A review on Bio-inspired machine learning methodologies for Wireless network security.

George Katsonis
3rd Year CSS Student
American College of Thessaloniki
Thessaloniki, Greece
giorgos_katsonis@hotmail.com

ABSTRACT: As network systems increase in complexity, the demands of performance and security have grown proportionally. Biological networking systems have proven to be extremely compliant with modern network requirements, leading them to be in the spotlight of research. This paper will demonstrate methodologies for the application of machine learning to examine the nodes of a Wireless Sensor Network (WSN) and identify potential hazards, inspired by the human immune system. Following these methodologies, a system of trust is created for the nodes of the network, allowing for a light, robust and versatile model of network security.

Keywords—Trust management, Wireless Sensor Networks, Biologically Inspired, Security, Machine learning, MLBTRM, EBTRM.

I. INTRODUCTION:

A. Bio-systems.

A system in biology consists of any number of participating agents and can be of any complexity. Systems can be of the microscopic or the macroscopic scale; however, the scale is a testament neither towards complexity, nor applicability in man-made constructs. Examples provided by Rathore and Jha [6] include the process of photosynthesis, as well as the organized flight of birds, both being considered biological systems. A biologically inspired system employs features from a biological system, expressing them in a computer system. There are many characteristics of biological systems that make them excellent inspiration for computer systems [6], some being:

- They achieve high complexity by utilizing only a simplistic ruleset.
- They are highly adaptable and resistant to individual or small group failures.
- They self-organize and self-repair.
- They are resistant to external influence (noise).

These characteristics directly respond to modern network challenges [6], such as:

- Scalability. The capacity to increase the number of nodes in a network as needed, without sacrificing functionality.
- Dynamism. The ability to respond to factors of variability such as bandwidth or varying network conditions. Dynamism also includes the capability to self-organize and self-evolve, ensuring survivability.
- Resource management. Allocating network resources in an efficient manner to avoid bottlenecks.

Meisel et al. [4] provide an excellent demonstration of the relation between biological and computer systems in figures 1 and 2.

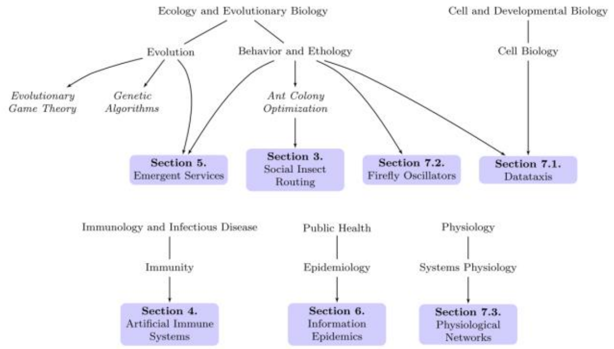

Ecology and Evolutionary Biology
Cell and Developmental Biology
Evolution
Behavior and Ethology
Cell Biology
Evolutionary Game Theory
Genetic Algorithms
Ant Colony Optimization
Section 5. Emergent Services
Section 3. Social Insect Routing
Section 7.2. Firefly Oscillators
Section 7.1. Datataxis
Immunology and Infectious Disease
Public Health
Physiology
Immunity
Epidemiology
Systems Physiology
Section 4. Artificial Immune Systems
Section 6. Information Epidemics
Section 7.3. Physiological Networks

Fig. 1. A taxonomy of network research inspired by biology for the topics covered in this survey, organized by the area of inspiration. Research topics are preceded by their section number.

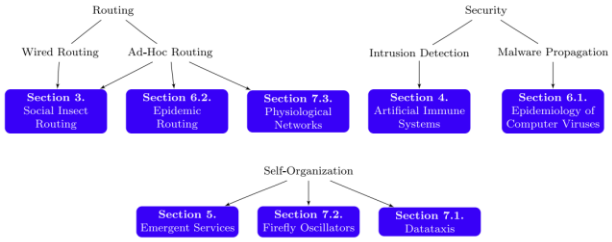

Routing
Security
Wired Routing
Ad-Hoc Routing
Intrusion Detection
Malware Propagation
Section 3. Social Insect Routing
Section 6.2. Epidemic Routing
Section 7.3. Physiological Networks
Section 4. Artificial Immune Systems
Section 6.1. Epidemiology of Computer Viruses
Self-Organization
Section 5. Emergent Services
Section 7.2. Firefly Oscillators
Section 7.1. Datataxis

Fig. 2. An alternative taxonomy of network research inspired by biology for the topics covered in this survey, organized by the area of application. Research topics are preceded by their section number.

