Network intrusion detection by the coevolutionary immune algorithm of artificial immune systems with clonal selection

T Salamatova, V Zhukov
Reshetnev Siberian State Aerospace University, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation
E-mail: salamatova_ta@mail.ru

Abstract. The paper presents the application of the artificial immune systems apparatus as a heuristic method of network intrusion detection for algorithmic provision of intrusion detection systems. The coevolutionary immune algorithm of artificial immune systems with clonal selection was elaborated. In testing different datasets the empirical results of evaluation of the algorithm effectiveness were achieved. To identify the degree of efficiency the algorithm was compared with analogs. The fundamental rules based of solutions generated by this algorithm are described in the article.

1. Introduction
Nowadays the qualitative and quantitative characteristics of network attacks on the informational objects are constantly changing. Under such conditions the role of intrusion detection systems (IDS) for information security is steadily becoming more and more important. IDS are becoming an integral part of the security infrastructure in public and commercial organizations (figure 1). In some cases applying IDS is a must, and it is governed by the requirements of legal acts and methodological documents.

![Figure 1. External Challenges Faced 2015.](image)

Comparative analysis of IDS showed that most of the Russian IDS use only signature methods for data analysis. The ability to update the fundamental rules database is available only for the developers. Users must buy such updates, they do not have the possibility to create and automatically
generate their own fundamental rules. IDS based only on signature methods do not allow detecting unknown threats of information security («zero day» attacks). As practice shows such IDS do not provide the required level of security for IS. This fact increases information risks and, as a consequence, financial losses. A possible way out of this situation is using adaptive heuristic intrusion detection algorithms in addition to IDS. It will allow not only detecting «zero day» attacks, but also automating a local content (update) of a fundamental rules base for information systems (IS). A heuristic algorithm built on the basis of computer models of artificial immune systems (which is based on the clonal selection methods and the theory of acquired immunity) is considered in this paper.

The paper consists of several sections. The 2nd section of this article is devoted to artificial immune systems. The modified algorithm of artificial immune systems (AIS) with clonal selection is discussed in Section 3. Section 4 evaluates its effectiveness (test and real data were evaluated). Section 5 is devoted to comparative analysis with similar algorithms according to the degree of their efficiency. Section 6 shows the use of generated information structures in the fundamental rules of intrusion detection systems. Finally, section 7 reflects the main results and suggestions for further research.

2. Artificial immune system
AIS is a complex adaptive structure which simulates some aspects of the immune system functioning in protecting organisms against external microorganisms (antigens (Greek ἄντι – against, γενεά – generating, born)) by protective cells of the immune system (antibodies (Greek ἄντι – against) or detectors). A characteristic which describes the strength of antibodies and antigens interaction is defined as the affinity (Latin affinis – related). Presently, such computational models of immune systems as clonal selection algorithms [2, 3], negative selection [4] and immune network algorithms are of interest [5, 6].

The application of clonal selection algorithm based on the methods of clonal selection and the theory of acquired immunity is inspected in the paper. Clonal-selection theory was formulated in 1957 by Frank Macfarlane Burnet in Australia and D. Talmagein in the USA independently of each other. This theory explains how the immune system resists alien macromolecules (antigens). Each individual’s antibody-producing cell system contains all the information necessary for the synthesis of any antibody (detector) before meeting the antigen. The antigen does not give information to these cells; it simply selects cells which synthesize antibodies corresponding to this antigen, and encourages them to multiply and generate such antibodies. Cells synthesizing such type of antibodies belong to one clone which consists of all the descendants of the parent cell (this parent cells acquired the genetic ability to respond to the antigen in a random process). This clone is numerically small before the antigen appearance. The presence of the antigen stimulates clone reproduction; it is able to synthesize the corresponding antibodies. The better this antigen will be recognized (affinity), the greater amount of offspring (clones) will be generated.

The heuristic algorithm built on the basis of synthesis of AIS clonal selection algorithm as well as evolutionary and coevolutionary strategies are presented in this work.

