# Preemptive Performance Monitoring of a Large Network of Wi-Fi Hotspots: An Artificial Immune System

Pheeha Machaka, Antoine Bagula

# Preemptive Performance Monitoring of a Large Network of Wi-Fi Hotspots: An Artificial Immune System

Pheeha Machaka[1], Antoine Bagula[2]

Intelligent Systems and Advanced Telecommunication Laboratory (ISAT)
Department of Computer Science, Room 317 Computer Science Building, 18 University Avenue, University of Cape Town, Rondebosch, Cape Town, South Africa, 7701
1. pheeha.machaka@uct.ac.za, 2.antoine.bagula@uct.ac.za

**Abstract.**
[1]This paper addresses the problem of network monitoring by proposing an Artificial Immune System (AIS) system to achieve situation recognition and monitoring in a large network of Wi-Fi hotspots as part of a highly scalable preemptive monitoring tool for wireless networks. Using a set of data extracted from a live network of Wi-Fi hotspots managed by an ISP, we integrated an AIS algorithm into a data collection system to detect anomalous performance and aberrant behavior in the ISP's network. The results reveal the efficiency of the AIS system in terms of both anomaly performance and aberrant behavior on several test case scenarios.

Keywords: Performance Monitoring, Non-Self Detectors, Artificial Immune Systems, Anomaly Detection.

## 1 Introduction

Wireless Fidelity (Wi-Fi) is a wireless networking technology that uses radio waves to provide high-speed wireless internet connections. It is based on the family of IEEE 802.11 standards and builds upon a fast, easy and inexpensive networking approach [1] that uses a client-server model where the access points (AP) also called hotspot, plays the role of server while its client devices range from laptops to cellular mobile devices and Personal Digital Assistants (PDAs). The Wi-Fi APs broadcast signals to Wi-Fi-capable client devices that detect and receive broadcast messages from hotspots within their AP's range, and thus connect to the Internet. These APs are currently installed in different business settings, such as coffee shops, restaurants, hotels and conference rooms to provide wireless Internet access to clients located in these settings.

Performance monitoring is an important task upon which large Wi-Fi network deployment depends. As traditionally implemented, performance monitoring is based on a reactive network approach where the operating system software only warns the network administrators when a problem occurs. This approach leads to both the halting of important network processes and the hampering of critical business processes of the organization. Pre-emptive network monitoring provides the potential to prevent the occurrence of faults by analyzing the status of the network components to create a fail-safe network status or allow a smooth migration from a faulty to fail-safe network status. The popularity of Wi-Fi technology has led to a large scale deployment of thousands of hotspots networks. These hotspots generate huge amounts of monitoring data which require efficient data handling methods to analyze data, recognize anomalous hidden patterns and implement fault tolerance. While statistical analysis has been deployed in many cases to address this issue, soft computing methods borrowed from the human immune system are emerging as powerful tools used in anomaly detection and security monitoring systems.

### 1.1 Related Work

There has been work done in the field of AIS, which has mostly focused on intrusion detection, detection of computer viruses and network security [2], [3], [4], and [5]. While [6] addresses anomaly detection for a refrigerator system, the works in [3] and [4] describe AIS mechanisms that differentiate between self and non-self performance with application to a dataset of executable files infected by computer viruses. In [7], a security authentication system inspired by the immune system using the negative selection algorithm is introduced. This system generates a set of anti-passwords (a negative image of the password set) used as a first line of authentication kept separate from the positive authentication system (secured). Using datasets with approximately 2500 most common passwords, the authors generated a set of anti-passwords that defined the non-self of the password authentication system with the objective of providing a proof-of-concept for negative authentication systems using AIS to reduce the effect of brute force attacks on authentication systems by hackers. The tool breakage detection tool developed in [8] aims to detect tooth breakage in different environments. Using this tool, a negative selection technique was successfully implemented to detect the tooth breakage from dynamic variations of the cutting force signal.

