Introducing Biological Mechanisms To Computer Security Students

Qinghai Gao
Department of Criminal Justice & Security Systems
State University of NewYork at Farmingdale
2350 Broadhollow Road, Farmingdale, NY 11735
Email: GaoQJ@farmingdale.edu

DR. QINGHAI GAO

Dr. Qinghai Gao is an Assistant Professor in the Department of Criminal Justice & Security Systems at Farmingdale State College/SUNY. Before joining Farmingdale, he taught in China University of Petroleum from 1992 to 1998. From 1998 to 2007 he taught as Adjunct in Brooklyn College, Lehman College, NYC College of Technology, College of Staten Island, and York College. Since 2001 he held positions in industry as Software Developer, Database Administrator, Network Engineer, Researcher, Consultant, and Information Security Specialist.
Dr. Gao received a PhD in computer science from the City University of New York in 2007. So far he has published one book and >20 articles. His present research interests include Digital Forensics, Network Security, Biometrics, Biological Information System, Forensic DNA Analysis, Cryptography and Steganography.

Introducing biological mechanisms to computer security students

Abstract:

Biology has broad impact on computer security. Many computer security approaches to defense originate from the observation of biological phenomena. However, students majoring in computer security are often lack of knowledge in these biological mechanisms and thus have difficulties in correlating the two seemingly unrelated fields. In this paper we introduce a few biological security mechanisms, such as the cryptographic and steganographic processing of biological information expressed in DNA, RNA, and protein sequences, the defensive actions against invading pathogens conducted by biological immune system, and the improved survivability of a species through bio-diversity, to bring students to a higher level in understanding biologically inspired security mechanisms and in drawing new security paradigms from the most recent developments in biology. Not only do we ask students in computer security to carry out comparative study of biological mechanisms and computer security algorithms, but also encourage faculty to develop new curriculum that can facilitate this objective. Due to the interdisciplinary nature of the subject, it is necessary to involve faculty from both computer science and biology to train students with the knowledge that are transferrable across professional boundaries.

1. Introduction

Biology has broad impact on computer technology. In particular, biological phenomena have been a rich source of inspiration for computer security professionals. Famous computer scientist Seymour Cray [1] once described a biological cell as a computer system consisting of the following components: several thousand microprocessors → ribosomes (RNA); Dynamic Random Access Memory (RAM) → DNA; program code organized into ~150,000 subroutines (genes); power supply →mitochondria.

Many concepts widely used in computer security such as virus, worm, and Trojan horse are borrowed from biology. Many emerging computer security techniques are invented as a result of observing the biological world.

However, students learning computer security often have difficulty understanding these borrowed biological mechanisms, which negatively affect their critical thinking skill and creativity. The main reason for the problem is due to their lack of knowledge in biological systems. One possible solution is to comparatively study biological mechanisms and their applications in computer security. Unfortunately, existing curricula for computer security and its related programs do not provide students with the necessary training.

In this paper we introduce a few biological defense mechanisms as an effort to stimulate interests in biology among computer security students.
Due to the interdisciplinary nature of the subject, it is necessary to involve faculty from both computer security and biology to co-develop new courses to train students.
As technology advances, using computer becomes an indispensible part of life for all intellectuals. It is necessary for all college graduates to have some common knowledge about

computer security to facilitate their daily routine and to protect their online privacy. Therefore, students majoring in biology can also benefit from these new courses.
The rest of the paper is organized as the following. Section 2 introduces intracellular defense mechanisms, focusing on natural cryptographic and steganographic processing of genetic information. In Section 3, we introduce intercellular defense – defense against invading pathogens by biological immune system. Section 4 discusses species defense – security through diversity. Section 5 summarizes the paper and proposes future work.

2. Intracellular defense - natural cryptographic and steganographic processing of genetic information

The Central Dogma of Biology mainly includes two biological sequence mappings. One is called *transcription* from DNA to RNA and the other is called *translation* from RNA to protein. In biology, the protein making process in most species typically contains the following three steps [2]:

S1: *Transcription* from DNA to pre-RNA
S2: *Splicing* of pre-mRNA to mRNA
S3: *Translation* of mRNA to protein

Fig. 1 shows the steps. In it the **pre-RNA** is the intermediate sequence in which the introns, labeled as I1, I2, I3 and I4, will be removed and the exons, labeled as E1, E2 and E3, will be ligated to form mRNA, the final template sequence for protein synthesis.

