*):

$$M_1 = p_1 l_1 m_1 \mid \beta(k-1) \parallel \beta(k+1) \mid \tag{5}$$

where p_1 – the probability P_α,

l_1 – cloning coefficient in the first step,

m_1 – mutation coefficient in the first step,

$\beta(k - 1)$ – the number of β-elements which have age is equal to k -1,

$\beta(k + 1)$ – the number of β-elements which have age is equal to k+1.

Iteration 2 is calculated based on the re-use of cloning and mutation operations taking the results obtained in the previous iteration, that is:

$$M_i = C(M(M_{i-1})) \mid P_\alpha = \\ C(M(C(M(M_{i-2})))) \mid P_\alpha = C(M...(M_1)...) \mid P_\alpha \quad (6)$$

Thus, the result of iteration 2 is as follows:

$$M_2 = C(M(M_1)) \mid P_\alpha = \bigcup_{i_1=1}^{b_{21} \le b_{11}} \beta_{i_1}(k-2) \times \\ \times \bigcup_{i_2=1}^{b_{22} \le b_{11} - b_{21}} \beta_{i_2}(k) \bigcup_{i_3=1}^{b_{23} \le b_{12}} \beta_{i_3}(k) \bigcup_{i_4=1}^{b_{24} \le b_{12} - b_{23}} \beta_{i_4}(k+1) \quad (7)$$

and a dimension of M_2 is finite and equals to:

$$\mid M_2 \mid = p_2 l_2 m_2 \mid \beta(k-2) \parallel \beta(k) \parallel \beta(k+1) \mid \quad (8)$$

where the parameters are similar like in Iteration 1.
For the r-th iteration, the set of M_r becomes:

$$M_2 = L(M(M_{r-1})) \mid P_\alpha = \bigcup_{i_1=1}^{b_{r1} \le b_{(r-1)1}} \beta_{i_1}(k-r) \times \\ \times \bigcup_{i_2=1}^{b_{r2} \le b_{(r-1)1} - b_{r1}} \beta_{i_2}(k-r+1) ... \bigcup_{i_2 r=1}^{b_{r2} r \le b_{(r-1)1} - \sum_{j=1}^{r-1} b_{rj}} \beta_{i_2 r}(k+r) \quad (9)$$

On this stage the β^m-elements are appeared because there are β-elements which have age equals $n = k+r$. The value of this age equals the threshold when $\beta \to \beta^m$.

The dimension M_r is finite and equals:

$$\mid M_r \mid = p_r l_r m_r \mid \beta(k-r) \parallel \beta(k-r+1) \mid ... \\ \mid \beta(k+r-1) \parallel \beta(k+r) \mid = p_r l_r m_r \mid \beta \parallel \beta^m \mid \quad (10)$$

where $|\beta|$ – number of all β-elements;
$|\beta^m|$ – number of all β^m-elements.

The relationship between the count of β-elements and β^m-elements is a value (coefficient) depending on cloning coefficient, mutation coefficient and affinity distance greater than or equal to a predetermined value. Thus, the proposition that for each $\varepsilon>0$ exists $\lambda \in [0,1]$ such that $|\beta| = \lambda|\beta^m|$ is proofed.

Now, we need to establish the relation between the number of simple β- elements and β^m-elements with the memory from the sets it's created. Another words, there is the task to determine the dependence of count β-elements and β^m-elements for the two finite sets which have nesting ratio. To do this, let state Proposition 2.

Proposition 2.
For any two finite sets S_1 и S_2, where $S_1 \subseteq S_2$, $\mid S_1 \mid \le \mid S_2 \mid$, $|\beta_1| \le |\beta_2|$ and $|\beta^m{}_1| \le |\beta^m{}_2|$.

Proof
The proof is reduced to a repetition of the iterations of the Proposition 1, but this time for two sets S_1 and S_2 with a comparison of the results.
Thus, there is a direct dependency between simple β-elements from β^m-elements which allows to estimate the amount of generated β-elements and β^m-elements. Besides, if we have finite sets we can predict the number of β-elements and β^m-elements. It helps us to design the information systems for classifying.

Conclusion

Thus, there is a dependency between the simple β-elements from β^m-elements within a finite set of emails or data blocks. It can help to predict the amount of memory allocated for the creation and storage the β-elements from β^m-elements in 2CAIS. The information systems which can be built using this approach can classify emails or data blocks on two classes: reliable class and non-reliable class. It can be very necessary for creating the intrusion detection systems based on its content.

References

1. Vasylyev VI. Intelligent Information Security Systems. Moscow, Mashinostroenie Publ., 2012; 20-22. [in Russian]
2. Vacca JR. Computer and Information Security Handbook. Newnes, 2012; 330-335.
3. Nunes L. Artificial Immune Systems: A New Computational Intelligence Approach. Springer Science & Business Media Publ., 2002; 2-4.
4. Haikin S. Neural networks. Moscow. Vilyams Publ., 2008; 32-34. [in Russian]
5. Abe S. Support Vector Machines for Pattern Classification. Springer Science & Business Media Publ., 2005; 39-40.
6. Kollias S. Artificial Neural Networks. Springer Science & Business Media Publ., 2006; 161-162.
7. Artificial Immune Systems and Their Applications, Edited by D. Dasgupta Springer Verlag Publ., 1999; 306 p.
8. Tarakanov AO. Immunocomputing: principles and applications. Springer Verlag, New York Publ., 2003; 193 p.
9. Vacca JR. Computer and Information Security Handbook. Newnes, 2012; 330-335.
10. Borger E. The Abstract State Machines Method for High-Level System Design and Analysis. Dipartimento di Informatica, Universita di Pisa Publ., 2007; 30-35.
11. Shim JK. Information Systems and Technology for the Noninformation Systems Executive. CRC Press Publ., 2000; 230-235.
12. Lunt TF. A real-time intrusion-detection expert system (IDES). Final Technical Report, 1992; 10-13.
13. Burlakov ME. The method of filtering incoming traffic on the basis of a two-layer recurrent neural network. Polzunovsky vestnik, 2012; 3(2): 215-219. [in Russian]
14. Burlakov ME., Osipov MN. Security audit of the local area network using a dynamic system neurons reacting to the sequence. Information counteraction to threats of terrorism, 2013; 20: 166-170. [in Russian]
15. Burlakov ME. On some optimization models of the artificial neural networks by genetic algorithms. Proceeding of the International Scientific Conference (PIT-2015), Samara: Samara Scientific Center of RAS Publ., 2014; 99-105. [in Russian]

16. Delvin D, O'Sullivan B. Satisfiability as a Classification Problem. University College Cork Publ. URL: http://www.cs.ucc.ie/~osullb/pubs/classification.pdf.
17. Fernandez-Delgado M, Cernadas E, Barro S. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems. University of Santiago de Compostela Publ. URL: http://jmlr.csail.mit.edu/papers/volume15/delgado14a/delgado14a.pdf.
18. Schapire R. Machine Learning Algorithms for Classification. Princeton University Publ. URL: http://www.cs.princeton. edu/~schapire/talks/picasso-minicourse.pdf.
19. Burlakov ME. Two-classification artificial immune system. Vestnik Samara State University, 2014; 7(118): 207-221. [in Russian]