Anomaly detection in WSN using Immune Inspired Algorithms

Delona C Johny
Dept. of Information Technology
Govt. Engineering College Bartonhill
Trivandrum, Kerala, India

Anju J S
Dept. of Information Technology
Govt. Engineering College Bartonhill
Trivandrum, Kerala, India

Abstract—Wireless Sensor Networks (WSNs) have been widely considered as one of the most important technologies for the twenty - first century. A WSN typically consists of a large number of low - cost, low - power, and multifunctional sensor nodes that are deployed in a region of interest.These sensors send the data to the central location.But some anomalies interrupt the normal flow of data.For solving these problems we used some bio-inspired mechanisms such as Negative Selection Algorithm(NSA) and Clonal Selection Algorithm(CSA).To implement CSA and NSA,first generate a detector set which containing only anomalous packets.These detector set are compared with test data and anomalous packet are identified.Anomalous data packets are used for further processing to identify specific anomalies. In this way, the number of wormholes, packets delayed, and packets dropped are calculated and identified. Simulations are performed on a large dataset such as wireless sensor dataset and the results show high accuracy of the proposed algorithm in detecting anomalies.

Keywords—Biological Immune System,Artificial Immune System,Negative Selection Algorithm,Clonal Selection Algorithm.

I. Introduction

The Biological Immune System (BIS),shown in figure 1, an integral part of the vertebrate immunity, is a dynamic, powerful, intelligent, and interconnection of different components of the body, working in totality to fight, defend, and prevent pathogenic organisms entrance into the body. BIS functions are protection from foreign invaders, and maintaining homeostasis.A role of the immune system is to protect our bodies from infectious agents such as viruses, bacteria, fungi and other parasites.To detect and eliminate pathogens efficiently,the immune system possesses a multi layered protection,detection and elimination architecture.

There are two basic types of immunity, innate and adaptive . The static system that identifies and eliminates definite harmful organisms is known as innate immune system.The system that remembers unknown foreign cells and react with them is known as adaptive immune system. The adaptive immune system is a combination of atom cells spread all over the body.There are two type of lymphocytes are present in the cells , that is T-cells and B-cells.These cells recognize and destroy specific substances which are entered into our body. Any substance that is capable of generating such a response from the lymphocytes is called an antigen or immunogen. A primary response is produced when the human immune system encounters an antigen for the first time. Large numbers of antibodies are created by the immune system in response to the antigen. These antibodies eliminates the antigen from human body.When the same antigen is encountered again after a period of days ,secondary response is produced by the immune system . Secondary immune response is specific to the antigen that first initiated the immune response and causes a very rapid growth in the quantity of B cells and antibodies.A faster response is attributed to memory cells remaining in the immune system, so that when similar antigen encountered, a new immunity does not need to be built up, it is already occur. AIS is inspired by the processes and principles of the BIS and takes advantage of its characteristics like memory and learning for solving various kinds of problems. AIS abstracts the structure and functions of BIS into computational systems. AIS do not create pattern for normal data but instead produce anomalous patterns by using normal data. These anomalous patterns are known as nonself. Hence, they perform only anomaly-based intrusion detection. Patterns that match with the nonself-patterns will be declared as anomalies.

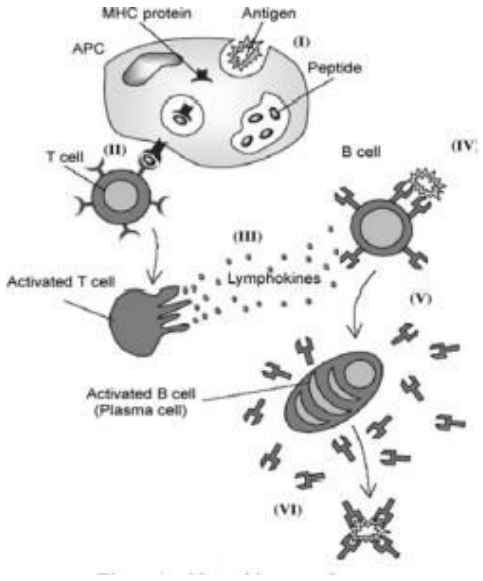

Fig. 1: Biological Immune System [1]