However, biologists also find a different path of making protein in some other species, as given in Fig. 2. It mainly consists of the following three steps:

S1: *Transcription* from DNA to RNA
S2': *Translation* of RNA to pre-Protein
S3': *Splicing* of pre-Protein to protein

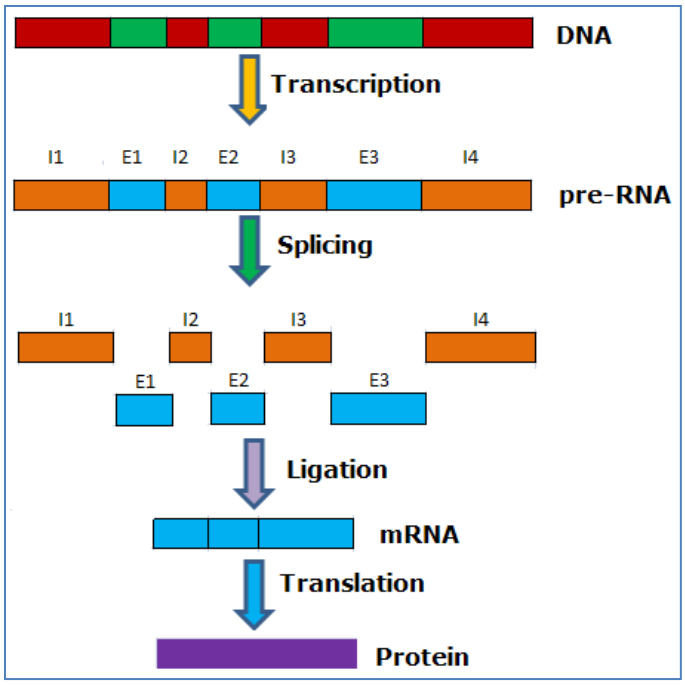

Fig. 1 Biological information flow with **pre-RNA** as the intermediate sequence

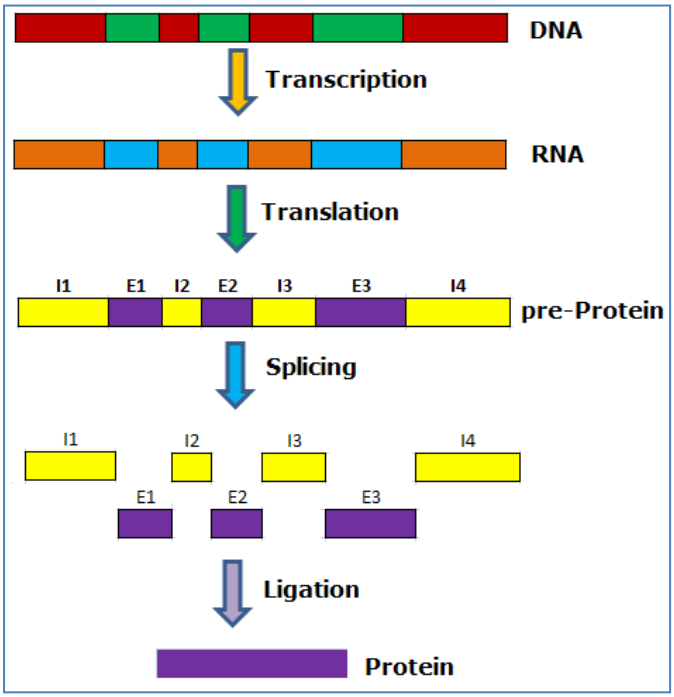

Fig. 2 Biological information flow with **pre-Protein** as the intermediate sequence

By comparing Fig. 1 and Fig. 2 we can see that the transcription steps of making protein in different species are exactly the same. The difference comes from the swapping of the translation step and the splicing step.