3. Modification algorithm of artificial immune systems with clonal selection
During the development and research of clonal selection algorithm AIS [7] it was determined that the algorithm can detect deliberate changes in the monitored data, stabilize the average number of errors of type I and II, and make them dependent on the resources allocated to the algorithm. However, it became clear that in the process of results averaging the error rate of the first and second type was quite high. The problem of forming a representative set of high affinity detectors has been determined. It became necessary to introduce additional mechanisms for solving the problem. It was proposed to replace the mechanism of detector reproduction and mutation (which was the standard for clonal selection algorithm) with the external optimization procedure. This procedure is based on the application of evolutionary algorithms strategy [8]. The results of experiments with evolutionary immune clonal selection algorithm showed that the proposed modifications reduced the number of
type I errors and nullified type II errors. However, the application of evolutionary strategies resulted in the problem of identification of its optimal settings, such as selection, recombination and mutation.

To solve the problem of automated index selection and setting the immune evolutionary algorithm parameters, the use of coevolutionary strategy was proposed (coevolution: Latin. co – compatibility, consistency, volutio – deployment). Coevolutionary immune algorithm with clonal selection AIS (coevolutionary algorithm) presents some independent, self-acting evolutionary immune algorithms with clonal selection AIS. These algorithms have different settings in competition for granted (limited) computing resource. The generalized scheme of coevolutionary algorithm is presented in figure 2.

![Figure 2. Flowchart of generation of detectors by coevolutionary immune algorithm with clonal selection AIS.](image)

Thanks to the coevolutionary strategy the necessity of a parameters random selection and / or the need of exhaustive combinations search during evolutionary immune algorithm separate testing have been obviated.

4. Experiments and Results

Empirical results of evaluation of coevolutionary algorithm effectiveness (on test data) were introduced [7, 8]. These results indicate that the developed algorithm can detect real data intentional changes (malicious traffic patterns). Also, the algorithm makes it possible to create information structures for automatic construction of fundamental rules (signatures) in the specified format. Effectiveness evaluation was conducted on a set of test data from the public database of network traffic KDD Cup 1999 (database «KDD’99») [10], collected by the Information and Computer Science University of California. The number of detected antigens in the test multiplicity (%), the number of the first and second type errors (%), the ratio of average number of the first and second type errors to the number of detectors and the ratio of abnormal events (antigens) database size to the size of the created detector database were used as evaluation criteria of evolutionary algorithm efficiency.
The next logical step is to evaluate the developed algorithm efficiency in the network traffic «real data» containing or not containing malicious content. Network analyzers (e.g. CommView, Wireshark) using a common format Packet Capture (pcap) as a storage format of the intercepted traffic are typically applied to create the dumps of network traffic. Multiplicity of data for research (patterns of network traffic in pcap format from international cyber-training Locked Shields [11] and DEF CON competition [12]) was formed as a result of open source studying. Figure 3 shows some part of effectiveness evaluation results of the coevolutionary algorithm for antigens unit \(|Z|=10\) when tested on real samples of malicious and the «normal» traffic pcap format (\(REAL\) column). Previously reached on the test base «KDD’99» results are compared (\(TEST\) column). \(D_{best}\) is multiplicity of detectors recorded into the detectors database. \(A\) is a set of antigens (abnormal events). \(E\) is test multiplicity. The results of the experiments averaged over multiple startups show that the percent of detection of malicious samples in the test data and the level of Type II error (100% and 0% respectively) remained the same. The level of the type I error has increased by 1% on the average. The average number of type I errors per one detector remained at the same level as that obtained when tested on the test database «KDD’99».

![Figure 3. The percentage ratio of the amount of detected antigens and the average number of type I errors by varying the set of test data \(E\) and \(|D_{best}|=|A|\cdot15\%\).](image)

Figure 4 shows the size of generated detectors database (Bank of detectors) for the given size of abnormal events database (Bank of antigens) and normal events database (Bank of normal events).

