Cryptography and Information Protection in the Living World

Naya Nagy\textsuperscript{1}, Marius Nagy\textsuperscript{1}, and Paul Hodor\textsuperscript{2}

\textsuperscript{1} College of Computer Engineering and Science
Prince Mohammad Bin Fahd University, Al Khobar, KSA
\{mnagy, mnagy\}@pmu.edu.sa
\textsuperscript{2} Booz Allen Hamilton, Rockville, MD, USA
hodor_paul@bah.com

Abstract. This paper explores parallels between concepts defined in cryptography and concepts of biology at different levels of organization. Cryptographic settings, including the presence of an eavesdropper are extensive in the realm of plants and animals. It also turns out that principles of information protection show strong similarities between the two disciplines: computer science and molecular biology. Biological information, as held by the DNA molecule, and digital information, as used in digital communication systems, are subject to analogous procedures of protection and repair when damaged.

Keywords: information protection, error correction, cryptography, mimicry, DNA, DNA repair

1 Introduction

Cryptography is a field that spreads human activities. The need to communicate privately, or secretly, enters various corners of human private and social life, such as financial transactions, personal privacy of communication, company secrets, information protection at the level of a country, a state, a group, or an organization. In fact, the possibility to communicate privately with another human is considered to be an individual freedom. It has the flavor of a human right.

The idea behind this paper is that the need for secret communication, or more generally, the existence of cryptographic needs is not inherently pertaining to humans. Cryptographic settings and cryptographic solutions can be encountered throughout the living world. The point of view may be that of an information-carrying molecule, a cell, an entire organism, a population, or an ecosystem. This paper explores various scenarios in which encryption/decryption, and cryptographic identities are part of vital processes.

2 The Players in Cryptography and Their Interests

We may consider the classic model of a cryptographic setting to be sufficient for the biological realities to be discussed here. This model involves three entities:
two communication partners and a third malevolent party that intends to corrupt
the communication. The corruption of the communication refers to its privacy
and the reliability and truthfulness of the exchanged messages.

2.1 The Good Players

By standard now, it is Alice and Bob that intend to communicate secretly and
reliably. Alice and Bob are usually equivalent partners. The communication is
symmetric and as such, Alice’s and Bob’s points of view are identical. For the
communication to comply to secrecy and reliability, Alice’s expectations are the
following:

1. Alice wants to get all messages from Bob. This means that the connection
   between Alice and Bob should be permanently working. Or else, if the con-
   nection is broken, both Alice and Bob should be aware of it. Any message
   sent by Bob should reach Alice.
2. Any message Alice gets from Bob is indeed sent by Bob. This is called
   authentication. Bob’s message may carry Bob’s unique signature. For log-
   ical completeness, any message that is not coming from Bob, is known to
   have another sender. That is, Alice recognizes the message to have a foreign
   sender.
3. For any message that Alice sends to Bob, Alice knows whether Bob has
   received the message. This is a handshake.
4. When Alice receives a message from Bob, the content of Bob’s message is
   complete and unaltered. No parts of the message were lost, no meaning has
   been altered or twisted.
5. Bob’s message is understandable to Alice. That is, they speak the same
   language, use the same alphabet, semantics, and syntax. Additionally, any
   new concept that Bob may start using, should first be defined to Alice, before
   its usage.

These expectations have been formulated for Alice. As Bob is an equivalent
entity, all of the above items apply symmetrically to Bob.

2.2 The Bad Player

The third party, called Eve, makes every effort to meddle, interrupt, or attack
the privacy of Alice’s and Bob’s communication. Thus any form of corrupting
the transfer of messages is in the domain of Eve. Eve may listen to the communi-
cation, or break the connection, or may have any other destructive behavior. The
interest of Eve may vary, depending on the practical setting and goals. Some of
Eve’s possible attacks may not be compatible with one other. For example, if Eve
chooses to interrupt a conversation by severing the connection, this means she
cannot gain any knowledge on what Alice and Bob would have communicated
to each other. It means she obviously cannot eavesdrop on the conversation.

Consider the following list of attacks that Eve may plan on the communica-

1. **Eavesdropping.** Eve may listen to the communication channel and read the encrypted messages.

2. **Tampering.** Eve may tamper with the content of a message. For example, if the message is a string of characters, Eve deletes a substring from the message and/or inserts a substring of her own into the string of the message.

3. **Inserting.** Eve may insert a false message. Eve may send a false message to Alice or Bob.

4. **Intercepting.** Eve may intercept a message sent from Bob to Alice and drop it.

5. **Masquerading.** Eve may masquerade as Bob and send messages to Alice pretending she was Bob. This is in the realm of identity theft.

6. **Disconnecting.** Eve may completely sever the connection between Alice and Bob so that no further messages can be transmitted.

The question lends itself to where we can find Eve in Biology. There is an interesting aspect to the parallel of Eve, as a cryptographic entity, and its counter character in Biology. We can find Eve at every level of biological scrutiny, that is to say, both at sub-cellular, as well as cellular and multicellular levels within the hierarchy of life.

The next section explores a few encounters of Eve in nature, as the above cryptographic identity. They show the range and diversity of parallels that can be drawn between cryptography and biology.

### 3 The presence of Eve in nature

#### 3.1 Eve at high levels of biological organization

Let us first explore cryptographic needs during the interaction of organisms such as animals or plants. Organisms typically communicate through visual, acoustic, or chemical signals. We think that many of Eve’s actions as described in subsection 2.2 can be found in interfering with a variety of modes of communication.

**Eavesdropping.** