Intrusion Detection Based on the Immune Human System

Kathia Regina L. Jucá† kathia@npd.ufsc.br Azzedine Boukerche‡1 boukerche@cs.unt.edu João Bosco M. Sobral† bosco@inf.ufsc.br

† Department of Informatics and Statistics Federal University of Santa Catarina Santa Catarina, Brasil
‡ Department of Computer Science University of North Texas Texas, USA

Abstract

In recent year, we have seen a growing interest in computational methods based upon natural phenomena with biologically inspired techniques, such as cellular automata, immune human systems, neural networks, DNA and molecular computing. Some of these solution techniques are classified under the realm of a general paradigm, called biocomputing. In this paper, we propose a security system for fraud detection of intruders and improper use of computer system operations. Our technique is based upon data analysis inspired by the natural immune human system. This paper shows how immune metaphors can be used efficiently to tackle the computer system security problem.

1. Introduction

This paper is motivated by the needs in the continuous changing computer network industry, and the security of the computer systems. As we all know, even the most modern computer systems have challenging and severe security problems. Most applications and operational systems are heavily affected by the security failures at different levels. If there are no verification and monitoring procedures in the computational system, the security services (such as access control, firewalls, etc.) become ineffective. Security policies define requisites and use rules predetermined by the computational and departmental services. An intrusion can be defined as a group of actions, which endanger integrity, confidentiality or availability of a given resource. In this paper, we propose a security system for fraud detection of intruders and improper use of computer system operations. Our system acts solely upon the usage log files and the registration activities of the users. Our technique is based upon data analysis inspired by the natural immune human system. In this paper, we show how immune metaphors can be used efficiently to tackle the security problem of computer system. We also describe how our scheme extracts salient features of the immune human system and maps them to computer system using a software package designed to identify security violations of the computer system as well as unusual activities according to usage log files.

1 A. Boukerche’s work was supported by Texas Advanced Research (ARP/ATP) Program and UNT Faculty Research Grants
The remainder of the paper is organized as follow. We present the basic background on both the human and artificial immune systems in section 2. In Section 3, we present our scheme and discuss the registration activities of the users as well as the computer system’s usage. Section 4 reports the experimental results we have obtained. In Section 5, we present the conclusion and our future work.

2. Background on Immunology

In this section, we shall present the basic background on immunology. The reader who is familiar with the basic concepts of immunology may skip this section.

The immune system [4,6,8,11,12] is a complex network of organs and cells responsible for the organism defenses against alien particles. One of the main features of the immune system is its capability to distinguish between self and non-self genes. Every cell of the organism possesses molecules in its structure which are characterized as self genes. Whereas, the molecules that constitute alien organisms are characterized as non-self genes.

2.1. Immune System Anatomy

The organs that compose the immune system can be found throughout the entire body. These organs are referred to as lymphoid organs since they are more concerned about the growth, development, and deployment of the lymphocytes cells.

• Lymphocytes – The lymphocytes are small white cells. Their responsibility relies upon specifying the activities to be taken by the immune system. The two major classes of lymphocytes are: B-cells, T-cells.

• Antibodies – Antibodies can be classified under a large family of molecules known as immunoglobulins. They are described as follows:
  - IgG – is the largest immunoglobulins in the blood, working effectively against micro-organisms;
  - IgM – usually possesses a star-shaped surface, and is very effective against bacteria;
  - IgA – founds in body fluids. It has the responsibility to defend the body’s entrances;
  - IgE – normally appears in small concentrations. It is involved mainly to defend the body against parasites;
  - IgD – founds almost exclusively in B-cells membranes, which regulates the activation of the cells.

• Immunity: Natural and Acquired – When an organism is exposed to a disease, B-cells and T-cells are activated, and some of these cells are able to memorize the disease. When the organism faces the same antigen in another circumstance, the immune system will recognize it and will quickly destroy it. They have a passive natural immunity though, being protected in their first months by antibodies received from their mothers. The IgG antibody, which moves through placenta, makes them immune against microbes from which the mother is immune. Also, children receive IgA antibodies from breast milk.

2.2. Immune System Properties

The immune system has four main properties: detection, diversity, learning and tolerance.

• Detection – detection (and recognition) occurs in an immune system when
chemical bonds between pathogen fragments and receptors on the surface of the lymphocyte.

- **Diversity** – Detection in the immune system is related with non-self elements of the organism. Thus, to ensure that at least some of lymphocytes will react to the pathogenic element, the immune system must have different types of receptors. The solution adopted by the body relies on a dynamic protection, via a continuous renewal of lymphocytes.

- **Learning** – The immune system must be capable of detecting the pathogen and eliminate them, as fast as possible. It includes principles which allows lymphocytes to learn and adapt themselves to specific foreigner protein structures, and to remember these structures, when needed as soon as possible. These principles are implemented by the B-cells.