The Biological Immune System (BIS) is of great interest to computer scientists, because it provides a unique and fascinating computational paradigm for solving complex problems. Inspired by BIS, artificial immune system (AIS) has become a vibrant and active research area . Currently, major types of AIS methods include Negative Selection Algorithm (NSA), Clonal Selection Algorithm[5].

Wireless Sensor Networks (WSNs) consist of a set of distributed wireless devices known as sensors. Sensors are used to examine and check physical conditions and pass their data to a central location [2]. It is important to detect whether the data transferred from source nodes reach the gateway properly without any interruption. Sensors are vulnerable to attacks and their security is highly important as they communicate very sensitive data. There are many interruptions, also known as anomalies, which disrupt the normal flow of the sensor data. These anomalies disturb the normal network flow in many ways including delayed packets, packets destroyed, and wormhole attacks. Therefore, it is highly desirable to detect such anomalies that disrupt the normal flow of the data in order to make the sensor network communication more reliable and consistent.For the solution we use bio-inspired algorithms.To implement Negative Selection Algorithm and Clonal Selection Algorithm,we use a large dataset.First generate a detector set and anomalous packet are classified as packet dropping,packet delaying and worm holes.

This paper introduces an anomaly detection techniques using immune inspired algorithms. Section II includes the related works based on anomaly detection .An Artificial immune based anomaly detection techniques are described in the next section. That include CSA and NSA. Section IV discusses the experiments and results. Followed by conclusion in Section V.

II. RELATED WORKS

Kruegel et al. [4] proposed a multisensory fusion approach where the outputs of different IDS sensors were aggregated to produce a single alarm. This approach is based on the assumption that any anomaly detection technique cannot classify a set of events as an intrusion with sufficient confidence. Although using Bayesian networks for intrusion detection or intruder behavior prediction can be effective in certain applications, their limitations should be considered in the actual implementation. Since the accuracy of this method is dependent on certain assumptions that are typically based on the behavioral model of the target system, deviating from those assumptions will decrease its accuracy. Selecting an inaccurate model will lead to an inaccurate detection system. Therefore, selecting an accurate model is the first step towards solving the problem. Unfortunately selecting an accurate behavioral model is not an easy task as typical systems and/or networks are complex.

Shyu et al. [5] proposed an anomaly detection scheme, where PCA was used as an outlier detection scheme and was applied to reduce the dimensionality of the audit data and arrive at a classifier that is a function of the principal components. They measured the Mahalanobis distance of each observation from the center of the data for anomaly detection. The Mahalanobis distance is computed based on the sum of squares of the standardized principal component scores. Shyu et al. evaluated their method over the KDD CUP99 data and have demonstrated that it exhibits better detection rate than other well known outlier based anomaly detection algorithms such as the Local Outlier FactorLOF approach, the Nearest Neighbor approach and the kth Nearest Neighbor approach.

Yeung et al. [6] describe the use of hidden Markov models for anomaly detection based on profiling system call sequences and shell command sequences. On training, their model computes the sample likelihood of an observed sequence using the forward or backward algorithm. A threshold on the probability, based on the minimum likelihood among all training sequences, was used to discriminate between normal and anomalous behavior. One major problem with this approach is that it lacks generalization and/or support for users who are not uniquely identified by the system under consideration.

III. AN ARTIFICIAL IMMUNE BASED ANOMALY DETECTION TECHNIQUES

An anomaly detection approach usually consists of two phases: a training phase and a testing phase.In this subsection,we present various artificial immune algorithms that have been proposed for anomaly detection [7].That include Negative Selection Algorithm and Clonal Selection Algorithm.