B. WSN

Wireless sensors are low-cost devices that respond to external stimuli and transmit response data through radio signals. These devices are often deployed en masse with the purpose of covering a large area with a surveillance network, called a Wireless Sensor Network or WSN [1]. There are two architectures for WSNs, flat configurations and clustered configurations as shown in figure 3 [6]. These networks are promising for a number of applications including military intelligence systems, building security and traffic monitoring; however, securing them has proven to be a challenge, as well as their current major drawback limiting their use.

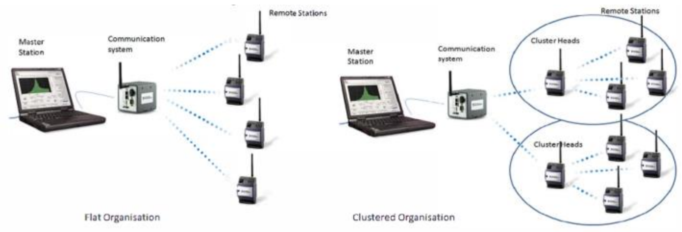


Fig 3. Architectures in WSN.

C. Mobile agents

A mobile agent is an autonomous, asynchronous, adaptive, and communicable program that is able to migrate between computational environments to accomplish specific tasks. After running, the mobile agent will access the local resources of the computer, interact with the local execution environment, sense dynamically the changes in the network and operate based on a set of predefined rules. Furthermore, agents are able to communicate with one another through messages to work collectively if necessary. Network congestion is minimized, due to the local access to the required data, as well as due to the encapsulation of the logic within the agent. While large scale applications of mobile agents are already in effect, the issue of security remains as mobile agents run locally on remote computers after traveling large scale networks [2].

D. Backbone

In a mobile backbone network, cluster heads have to be pre-deployed as backbone nodes that communicate through long ranges using radio signals instead of a message relay. Nodes share a communication channel based on their level and can only communicate to nodes belonging to the same cluster, while inter-cluster communication can be performed only through a higher-level backbone. As such, every cluster constitutes a separate mini network, connected to other clusters through the backbone. [1]

E. ARTM

ARTM, standing for Agent-based trust and reputation management scheme, allows for local trust and reputation management with minimal overhead in terms of extra messages and time delay. It is built on a clustered WSN with backbone, and its base is a mobile agent system. Every node stores its own trust (t-instrument) and reputation (r-certificate) locally, along with an agent responsible for managing those metrics for the hosting node. Before a transaction between two nodes can take place, one asks for the r-certificate of the other through queries performed by the mobile agents, that also decide whether or not to initiate the transaction. After the transaction is complete, the t-instruments are created and delivered to the nodes based on their quality of service. Having accumulated enough t-instruments, the agent of every node updates the r-certificate. As the agents run through the entire network it is imperative that (1) they are created and issued by a trusted authority and (2) they are not susceptible to analysis or manipulation [2].

F. Human Immune System

The human immune system is extremely complex, and the intricacies surrounding it are all yet to be uncovered. For the scope of this review, there will be a brief overview of the relevant mechanisms that are imitated by the bio-inspired systems examined. The two major categorizations of the immune system consist of the adaptive immune system and the innate immune system. The innate immune system focuses on prevention of infection and generalized defenses; however, for this study we will focus on the adaptive category. White blood cells can be categorized as phagocytes and leukocytes. Phagocytes are part of the innate immune system. The adaptive immune system is comprised of leukocytes, that can be further categorized as T-cells and B-cells. These categorizations are better illustrated in figures 4 and 5 [6].