- **Tolerance** – The molecules that mark a cell as self-gene are contained in the chromosome sections also known as Major Histocompatibility Complex (MHC). There are two types of MHC.
  - **MHC Class I**: antibodies are self characteristics that are present in every cell of an organism;
  - **MHC Class II**: antibodies are found in B-cells, phagocytes and other cells that are responsible for the antigens that aid T-cells.

### 2.3. Immune System vs. Computer System

There is a strong correlation between an immune system and a computer system. The immune system protects the body from pathogens elements in the same way the computer security system protects the computer from malicious users. In this section, we will show how the immune system paradigm can be adapted to the security problem. An adequate security system management policy has long been an important issue. The security plan must encompass all of the elements that make up the network, and provide important services [1,2,5,6] which include: (1) **Confidentiality**: Data access must only be allowed to authorized users; (2) **Integrity**: Data has to be protected from intentional or accidental corruption; (3) **Availability**: Information, such as the computer, must be accessible when necessary and as desired; (4) **Correction**: False alarms from an incorrect classification of computational events must be minimized; and (5) **Accountability**: the security system must be configured to preserve sufficient information from the intrusion that can be permitted to trace the origin of the attack.

The security policy used by the immune system is specified mainly by a natural selection phenomena, and only few computational aspects of security are of great importance, such as: (1) **Disposability**: this aspect allows the body to continue working even under attack of a pathogen; (2) **Correction**: this mechanism prevents the immune system against attacks of the (body) cells; (3) **Integrity**: this is a way to guarantee that the genetic codes, present in the cells, will not be corrupted by any pathogen; (4) **Accountability**: these are means adopted by the immune system to identify, find and destroy the pathological agents; and (5) **Confidentiality**: This is not a very important aspect to the immune system. In the human body there are no concepts of secret data or confidential information.

As we can see, both immune and computer systems share a common security concerns. They are both intended to protect their corresponding system...
against attacks and intrusions that violate the (established) security policy.

2.4. Intrusion Detection

An intrusion can be defined as “an act of a person of proxy attempting to break into or misuse your system in violation of an established policy”. Intrusion can be classified as: (i) misuse intrusions, i.e., well defined attacks against known system vulnerabilities; and (ii) anomaly intrusions, i.e., activities based on a deviation from normal system usage patterns. Intrusion detection systems (IDS) are one of the latest security tools in the battle against these attacks [10,13].

3. Operational Environment

Earlier intrusion detection systems (IDS) focused on the log registration analysis, which are produced by the operational system and other applications, as cited in [3,9,10]. UNIX operating system uses syslog to generate, store and register its activities. The amount of information contained in these registrations varies, and are directly related to the category of services offered by the host and the number of system accesses. Many of these messages can be of no importance, which unfortunately can hide the visualization of events that might be of great importance [1].

In order to monitor the execution of a program, we choose to analyze the log registration file originated from the execution of the program. Presently, we will adopt a variation from the architecture presented in [12].

In the artificial immune system we consider that each process in a host corresponds to a cell. At a given time, it is possible to have many different processes in program execution. Thus, we say that the host is observed at a multi cellular organism. In this host, there is a process that acts as a lymphocyte in the recognition process of nonself elements. The immune system defense architecture is multileveled and provides protection mechanisms at most of the levels [4,8]. The most external levels are represented by a control and monitoring mechanisms that act like the skin and mucous surfaces, thereby establishing a physical surface. The traditional security mechanisms such as archive system, where archive permits and control accesses, are considered as the innate immune system. The acquired immune system is implemented by a program that acts as lymphocyte, and by the patches as upgrades that are considered as vaccines. Similar to the human body, when a process in the artificial organism has its function changed, this organism is ill.

3.1. Artificial Immune System Anatomy

Similar to the human system, the computational immune system is divided into innate and acquired immune system. The innate computational immune system is a component from UNIX’s root filesystem, much alike the file permission and access modes to archives that are established during the system installation. Adopting an immune analogy, these security mechanisms are IgG antibodies. Just after the operating system is installed, it can be compared to a new-born child, which possesses a weak immune system. The operating system responds only to the violations and weaknesses that the software designers were aware of during the period of release of that software version. By analogy, we can consider the
corrections in a given version of the operating system that is represented by the release as IgA antibodies. Passive immunity in modern computational system lasts for only a few hours, unlike the human immune system that protects the new-born for a few months. For this reason it is necessary that the computational and operating system receives a complementary immunization just after it installation. Complementary immunization such as installation of complementary security and software services, and patches can be made by the administrator. With a complementary immunization, the artificial organism begins the assembly of the acquired immune system. The degree and duration of immunity in a computational system will depend on certain items that characterize the security.