Negative Selection Algorithm (NSA) revolves around the most interesting feature of self/non-self identification, pioneered by Forrest and her group [3]. NSA revolves around the theory of T-cell maturation in thymus and self tolerance of immune system. NSA generates detectors and eliminates those ones that detect self. This results in a group of detectors that has the potential to detect non-self. Different variations in NSA have been suggested over a period of time from its inception to solve problems of various domains including anomaly detection, fault detection and optimization.

Clonal Selection theory illustrates creation of immune cells when activated in the presence of an antigen. T-cells when encounters an antigen stimulate B-cells which are capable of matching antigens. This moment these B cells are stimulated, they initiate the process to produce clones of themselves. These clones are not the precise copy of the previous ones. Actually they undergo mutation to enhance their uniqueness to address and match different antigens . This B-cell become specific and becomes memory cell to facilitate adaptive immune response,so that it can react faster to the same

antigen if encountered in future .

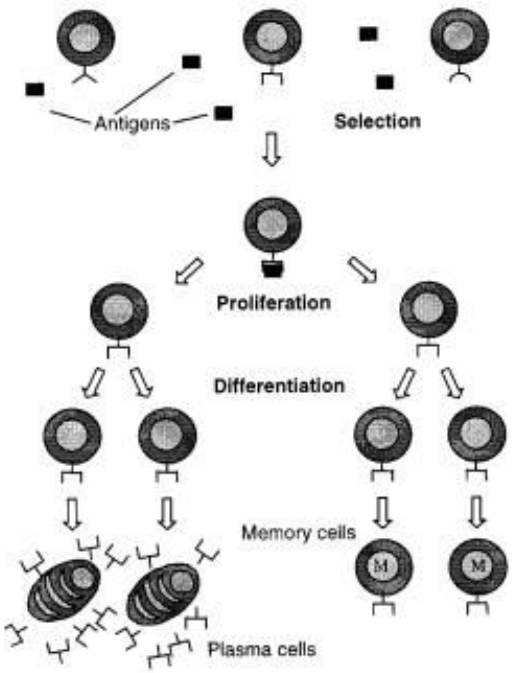

Fig. 2: Clonal Selection Theory
[8]

IV. EXPERIMENTS AND RESULTS

In order to determine the performance and possible advantages of our approach, we performed the experiments with a dataset such as sensor network dataset.Simulations are executed on Matlab and it took 8-10 secs to run.

The training data were either partially or completely composed of the elements of the normal class. For NSA the distance function was based on the Euclidean distance and in the case of CSA,it is based on affinity calculations.

In this section we discussed two artificial immune system mechanism for anomaly detection.That include Negative Selection Algorithm and Clonal Selection Algorithm.

A. NSA in Anomaly detection

Negative selection approach is frequently used in the anomaly detection domain because of astonishing similarities between requirements of problem space and features offered by NSA.

1) Detector generation and Anomaly detection: In our first experiment, we implemented simple NSA for a small dataset having normal packets only. We inserted anomalies at runtime and then detected the anomalies. Total anomalies inserted are 10. Simulations are executed in Matlab R2013a and it took 8 10 seconds to run.

TABLE I: Result of proposed NSA based Anomaly detection

Sr.Number	Packet Delayed	Packet Dropped	Wormholes
Iteration 1	20	31	18
Iteration 2	22	32	19
Iteration 3	20	30	17
Iteration 4	24	28	18
Iteration 5	20	28	19

In our second experiment,we implemented proposed NSA for large dataset such as sensor network dataset.First generate random detector set(d).After that this random detectors are compared with self set(S) by using matching rule.Here we use Euclidean distance(D) as the matching function.If the random detector match with self set,it is rejected.Otherwise it is added to the detector set(DS).In this stage we set a threshold value.If the Euclidian distance between Detector set(DS) and Training set(TS) is less than threshold value,it is discarded.After getting the detector set(DS) ,we proposed an injection feature to the detector set(DS).By this method the detector set(DS) can updated at any stage.