Building upon the AIS' success, this paper presents a network monitoring tool that uses the Artificial Immune System (AIS) to achieve pre-emptive monitoring of a large scale live ISP network operating in the city of Cape Town in South Africa. Some of the main goals of this paper consist of providing answers to the following questions:

- What kind of data can be extracted using AIS methods on the network performance?
- How useful is the data that will be extracted?
- How does AIS perform under different network profiles?

- Finally, can we use AIS for network performance monitoring?

The remainder of this paper is organized as follows. Section 2 describes the development of the AIS system while section 3 describes our experimental setting. Section 4 presents the experimental results while our conclusions are presented in section 5.

## 2 Development of AIS

The Human Immune System (HIS) is a robust, complex network of specialized cells and organs that defend the human body against a sea of harmful microorganisms called antigens. Building upon its capability of differentiating the cells and molecules of the body into self (what belongs) and non-self (what does not belong) [9] and [11], the HIS has led to the development of the Artificial Immune System (AIS) computational paradigm that may solve complex computational problems.

### 2.1 Human Immune System (HIS)

As depicted by Figure 1, once the body recognizes an invasion by antigens (I-III), the immunological response (IV-VI) consists of neutralizing or destroying the invading antigens through the release and use of B-cells and T-cells (lymphocytes that originate from the bone marrow and play an important role in the immunological response). The B and T-cells will then destroy the invading antigens and remove them from the body.

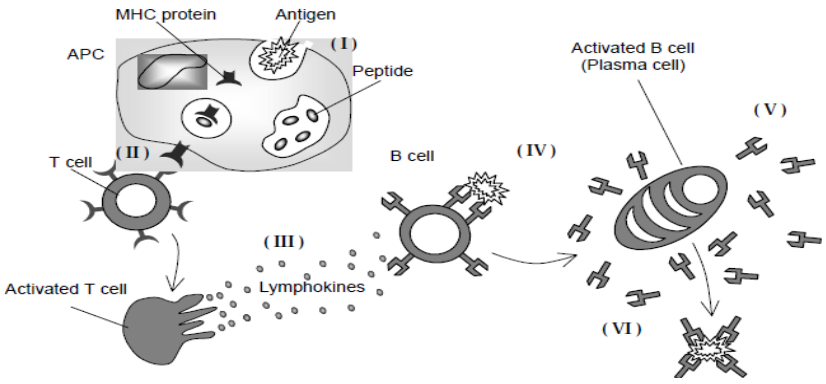


**Fig. 1.** Pictorial representation of the essence of the acquired immune system mechanism (taken from [10])

### 2.2 Artificial Immune System (AIS) and Negative Selection

A biologically inspired computational technique, called Artificial Immune System (AIS) has emerged from the HIS metaphor. The HIS features that are of particularly relevant to AIS include matching, diversity and distributed control. These features

have been implemented in many AIS mechanisms such as the immune system's Immune Network Theory, Negative Selection mechanism and Clonal Selection Principle to solve real world science and engineering problems. The focus of this paper lies on the negative selection mechanism of the immune system with its application to pattern recognition in Wi-Fi network monitoring.

**Negative Selection Mechanism.** The HIS mechanism provides tolerance for self cells, and reacts against unknown antigens. The HIS generates antibodies through a pseudo random rearrangement process and undergoes a censoring process in the thymus as described by [9] and [11]. Only those antibodies that do not bind to the self-protein are allowed to leave the thymus to circulate throughout the body to perform immunological functions and protect the body against foreign antigens. This mechanism gave developments to the Negative Selection Algorithm (NSA) described below.

**Negative Selection Algorithm.** In [3], the negative selection algorithm (NSA) was first proposed as a means of detecting unknown or illegal strings for virus detection in computer systems. This was inspired by a change detection mechanism which is based on the way the human immune system distinguishes self cells from non-self. The algorithm follows the two steps as shown by Figure 2.