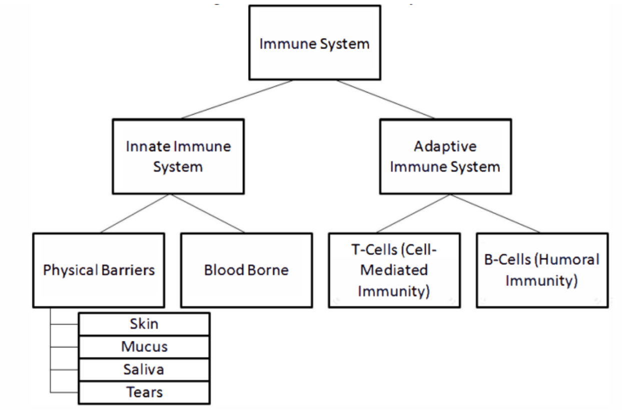


Fig 4: Human Immune System

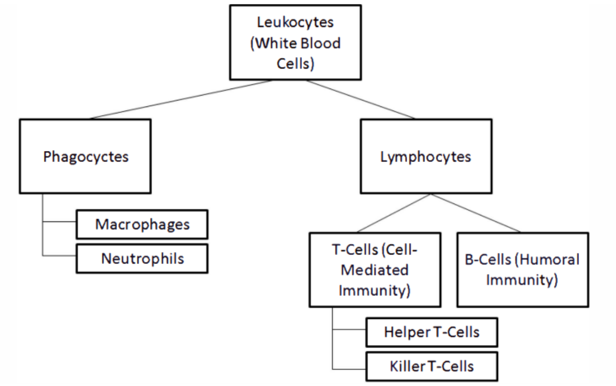

Leukocytes (White Blood Cells)
Phagocytes
Lymphocytes
Macrophages
Neutrophils
T-Cells (Cell-Mediated Immunity)
B-Cells (Humoral Immunity)
Helper T-Cells
Killer T-Cells

Fig 5: Categorization of Leukocytes

Lymph nodes are special tissue that hosts T and B cells, along with cells of the innate immune system called leukocytes and macrophages. These nodes are connected to the lymphatic circulation of the body and assist in the coordination of the two branches of the immune system in order to fend off pathogen attacks.

The B-cells are part of the humoral immune system, and their role is to create antibodies in response to the antigens created by pathogenic microorganisms. The antibodies are essentially markers that signify a target for destruction. After the target is marked, T-cells, part of the cellular immune system, are deployed to destroy it. This coordinated "search and destroy" system is the essential core that will be replicated by the methodologies to follow. Naturally, there are more classifications and categorization of the complex interactions between the parts of the immune system that were not mentioned here, as they remain outside the scope of this paper [6].

II. RESEARCH METHODOLOGY

In this part of the paper, we will present an overview of the two dominant bio-inspired methodologies for WSN security. The first is the focus of this paper and is called MLBTRM or Machine Learning based Bio-inspired Trust and Reputation Model, deriving inspiration from the human immune system. The second, is the EBTRM-WSN, or Enhanced Bio-Inspired Trust and Reputation Model for WSN, inspired from the organization of ant colonies.

A. MLBTRM

The MLBTRM is a three-step model designed to evaluate nodes inside a WSN as demonstrated in figure 6. These steps are, in order, the evaluation of the data by machine learning algorithms to determine if they are fraudulent, the generation of virtual antibodies by the cluster heads that are then transmitted to the sensor nodes and, finally, the generation of trust values for the nodes by the gateway [5].

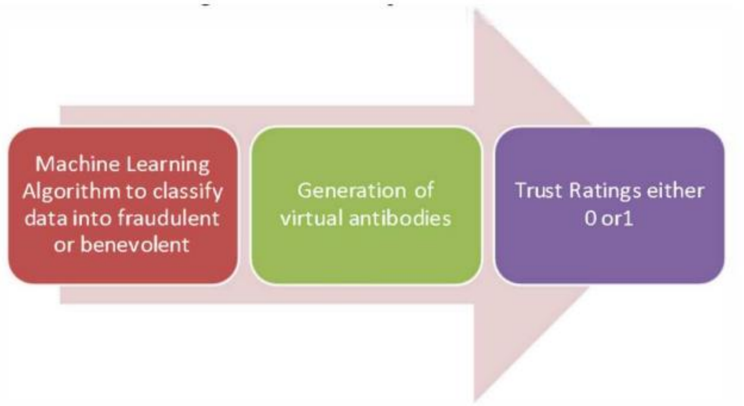
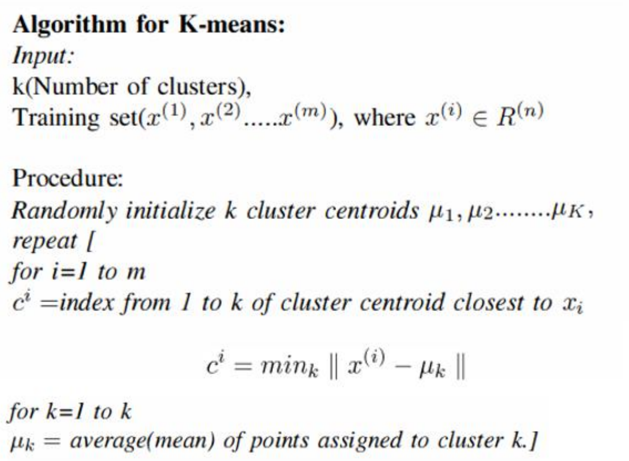