Detectors (d1, d2, d3,..dn) from Detector Sets (DS1, DS2,..DSN) are compared with the samples of Test Set . Each Detector Set (DS) uses its own detectors (d1, d2, d3,..dn) to detect the anomalies of the incoming Test Set .Then find the Euclidean distance between detectors(d) of detector set(DS) and Test set.Before that we set an Affinity threshold value.If the distance(D) is greater than Affinity threshold,the test sample is marked as anomaly. As a result Detector (d) increases count value for this sample of Test set . Information about count of test sample with a count value is shared with other detectors. Other detectors in Detector Set DS also compares the test sample and contributes to count. This process is repeated till all the Detector Sets(DS) test the Test Set. If the final count for a packet increases over the count Threshold (CT), then the packet is marked as an anomaly.

Detection Rate and False Alarm rate is evaluated to determine the proficiency of anomaly detection system. Detection rate is the number of anomalies detected out of the available ones, while False alarm rate represent how many normal is identified as an anomaly. Any system with high detection rate and low false alarm is considered to be good. Detection Rate (DR) is calculated by:TP/(TP + FN) * 100 where TP means True Positive and FN stands for False Negative. Higher detection rate is always desired. False Positive Rate is obtained by FP/(TN + FP) *100, where FP means false positive and TN represents true negative. Lesser false positive means that the system is properly detecting self as self.The proposed NSA shows high detection rate and less false alarm rate as compared with simple NSA. Result of second experiment is shown in table 1.But in simple NSA detector set is compared with test set by using string matching rule.

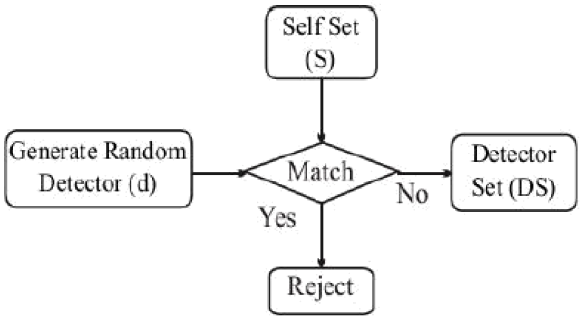

Fig. 3: Detector set Generation

B. CSA in Anomaly Detection

Immune Algorithm is derived through the study of immune response of Biological Immune System. It models how antibodies of the immune system learn the features of the intruding antigen . Clonal Selection Algorithm is a special class of Artificial Immune Systems. In this work, CLONALG algorithm which was originally proposed by De Castro and Van Zuben [9] is used. The algorithm starts by defining a purpose function f(x) which needs to be optimized. Some possible candidate solutions are created, antibodies will be used in the purpose function to calculate their affinity and this will determine the ones which will be cloned for the next step. The cloned values are changed, mutated with a predefined ratio and the affinities are recalculated and sorted. After certain evaluations of affinity, affinity with the smallest value is the solution closest to our problem.Block diagram of CSA is shown in figure 4.

Clonal Selection Algorithm can be listed as follows:

1) Generate a set of antibodies (generally created in a ran-dom manner) which are the current candidate solutions of a problem.
2) Calculate the affinity values of each candidate solutions.
3) Sort the antibodies starting from the lowest affinity. Lowest affinity means that a better matching between antibody and antigen.
4) Clone the better matching antibodies more with some predefined ratio
5) Mutate the antibodies with some predefined ratio. This ratio is obtained in a way that better matching clones mutated less and weakly matching clones mutated much more in order to reach the optimal solution.
6) Calculate the new affinity values of each antibody.
7) Repeat 3 through 6 while the minimum error criterion is not met