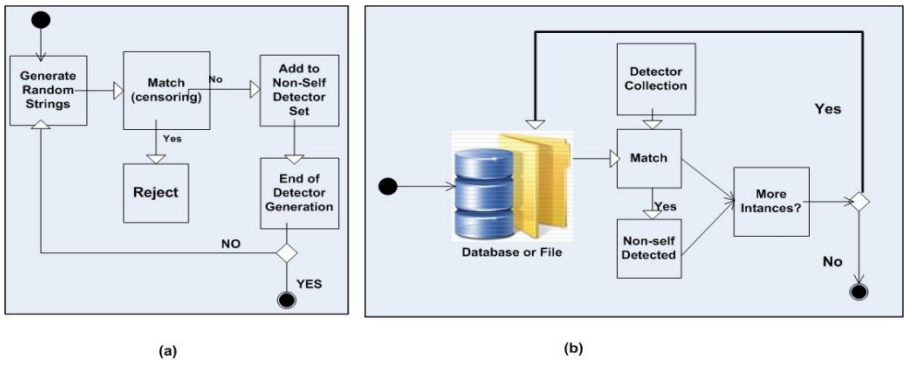


**Fig. 2.** (a) Generation of valid detector set (stage 1); (b) Monitoring the protected strings for changes (stage 2). (Adapted from [3])

1. **Generate a set of detectors.** Each detector is a string representation that does not match any of the protected data (self). This is the censoring stage shown in figure 2(a). The number of detectors generated will vary according to combination of total number of monitored variables and those that are undesired variables. If a non-self detector is found, it is added to the detector set. This process is repeated until the desired number of detectors is reached or no more detectors are found.
2. **Monitor the protected strings.** The AIS continuously monitor for changes in the system, by matching encoded instances from the database/system. If a non-self

detector is activated, the correct action will be taken, thus creating an alert and logging this with the monitoring system as show in figure 2(b).

# 3 The Experimental Model

## 3.1 The Wi-Fi Network

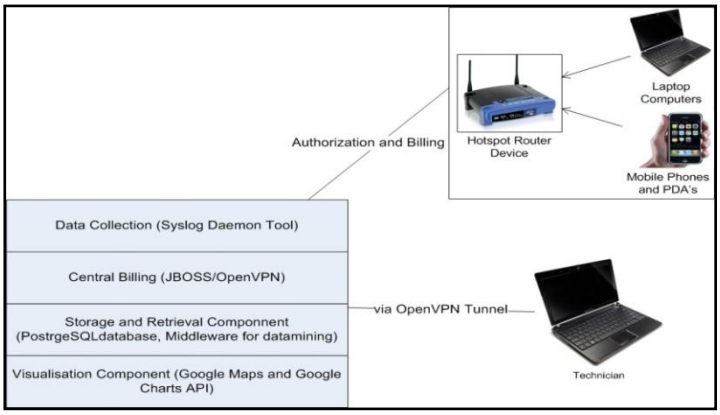


**Fig. 3.** The Wi-Fi network monitoring tool.

As depicted by Figure 3, a Wi-Fi network monitoring tool was developed around WRT54GL routers with three components: a storage component, a visualization component and a data collection component. The "Storage component" provides storage for the enormous amount of data that is harnessed during network monitoring. A postgreSQL database was used for this component. The "Visualisation Component" provides visual display of the performance of the network. Google Maps was used as a visual tool as the network monitored covered a large metropolitan city of Cape Town. Google Chart Tools were used to visualize performance statistics and individual router performance. Of most importance was the "Data Collection component" which provided monitoring and data gathering capabilities for the monitoring tool. The data gathering component was implemented using the "Syslog protocol" program installed on each of the 615 Cisco WRT54GL router devices in more than 400 hotspots in the Cape Town network. The program was left to run from 2009-07-03 12h00 to 2009-09-02 04h00 collecting monitoring data at every hour's interval, thus leading to up to 356537 items in the experiment dataset .