Viewed from the perspective of information security, the sequence mapping in the transcription step does not increase security due to the 1-to-1 correspondence between the chemical bases of

DNA and those of RNA. However, both the splicing step and the translation step transform the biological information in security-enhanced manner.

●Enhance security with splicing process

In Fig. 1, the pre-RNA sequence contains introns (I1, I2, I3, and I4) and exons (E1, E2, and E3). However, for a given pre-RNA sequence the number of introns, the number of exons, and the boundaries between introns and exons are non-deterministic. That is to say, the same pre-RNA sequence can be spliced into different mRNA sequences under different circumstances. Thus different proteins can be produced.

Similarly in Fig. 2, the pre-Protein sequence contains inteins (I1, I2, I3, and I4) and exteins (E1, E2, and E3). However, for a given pre-Protein sequence the number of inteins, the number of exteins, and the boundaries between inteins and exteins are non-deterministic. One pre-Protein sequence can be spliced into different final protein sequence.

In information security, given a sequence containing dummy letters (stegotext), it is a steganographic problem on how to find and remove them to recover the original plaintext.

●Enhance security with translation process

In computer security, the high-level language source code can be compared to the DNA sequence; the intermediate assembly language code can be considered as the RNA sequence; and the binary machine language code can be thought of as the protein sequence. For security purpose the binary/machine language code can be reversely engineered to its source code with a decompiler.

In biology, RNA is a base 4 system and protein is a base 20 system. The mapping from RNA to protein is a substitution cipher, in which every three consecutive symbols from RNA alphabet will be replaced with one symbol from the protein alphabet. With direct translation, one RNA sequence can only map to a unique protein sequence.

In the natural genetic code[2], the 64 codons (triplets of 4 letter A, U, G, C, 4x4x4=64) codes for 20 amino acids. On average, the ratio of the number of RNA codons to the number of amino acid is about 3:1 (64:20). The number of possible RNA sequences for a given protein sequence of length n will be up-bounded by 3^n. Therefore, it is difficult to reversely translate a protein sequence back into its corresponding RNA sequence.

3. Intercellular defense – defense against invading pathogens by biological immune system

According to Dasgupta [3], "*The biological immune system is a complex adaptive system that has evolved in vertebrates to protect them from invading pathogens. To accomplish its tasks, the immune system has evolved sophisticated pattern recognition and response mechanisms following various differential pathways, i.e. depending on the type of enemy, the way it enters the body and the damage it causes, the immune system uses various response mechanisms either to destroy the invader or to neutralize its effects*."

Biological immune system has three defensive mechanisms to eliminate invading pathogens: Clonal Selection, Negative Selection, and Immune Network [3].

●Clonal Selection

In general, an organism has no predetermined way to know what kind of pathogen would invade its body so that it can produce antigens to defend against the invasion. Therefore, one cell of an immune system in an organism will produce a number of daughter cells. However, unlike most somatic cells these daughter cells are not the exact copy of the mother cell. Instead, they are the randomly mutated version of the mother cell. Upon the invading of pathogen, only these daughter cells that are effective at eliminating the pathogen will multiply at a high rate. The other daughter cells that do not contribute the elimination process will die out.

The clonal selection mechanism has many applications in computer security, such as antivirus software design and network intrusion detection. In particular it provides an effective way to detect and remove new species of computer virus and network threats.

●Negative selection

As mentioned in the previous section, an organism produces random cells to defend against pathogens. However, these randomly generated cells could be harmful to the survival of its own functional cells. Fortunately, the immune system has evolved a mechanism to distinguish between pathogens and self cells – negative selection. Fig. 3 illustrates the processes of negative selection.

Similar to clonal selection, negative selection mechanism has been applied in computer virus detection and network intrusion detection.