![Figure 4. The size of obtained detectors database at specified test databases of normal and abnormal events.](image)
Thus, the ratio of the abnormal events database to detectors database is about 3:1.

5. Comparison of the results of coevolutionary algorithm research with the other results.

To evaluate the effectiveness of the coevolutionary algorithm developed by the authors it was decided to conduct a comparative analysis of the results of the algorithm and the results of other algorithms (running on the test set «KDD’99»base).

Shirazi H. M. with colleagues proposed an IDS model where collected data analysis is carried out by the algorithm based on Memetic algorithm and Bayesian networks [13]. Also, the authors additionally solve the classification of attacks problem in the research. The model was tested by the «KDD’99» database. Detection Rate (DR) and False Acceptance Rate (FAR) were used as the criteria of effectiveness calculated as follows:

\[
DR = \frac{TP}{TP + FN} \cdot 100\% \quad \text{and} \quad FAR = \frac{FP}{FP + TN} \cdot 100\%,
\]

where \( TP \) (True Positive) is the situation when a signature is fired properly as soon as an attack is detected and an alarm is generated; \( FP \) (False Positive) is the situation when normal traffic causes the signature to raise an alarm; \( TN \) (True Negative) is the situation when normal traffic does not cause the signature to raise an alarm; \( FN \) (False Negative) is the situation when a signature is not fired when an attack is detected.

Figure 5 presents a comparative analysis of the coevolutionary algorithm with the results of research of Shirazi H. M. and other results of scientific researches and works solving a similar problem of intrusion detection mentioned by Shirazi H. M.

![Figure 5. Comparison of the research results of Shirazi H. M. with the other results and the coevolutionary algorithm.](image)

6. The format of the decision rule

The detectors generated by the research coevolutionary algorithm can be used for creation of the IDS decision rules and, in the future, in some other means and systems of information protection (e.g., antivirus software). A lot of modern IDS support the «Snort» signature format of Cisco Company [14], so it is possible to specify the location of the generated coevolutionary algorithm detector as the example of its constructing.
«Snort» signatures (the decision rules) consist of two parts: a rule header and a body (parameters) of a rule. A rule header contains a description of the action, a data transfer protocol, IP-address, a network mask, a source and destination ports. Rule parameters store a warning message and inform what part of the detected packet should be processed in case of rules activation. There are four categories of rule parameters in the modern «Snort»: general, payload, non-payload and post-detection. During the analysis of the interaction of the coevolutionary algorithm detectors and the format of «Snort» signatures it was found that the generated detectors can replace the last three parameters in «Snort» signature structure (figure 6).

![Figure 4. The detector location in the IDS «Snort» signature format.](image)

For example, there is a signature (figure 7):

```
alert tcp any any <> any 80 (msg:"MD5 Alert"; protected_content:"293C9EA246FF9985DC6F62A650F78986"; hash:md5; offset:0; length:4;).
```

In this case, there are two types of parameters involved in the body of the rule: general (msg) and payload (protected_content, hash, offset, length). The msg option sets the rules alert and log a message which is linked to the packages in the dump or is displayed in triggered alert. Keyword protected_content as well as content search the matches in the lines of certain information bits but the search is performed in hashed packets whose content is compared with a predetermined value in the rule. Keyword hash is used to specify the type of hash (MD5, SHA256 и SHA512). Keyword offset allows specifying a search range set value (value from -65535 to 65535). Keyword length is used to indicate the initial length of hashed content specified in the rule (value from 0 to 65536).

```
A signature: 
alert tcp any any <> any 80 (msg:"MD5 Alert";
protected_content:"293C9EA246FF9985DC6F62A650F78986"; hash:md5; offset:0; length:4;)
```

![Figure 5. The synthesis submission the IDS «Snort» signature and the generated detectors.](image)