## 3.2 Performance Metrics

We conducted a set of experiments based on three performance metrics which are usually used in performance evaluation of Wi-Fi networks. These include:

**Uptime and Downtime.** This metric measure the time a device has been up and running. It reveals the availability, stability and reliability of the communication device when used in the network.

**Load Average.** Measures the "congestion rate" for the device based on the number of users connected to the device.

**Radio Noise and Channel.** Wi-Fi uses the 2.4 GHz spectrum band which is shared with other devices like cell phones, GPS, RFID tags and Bluetooth devices. Note that the proliferation of devices using the free 2.4 GHz ISM band leads to more congested and noisy Wi-Fi devices.

### 3.3 Experimental Plan

Our experimental plan consisted of the four stages shown in Figure 4.

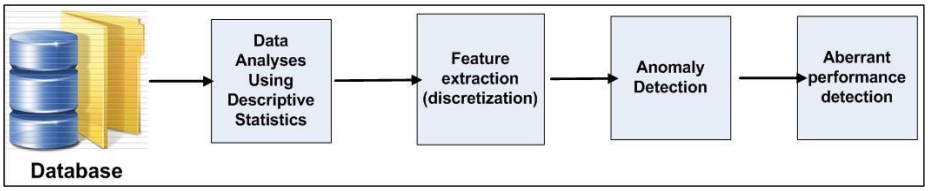


**Fig. 4.** Experimental Plan

**Data analysis.** In this stage, descriptive statistics were pulled out of the dataset to reveal the network's performance level in terms of time series characteristics and descriptive summary statistics. This allowed a summary of the observations by considering their central tendency and statistical dispersion.

**Feature extraction.** At this stage, the readings/observations received were discretized or reduced for further processing.

**Anomaly Detection.** This stage aimed at discovering those readings/observations associated to unacceptable device performance.

**Aberrant behavior.** This stage uses the observations obtained from the "Data Analyses stage" to detect aberrant behavior. An observation will fall into the aberrant behavior category when its observed values fall outside the statistical confidence band, an outlier. A total deviation of the observation will depend on a defined Delta ($\delta$) parameter.

### 3.4 The AIS Implementation

**Encoding.** Using numeric attributes to encode the three performance metrics, we considered categorical data for our pattern recognition. Therefore, the dataset was discretized for further processing and the following three categories were considered as performance levels achievable by the three performance metrics:

```
1 → LOW LEVELS
2 → MODERATE LEVELS
3 → HIGH LEVELS
```

**Selection.** Each of the three performance metrics was encoded in the following way:

*Noise.* In the dataset, the noise performance was multiplied by 100 (for calibration purposes) and noise values were categorized using the following encoding:

```
If             noise < -8500  then encode → 1
Else if -8500 < noise < -7000  then encode → 2
Else if        noise >-7000   then encode → 3.
```

*Load.* The load expressed the UNIX kernel load average multiplied by 100. The load values were categorized using the following encoding:

```
If           load <70    then 1(LOW)
Else if 70 <  load<100   then 2 (MODERATE)
Else if      load>100    then 3 (HIGH)
```

*Uptime or Downtime.* The Uptime is the device uptime considered as the current uptime (in minutes) since last reboot. The downtime rate per router is defined by

$$\text{Downtime rate} = \frac{\text{number of reading with uptime=<60 (indicating a device reset)}}{\text{Total number of readings}}$$

and the Downtime was categorized using the following encoding:

```
If           downtime-rate> 1%  then 1 (LOW)
Else if 5% <  downtime-rate< 1% then 2 (MODERATE)
Else if      downtime-rate< 5% then 3 (HIGH)
```

**Similarity Measure or Fitness Function.** A fitness function determines if an antigen represented by a string "NLD", where N, L, D express the Noise, Load and Downtime respectively, is fit for further processing. Thus to define the non-self, the following fitness function was used for a three variable antigen "NLD".

```
Fitness Function (f) = (N × L × D) mod 3
For non-self,    (f) = 0.
```

**Aberrant Performance Detection.** To detect aberrant behavior in performance, statistical confidence bands were used. They measured deviations of an observation in a time series. A deviation depended on the Delta (**δ**) parameter whose sensible values were taken between 2 and 3 [12].