Machine Learning Algorithm to classify data into fraudulent or benevolent
Generation of virtual antibodies
Trust Ratings either 0 or1

Fig 6: Trust reputation model of MLBTRM [5]

The evaluation of the data is performed by the IDS, or Intrusion Detection System. The IDS operates as a secondary barrier of defense against attacks, the first being conventional intrusion prevention methods, and dynamically monitors network traffic, deciding if it detects symptoms of an attack. It operates on two levels. The first is Misuse Detection, where it relates new incoming traffic to already recorded signatures from the database, this approach does not implement machine learning. The second level is Anomaly Detection. Using machine learning techniques, the IDS will detect abnormal activities from already defined profiles in order to detect an attack. This approach can be implemented using supervised (prelabeled classes), semi-supervised (one prelabeled class) or unsupervised (no labeled classes) [6].

Unsupervised learning attempts to discern structures from random data and then use the K-means algorithm to create clusters.



```
Algorithm for K-means:
Input:
k(Number of clusters),
Training set($x^{(1)}, x^{(2)}.....x^{(m)}$), where $x^{(i)} \in R^{(n)}$

Procedure:
Randomly initialize k cluster centroids $\mu_1, \mu_2........\mu_K$,
repeat [
for i=1 to m
$c^i$ =index from 1 to k of cluster centroid closest to $x_i$

    $c^i = min_k \parallel x^{(i)} - \mu_k \parallel$

for k=1 to k
$\mu_k$ = average(mean) of points assigned to cluster k.]
```

Supervised learning uses classified data to describe a pattern and create a decision boundary. SMV, or Support Vector Machine, is used for this purpose and can make use of the clusters provided by the K-means algorithm to generate training data for the model. In figure 8 we can see the areas defined by K-means clusters are being used to classify the training data, with data falling into the red area being marked as fraudulent and data falling into the green area being marked good [6].

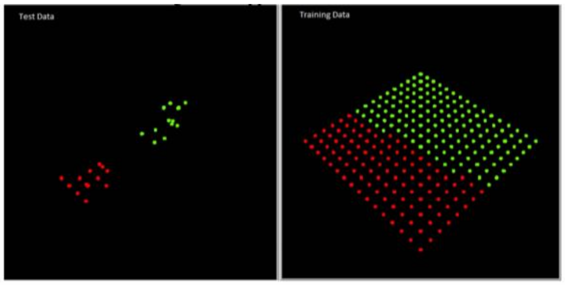


Fig 7: Support Vector Machine

Finally, an Anomaly Detection Algorithm is implemented in order to classify data that rest on top of the decision boundary, in order to further refine our selection. Following this process, a threshold is established that is used to classify all following data quickly and efficiently [6].

After the classification is complete the cluster head creates the virtual antibodies and transmits them to the sensor nodes. After some time, the gateway generates the trust for each node based on the antibodies it detects and, finally, if the trust of a node falls below a predefined threshold, the gateway sends a radio signal that terminates the connection to this node [5].

The correlation between MLBTRM and the human immune system is more clearly demonstrated in the following table [6].

TABLE I
RELATIONSHIP BETWEEN IMMUNE SYSTEM AND WSN

Immune System	WSN
B-cells	Cluster Head
Lymph Node	Gateway Nodes
Antigen	Corrupted data from nodes
Antibody	virtual antibodies
Normal cells in human body	Sensor nodes
T-cells attacking the infected cells	Disable radios on fraudulent nodes

B. EBTRM-WSN

Like the original BTRM-WSN, the enhanced version is based on an algorithm, inspired by how ants optimize their pathing, called Ant Colony System, or ACS. The way ants navigate is based on a pheromone secretion system. Every ant, while traveling douses its path with a pheromone that is detectable by other ants. The more ants travel through a specific path, the denser the pheromone becomes, and the more likely it is for a new ant to follow this path. After a while, there are several mainline paths of transport, supported by adjacent and branching paths in order to create an efficient network.