Classification task involve separating different classes based on Euclidean distance.The idea of fuzzy C means clustering technique has been adopted where the main purpose is to evolve generalized memory cells that are able to capture

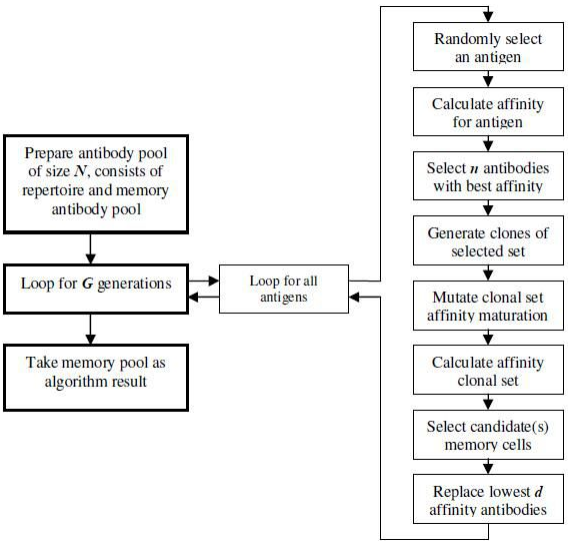

Fig. 4: Clonal Selection Algorithm

(detect) many of the similar structure of antigens. In fuzzy based clustering,dataset is grouped into n clusters with every datapoint in the dataset belonging to every cluster to a certain degree .The point which close to the center datapoint has the highest degree.The datapoint which is far away from center has lowest degree.

V. CONCLUSION

Artificial Immune System (AIS) is an active research area and researchers have been using AIS for network anomaly detection as well as other optimization problems.This paper presented a Negative Selection Algorithm (NSA) and a Clonal Selection Algorithm(CSA) for anomaly detection in Wireless Sensor Networks (WSNs).We first implemented simple NSA and tested on the small dataset having random anomalies, and results were calculated.Then, the proposed NSA is implemented for a large dataset such as Wireless Sensor dataset having normal and anomalous packets. CSA is implemented for the detector generation, and anomalies are identified using the detector set.For classification , CSA used Fuzzy C means clustering.Compared with simple NSA,both CSA and the proposed NSA shows high detection rate and less false alarm rate.

REFERENCES

[1] J. Shahabi, “A multi-set artificial immune system for searching optima in dynamic environments,” Ph.D. dissertation, Eastern Mediterranean University (EMU), 2012.
[2] R. Rizwan, F. A. Khan, H. Abbas, and S. H. Chauhdary, “Anomaly detection in wireless sensor networks using immune-based bioinspired mechanism,” International Journal of Distributed Sensor Networks, 2015.
[3] M. A. M. Ali and M. A. Maarof, “Malware detection techniques using artificial immune system,” in Proceedings of the International Conference on IT Convergence and Security 2011. Springer, 2012, pp. 575–587.
[4] C. Kruegel, D. Mutz, W. Robertson, and F. Valeur, “Bayesian event classification for intrusion detection,” in Computer Security Applications Conference, 2003. Proceedings. 19th Annual. IEEE, 2003, pp. 14–23.

[5] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang, "A novel anomaly detection scheme based on principal component classifier," DTIC Document, Tech. Rep., 2003.
[6] D.-Y. Yeung and Y. Ding, "Host-based intrusion detection using dynamic and static behavioral models," Pattern recognition, vol. 36, no. 1, pp. 229–243, 2003.
[7] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," Computer networks, vol. 51, no. 12, pp. 3448–3470, 2007.
[8] J. Brownlee, "Clonal selection theory & clonalg-the clonal selection classification algorithm (csca)," Swinburne University of Technology, 2005.
[9] L. De Charsto and J. Zuben, "Learning and optimization using clonal selection principle," IEEE Trans on Evolutionary Computation, Special issue on Artificial Immune Systems, vol. 6, no. 3, pp. 239–251, 2002.