### 3.5 Test Cases

We conducted experiments using four test case scenarios revealing Wi-Fi operating constraints from loose (e.g. rural setting where QoS is not a issue) to the most stringent(e.g. suburban setting where modern applications demand QoS). In the experiments conducted, four sets of test cases were devised. These test cases are defined by Table 1 in terms of Noise, Load, Downtime rate (overall time the device was down) and the confidence band **δ.**

**Table 1.** Parameter settings for each experiment test case

| Test Cases | Noise | | Load | Downtime | Confidence Band (δ) |
|---|---|---|---|---|---|
| 1 | high | -7000 | 100 | 4% | 1.00 |
| | moderate | -8000 | 70 | | 2.00 |
| | low | -9000 | 0 | | 3.00 |
| 2 | high | -7250 | 100 | 3% | 0.25 |
| | moderate | -8250 | 75 | | 1.75 |
| | low | -9250 | 0 | | 2.75 |
| 3 | high | -7500 | 100 | 2% | 0.50 |
| | moderate | -8500 | 80 | | 1.50 |
| | low | -9500 | 0 | | 2.50 |
| 4 | high | -7750 | 100 | 1% | 0.25 |
| | moderate | -8750 | 85 | | 1.25 |
| | low | -9750 | 0 | | 2.25 |

## 4 Experimental Results

### 4.1 Random Detector Generation stage

We conducted a first set of experiments to find the set of random detectors generated by the first and second stages of the NSA. These detectors fit the non-self description of the NSA problem. As depicted by Table 2, only 18 possible anti-detectors were found by our AIS system.

### 4.2 Anomaly Detection

Using the parameters described in the test cases of Table 1, we conducted another set of experiments to detect anomalous network performance in terms of percentage

of different types of faults (antigens) found by the AIS system and probabilities of finding these faults. As depicted by Table 2, the experimental results revealed that the probabilities of finding non-self detectors (anomalous behavior) increase slightly as the AIS parameters become stricter. This is shown by the increase in detection probability in Table 3 for each test case.

**Table 2.** Non-self Detectors for the experiment

| Antibody Code | Noise | Load | Uptime | | Antibody Code | Noise | Load | Uptime |
|---|---|---|---|---|---|---|---|---|
| 113 | low | low | high | | 311 | high | low | low |
| 123 | low | moderate | high | | 312 | high | low | moderate |
| 131 | low | high | low | | 313 | high | low | high |
| 132 | low | high | moderate | | 321 | high | moderate | low |
| 133 | low | high | high | | 322 | high | moderate | moderate |
| 213 | moderate | low | high | | 323 | high | moderate | high |
| 223 | moderate | moderate | high | | 331 | high | high | low |
| 231 | moderate | high | low | | 332 | high | high | moderate |
| 232 | moderate | high | moderate | | 333 | high | high | high |
| 233 | moderate | high | high | | | | | |

**Table 3.** Detection Rate of Anomaly and Aberrant Faults

| Test Case | Anomaly | Aberrant |
|---|---|---|
| 1 | 70.53% | 29.47% |
| 2 | 71.64% | 28.36% |
| 3 | 72.45% | 27.55% |
| 4 | 73.41% | 26.59% |

While the first stage of the AIS experiment revealed a set of 18 anti-detectors generated, only 8 antigens were detected as anomalous behavior in the network's performance. This reveals that most of the antigen space was covered by the anti-detectors. The results depicted by Figure 5 were obtained from the following test cases:

**Test Case 1.** It represented the most loose but yet acceptable constraints in terms of network performance. In this test case experiment, above 99% of the faults were of type "311" representing high noise anomalous performance in the network.