In ACS, agents are used as artificial ants, moving between the nodes, and evaluating them. Every time an ant reaches a node, it can chose based on the reputability of that and the adjacent nodes, whether it will continue traveling or remain and report its pathing back to the server. The server will then calculate a current best solution based on the paths with the most pheromone and will compare it with the global best solution. Should the new solution be more efficient than the previous, it will be tested and given a score based on the quality of the service provided. Through this process, the pheromone values will be modified for later iterations, repeating the cycle, and tracing any dynamic changes or attacks through this process [4].

The Enhanced BTRM-WSN proposed in [2] expands on the original by taking into consideration the length of the path and the number of ants that reached the same solution, aside the pheromone measurements. Thus, the following equation is used to calculate the quality of the path:

$$Q(S_k) = \frac{\overline{t_k}}{Length(S_k)^{PLF}} \cdot \%A_k$$

Where S_k is the solution returned from an ant k, $\overline{t_k}$ is the average pheromone of the path, P LF ∈ [0, 1] represents the path factor and %Ak is the percentage of ants that have selected the exact same solution as ant k.

This model also applies a peer trust system to generate trust values based on the satisfaction a peer has with other peers in their transactions. Ultimately, the server takes into consideration both the quality of the path and the trustworthiness of the peers when selecting the optimal path [2].

III. MAIN BODY/ PERFORMANCE ANALYSIS

TRMSim-WSN [7]

In order to conduct a performance evaluation on WSNs this open-source simulator is used by papers [2][3][6]. The reason it is favored, is due to its ability to easily compare models on WSNs using the same parameters, and thus provide objective comparisons between different solutions. In the evaluation performed by [3], there is a comparison between EBRTM-NSW, BRTM-NSW, and Peer Trust System. The metrics investigated are accuracy, path length, and energy consumption.

Accuracy is representative of the level of security, or reliability of the model, meaning the percentage of successfully selecting trustworthy sensors out of the total number of transactions.

Path length represents the average number of hops leading to the most trustworthy sensors as selected by the client. It is assumed that a shorter path is desirable for two reasons. First, the fewer number of intermediates increases the overall security of the network as it decreases the chances of a compromised sensor being included in the path. Second, a shorter path indicates a higher ease towards identifying trustworthy nodes, and thus leads to a shorter response time between the client and the server.

Energy consumption is extremely crucial, as the WSNs are characterized by low power sensors that are difficult to recharge. This metric considers the following actions that consume energy: 1) server nodes sending response services; 2) client nodes sending request messages; 3) relay nodes which do not provide services; 4) energy consumed by malicious nodes which provide bad services; and 5) the energy to execute the trustworthy sensor searching process of a certain trust and reputation system.

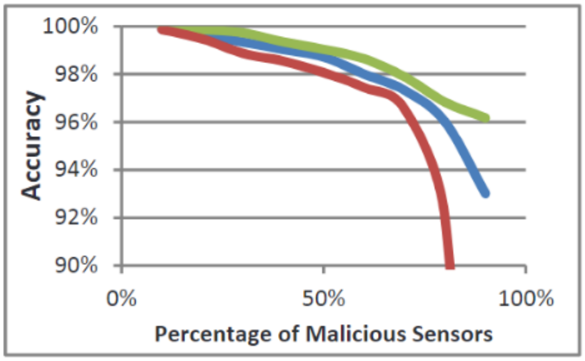


Fig 8: Accuracy in searching for trustworthy sensors.

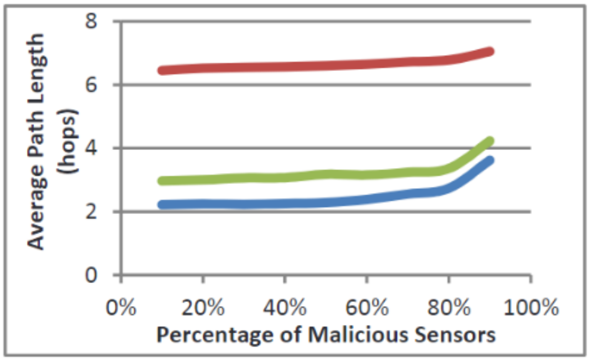


Fig 9: Average path length leading to trustworthy sensors.