The analysis of the individual components of the signature parameters shows that the rule demonstrates text search «HTTP» in packages of length of 4 bytes at zero bias. The detection of the text will activate «MD5 Alert» warning.
In using the information structure of the detector generated by the training coevolutionary algorithm previously the signature will acquire the following form:

```
alerttcp any any<> any 80 (msg: «MD5 Alert»; 0111110011011010101101101000110100111000111011010010111001110010011001001001111011110011101001100100111001100111010001001111011<;)
```

where the last value reflects the affinity value whose exceeding is the rule activation condition (in this case "<" is 60 encoded value in the Unicode format). It isn’t necessary to set parameters of payload type using this signature.

Today it is possible to point out the foreign solutions IDS such as «Suricata», «HPE TippingPoint Next Generation Intrusion Prevention Systems», «The Huawei Network Intelligent Protection system», «IBM Security Network IPS», «Cisco IPS», «McAfee Network Security Platform v.8.x» and Russian solutions such as «ViPNet IDS», «User Gate Proxy&Firewall 6.0 VPNGOST», «Dionis-NX» of Factor-TC company, «Outpost» and «Attack Detection «Continent» which support the «Snort» signature format (import and/or export of decision rules).

7. Conclusion

AIS were selected among the existing and rapidly developing methods for the development of a model-algorithmic support system of the heuristic detection methods for IDS. AIS have such important properties as adaptability, self-regulation and self-learning, due to which IDS allow making up the base of decision rules («learning mode») automatically. These data bases will be periodically updated as long as experience is accumulated («operation mode») that in the end could help to reduce the cost of creating the bases of IDS decision rules and updating them.

In general, coevolutionary algorithm proved its efficiency both in the test and "real" data sets as well as the problem of the automatic creation of information structures for the construction of rules to build the IDS decision rules base. However it is necessary to continue research of the coevolutionary algorithm aimed at reducing the errors level of type I (improve the quality of the decisions) and optimizing the algorithm for the rate of its convergence and resource. Further, the authors plan to submit comparative analysis of the research results with the results of other algorithms on the «real» data set and to develop recommendations for improving the algorithm.

References

[1] Annual Security Report Available at: http://www.cisco.com/c/m/en_us/offers/sc04/2016-annual-security-report/index.html
[2] Burnet F 1957 Australian Journal of Science 20 pp 67-69
[3] Talmage D 1957 Annual Review of Medicine 8 pp 239-256
[4] Forrest S, Perelson A 1994 Self-nonself discrimination in a computer Proceedings of the IEEE Symposium on Research in Security and Privacy pp 202-212
[5] Jerne N 1974 Towards a network theory of the immune system Annales d’Immunologie 125 pp 373-389
[6] Kim J and Bentley P 2001 The Artificial Immune System for Network Intrusion Detection: An Investigation of Clonal Selection with Negative Selection Operator Proceedings of the Congress on Evolutionary Computation pp 1244-1252
[7] Zhukov V and Salamatova T 2014 The detection of network intrusion by evolutionary immune algorithm with clonal selection Vestnik SibGAU 4 (56) pp 41-47
[8] Zhukov V and Salamatova T 2014 Detection of information security incidents modified algorithm of artificial immune system with clonal selection In the World of Scientific Discoveries 6.1 (54) pp 497-517
[9] Salamatova T 2015 The determination of competitive search strategy in the coevolutionary immune algorithm with clonal selection Materials XIX International scientific conference Reshetnev reading 2 pp 296-298
[10] KDD Cp 99 Intrusion detection data set Available at: http://kdd.ics.uci.edu/
[11] NATO Cooperative Cyber Defence Centre of Excellence Available at: https://ccdcoe.org/index.html

[12] DEF CON Communications Available at: https://defcon.org/

[13] Shirazi H and Namadchian A 2012 Combined anomaly base intrusion detection using memetic algorithm and Bayesian networks. International journal of machine learning and computing 2 (5) pp 706-710

[14] Snort. Available at: https://snort.org/