**Test Case 2.** Case 2 revealed similar performance pattern as Case 1 by showing 99% of type "311" faults: high noise anomalous behavior.

**Test Case 3.** Case 3 revealed a slight shift from the results observed in the test cases 1 and 2. The results revealed more than 99% detectors of type "321" found by the AIS system. This behavior expresses a high level of noise in the network with moderate load performance and low uptime. This kind of shift in anomaly is due to the fact that the load constraints were made stricter, thus picking up moderate load performance.

**Test Case 4.** The performance constraints for this part of the experiment were stricter for all variables. Test case 1, 2, and 3 followed a similar trend where a 99% detection rate was found for those devices with high levels of noise. In Case 4, more than 99% detection rate was found for type "113" detectors, representing those devices that experience very high levels of downtime.

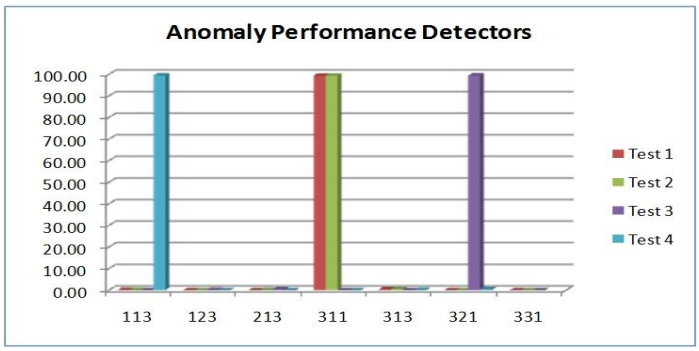


**Fig. 5.** Bar graph of results showing detection rate for the anomaly detection stage of the AIS experiment.

### 4.3 Aberrant Behavior Detection

We conducted another set of experiments to detect changes in normal performance that are significant enough to be regarded as outlying performance. The parameter that was used to measure deviation from normal performance is the Delta (**δ**) parameter. This parameter was set to different values for each test case as described in Table 1. Our experimental results revealed that 11 out of 18 detectors were discovered by the AIS system. These results depicted by Figure 6 reveal that:

**Test Cases.** In all the four cases, a significant percentage of Aberrant Performance detectors of type "113" were successfully detected with a detection rate of ~63.0%. This "113" detector indicates high downtime rate and implies that ~63.0% of the network devices suffered from high variations in the downtime rate. The second most significant aberrant performance detector was of type "313" with a detector rate of 29%. The "313" detector represent devices which have experienced high variations in downtime together with high variations in the level of noise.

The "311" detector type represent devices that are experiencing aberrant behavior in noise levels. Approximately 2.2% of the devices were experienced high noise level variations while they were performing well in terms of load and uptime.

**Changes in detection.** We conducted another experiment by changing the delta and downtime parameters to evaluate how the change will impact on the detection rate and device type. The results revealed a reduction in the type "113" detectors for each test case, for an average reduction of 0.75%. As the granularity of delta parameter was made coarse, a reduction in detector rate for detector type "313" was discovered and from one test case to the other, the average reduction was 0.25%.

When varying the AIS parameters, for each test case, a different set of results was produced by the AIS system, indicating that different parameter settings lead to the identification of different aberrant behaviors.