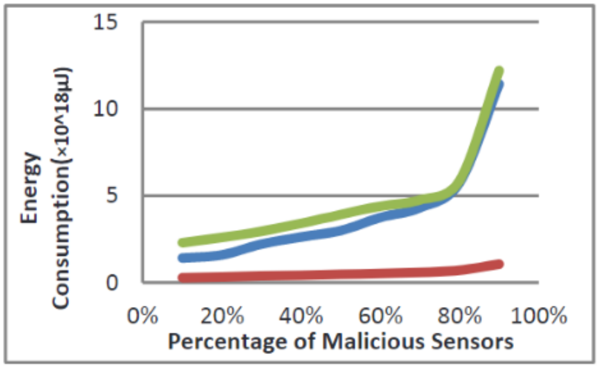


Fig 10: Overall network energy consumption.

Comparison between: BTRM-WSN (Blue ---); Peer Trust (Red ---); and EBTRM (Green ---) [3]

From the above figures we can see that EBTRM outperforms the other models in two out of three metrics with its performance increasing as the percentage of malicious censors increases. According to the authors, the difference between BTRM-WSN and EBTRM on the average path length is negligible, with both vastly outperforming Peer Trust.

Unfortunatly, there is no proper evaluation testing for MLBTRM in WSNs; however, using TRMSim-WSN, a team of researchers could feasibly conduct a series of comparison experiments to determine its relevant efficiency.

IV. CONCLUSIONS

In this paper we examined the literature around Bio-inspired security systems for WSNs. We provided a detailed overview on all the relevant terms and methodologies, for both MLBTRM and EBTRM, and examined the latter for efficiency, determining its superiority to both BTRM and Peer Trust. We illustrated the connections between the models and the biological systems they were inspired from, namely the human immune system and the organization of ant colonies, while we also briefed the relevant algorithms applied. Naturally, many complexities were not included, so as to maintain the scope of the paper; still, the goal of introducing the reader to the topic was sufficiently achieved.

BIBLIOGRAPHY:

[1]

A. Boukerch, L. Xu, and K. EL-Khatib, “Trust-based security for wireless ad hoc and sensor networks,” *Computer Communications*, vol. 30, no. 11–12, pp. 2413–2427, Sep. 2007, doi: 10.1016/j.comcom.2007.04.022.

[2]

F. Gomez Marmol and G. Martinez Perez, “Providing trust in wireless sensor networks using a bio-inspired technique. Telecommunication Systems 46, 163-180,” *Telecommunication Systems*, vol. 46, pp. 163–180, Feb. 2011, doi: 10.1007/s11235-010-9281-7.

[3]

H. Marzi and M. Li, “An Enhanced Bio-inspired Trust and Reputation Model for Wireless Sensor Network,” *Procedia Computer Science*, vol. 19, pp. 1159–1166, Jan. 2013, doi: 10.1016/j.procs.2013.06.165.

[4]

M. Meisel, V. Pappas, and L. Zhang, “A taxonomy of biologically inspired research in computer networking,” *Computer Networks*, vol. 54, no. 6, pp. 901–916, Apr. 2010, doi: 10.1016/j.comnet.2009.08.022.

[5]

H. Nunoo-Mensah, K. O. Boateng, and J. D. Gadze, “The adoption of socio- and bio-inspired algorithms for trust models in wireless sensor networks: A survey,” *Int J Commun Syst*, vol. 31, no. 7, p. e3444, May 2018, doi: 10.1002/dac.3444.

[6]

H. Rathore and S. Jha, *Bio-inspired machine learning based Wireless Sensor Network security*. 2013, p. 146. doi: 10.1109/NaBIC.2013.6617852.

[7]

H. Nunoo-Mensah, K. Boateng, D. Gadze, and G. Klogo, “SoTRMSim: Sociopsychological Trust and Reputation Models Simulator for Wireless Sensor Networks,” vol. 12, p. 4, Oct. 2018, doi: 10.5120/ijais2018451775.