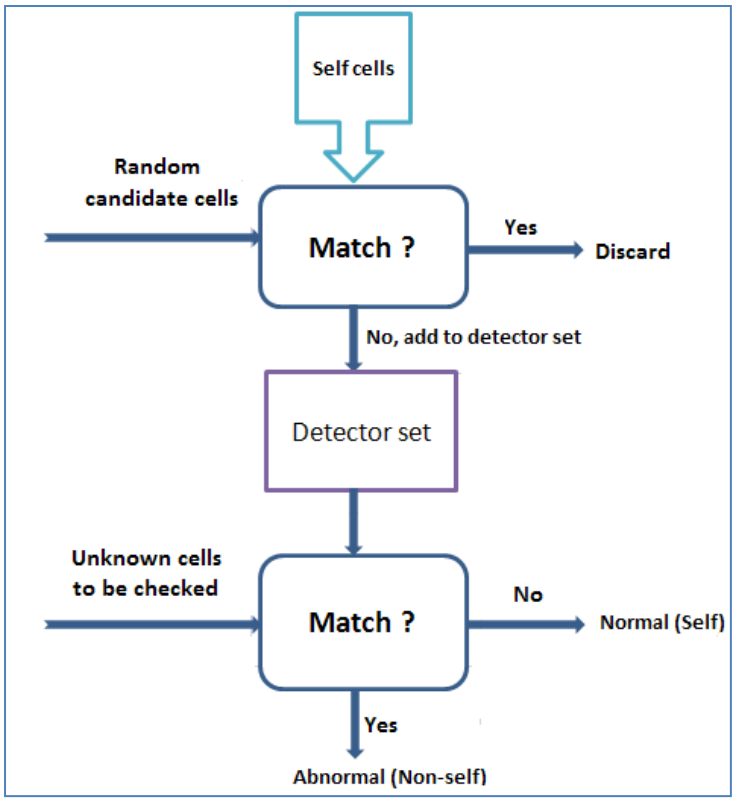

Fig. 3 Negative selection mechanism (Adopted from [3])

●Immune network

It is believed that the immune cells maintain an interconnected network for pathogen recognition and response. And the idea has been adopted in computer security for designing distributed antivirus and intrusion detection network.

4. Species defense - Improved survivability through bio-diversity

In nature, it has been observed that the outbreak of one strain of deadly and fast-spreading viruses or bacteria can decimate or wipe out an entire species. And very often the culprit virus or bacteria only harm a specific species and have no significant effects on other non-closely related species.

In computer security malwares remain the major threats to many systems. However, it is found that one computer virus or worm can only be harmful to one specific type of operating system, such as Windows. The same virus or worm could not effectively cause damage to a Linux or Mac machine because they have different "genetic codes".

Two decades ago, Forrest et al. [4] proposed idea of improving security through introducing diversity for computer operating systems. Although the idea has been applied in certain software design, security through diversity is still an active and promising research topic.

5. Summary and future work

In this paper we proposed the idea of introducing biological phenomena to students learning computer security. We looked at a few biological defense mechanisms, including intracellular defense, intercellular defense, and species defense, and briefly surveyed how these mechanisms have been applied in some way in the field of computer security.

Recently, there are many security-related new developments in biology that have yet to be explored by computer security professionals. It is beneficial for students majoring in computer security to understand these developments so that they may bring new ideas and security paradigms to the field of computer security. Not only do we ask students in computer security to carry out comparative study of biological mechanisms and computer security algorithms, but also welcome faculty to develop new curriculum that can facilitate this objective. Due to the interdisciplinary nature of the subject, it is necessary to involve faculty from computer security and faculty from biology to break the traditional professional boundaries.

Reference

[1] Cray, S. (1996). "An imaginary tour of a biological computer (why computer professionals and molecular biologists should start collaborating)", Remarks of Seymour Cray to the Shannon Center for Advanced Studies, University of Virginia. Retrieved on September 23, 2011 from http://www.cccp2000.com/cray.html

[2] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., &Walter, P. (2007). *Molecular Biology of the Cell* (5th edition), Garland Science.

[3] Dasgupta, D. (2006). Advances in artificial immune systems. *IEEE Computational Intelligence Magazine*, 1(4): 40-49.

[4] Forrest, S., Somayaji, A., & Ackley, D. (1997). Building Diverse Computer Systems. *Workshop on Hot Topics in Operating Systems*, pp. 67–72.