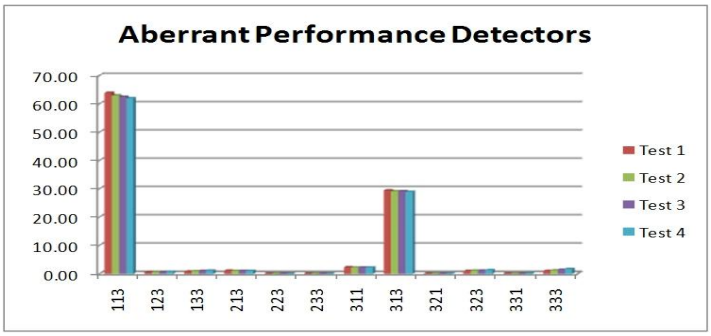


**Fig. 6.** Aberrant Performance detection stage of the AIs system.

### 4.4 Comparison with existing Monitoring tools

**MRTG (**http://oss.oetiker.ch/mrtg/**).** It monitors traffic load on network links and uses Simple Network Management Protocol(SNMP) to gather data from device. It uses the SNMP protocol which was found to be more bandwidth intensive in comparison to the Syslog protocol used in our monitoring tool.

**RRDTool(http://www.mrtg.org/rrdtool/).** It uses Round Robin Database, that stores data and also allows graphing capabilities, but the graphing capabilities are not as scalable and powerful as those presented by our monitoring tool, which includes powerful and intelligent techniques of analyzing and presenting data in a scalable fashion.

## 5 Conclusion

Building upon a hybrid solution using descriptive statistics and AIS, this paper presents a preemptive Wi-Fi network monitoring system that uses the negative selection algorithm (NSA) to achieve pattern recognition in an ISP network that provides hotspot access to clients in different settings in Cape Town. We present the main steps of our AIS implementation model and show through experimentation that the AIS system can reveal hidden pattern in a large network of Wi-Fi Hotspots in terms of anomalous and aberrant behavior. When deployed in a preemptive monitoring setting, the information resulting from our AIS system can further be used in future situation awareness to predict faulty scenarios in the Wi-Fi network and take preventive measures. This has been reserved for future work.

## 6 References

1. Vaughan-Nichols SJ. The challenge of wi-fi roaming. Computer 2003; 36: 17-19.
2. Luther K, Bye R, Alpcan T, Muller A, and Albayrak S. A cooperative AIS framework for intrusion detection. 2007; IEEE International Conference on Communication.
3. Forrest S, Perelson AS, Allen L, and Cherukuri R. Self-nonself discrimination in a computer. 1994, IEEE Computer Society Symposium on Security and Privacy.
4. Forrest S, Hofmeyr SA, Somayaji A, and Longstaff TA. A sense of self for unix processes. 1996 IEEE Symposium on Security and Privacy.
5. Dasgupta D, Gonzalez F. An immunity-based technique to characterize intrusions in computer networks. 2002 Evolutionary Computation, IEEE Transactions on Evolutionary Computation.
6. Taylor D, Corne D. An Investigation of the Negative Selection Algorithm for Fault Detection in Refrigeration Systems. In: 2003 Lecture Notes in Computer Science: Artificial Immune Systems. Berlin / Heidelberg: Springer, 2003: 34-45.
7. Dasgupta D, and Saha S. A biologically inspired password authentication system. 2009; Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies 5: .
8. Dasgupta S, Forrest S. Tool breakage detection in milling operations using a negative-selection algorithm. 1995; Technical Report Technical Report No CS95- 5, Department of Computer Science, University of New Mexico.
9. Schindler LW. Understanding the Immune System. United States Of America: DIANE Publishing. 1991: 40.
10. De Castro LN, Von Zuben FJ. Artificial immune systems: Part II–A survey of applications. 2000; Technical Report RT DCA 02/00 Universidade Catolica de Santos, Coordenacao de Pos-Graduacao e Pesquisa (COPOP) .
11. Jeanne Kelly. Understanding the Immune System - How It Works . United States: National Institute of Allergy and Infectious Diseases Science Education. 2007: 60.
12. Robinson S. Simulation: The Practice of Model Development and Use 1st Edition. Chichester, England: John Wiley & Sons Ltd. 2004: 339.

The financial assistance of the National Research Foundation (NRF) and Telkom SA Centre of Excellence (CoE) towards this research is hereby acknowledged.