The approaches adopted to restrict the access to some services were shifted to use the following tools: TCP WRAPPER [14], responsible for filtering the access to the programs, provides the same function of immunoglobulins. The antibody IgA is responsible for patrolling the entrances of the organism. The entrances of the organism are represented by TCP and UDP ports, which are used by the services offered by the host.

The largest property of the immune system human is the capacity to distinguish the self-elements from the nonself ones. Each cell of the human body contains MHC chromosomes, which identify them to the immune system as self. These characteristics are provided to the computational and immune system via the LogCheck [15], a software that plays similar role of the T-cells. The LogCheck software filters the registrations of the activities within a host, and generates reports to the administrator of the logs of the activities. The distinction between an activity that corresponds to self activity (or element) and that is nonself one is based upon the following four files:

**logcheck.hacking** – This file is composed of keywords that characterize an MHC antigen class I, and its occurrence unchains the immune answering system.

**logcheck.violations** - This file contains keywords that characterizes an MHC antigen class II, and the code portion exposes antigen.

**logcheck.violations.ignore** - In this file they recognize keywords that are the reverse of the existent words in the file logcheck.violations

**logcheck.ignore** – This file contains keywords which does not represent a violation of host’s safety.

### 4. Experimental Results

In our implementation, the software LogCheck[15] was installed in two hosts with different operating systems and registration services. The chosen hosts were the server of two different departments: Center of Physical and Mathematical Sciences (CFM) and the server of the Center of Communication and Expression (CCE). The hosts possess the following characteristics: The machine server of CFM is a machine Sun with the Solaris OS. The machine server of CCE is a PowerPC IBM, using AIX operating system. Both servers provide the following services: DNS, SMTP, POP3, FTP.

The activities in these servers were monitored for five consecutive months, and the operational ambient didn't suffer any interference in the characteristics of its basic activities. Thus, the registrations can be considered trustworthy, because they reflect a normal situation. The amount of information contained in this group is quite
large, and thereby making its analysis very difficult.

In Figure 1, we present the results we have obtained for the CFM server machine. As we can see, our results indicate that our system helps to reduce significantly the number of the registration activities.

In Figure 2, we portray the results we have obtained for CCE center. As we can see a significant reduction of the number of registrations is obtained. We can observe in Figure 2, that we obtained about 66% of reduction in the month of June, for instance.

As we can see, in both monitored servers, CFM and CCE, a significant reduction in the number of user registrations during the five months of our experiments has been obtained. This will have a profound impact in reducing the size of the reports that can be used to detect malicious intruders, and thereby speeding up the detection of such intrusions. However, during the period when we monitor the system and collect the registration information, we have obtained only the reports that correspond to the violations of safety and the unusual events. The registrations belonging to this group also belong to the activities group of type nonself. The administrator shall receive a notification, in the form of a mail, with the report of the activities of type nonself genes, thereby minimizing his security management work (i.e., intrusion detection). During the period of monitoring and user activity registrations collection, only the security violations and unusual events were detected. The data are represented in Figures 3 and 4. They show the security violations and the unusual events that have occurred for both CFM and CCE machines and during the period March to July.

Our results indicate clearly that it is possible to monitor and observe the process the hosts’ activities in an efficient and less costly way.

5 Conclusion

This paper describes a new approach to secure monitoring computer systems. The main idea is to specify what services are critical in such systems, and monitor their utilization while searching for anomalies.
This is done using of the Unix activities registration (syslog) and the proprieties of the human immune system.

Although many intrusion detection systems have been developed by now, to the best of our knowledge very little work was done for a Unix Operating System environment using bio-inspired techniques such as human immune system. In this paper, we discuss an efficient intrusion detection model based upon the human immune system, and discuss the 1-1 correspondence between the basic characteristics of Unix operating system and three parts from the human immune system.

Our model is based upon the intrusion detection by anomaly. The hosts monitor the data set (registration activities: attacks, security violations and security events). In our experiments, we focused upon the security violations and securing the events. These ones were detected by finding out either a non typical action or by violating the (computer system) security policies.

Our results indicate clearly that our system was very successful in reducing by 50% the size of the reports the administrators use to identify potential intruders. This system will have a profound impact in detecting intruders (equivalent of diseases in the human system), and thereby reacting to these intrusions in a very affective and efficient way.

In the future, we plan to expand and improve our system. First, we will develop efficient cryptography mechanisms that can be added to improve the efficiency of the communication between the administrator (host) and the rest of the machines available. Second, we shall protect the registration activities in both original and final forms - thereby allowing further data analysis. Finally, we plan to extend our tool to intrusion detection by abuse, i.e., recognizing the traces of the intrusions, and thereby appropriate actions can be taken against these intruders.
References


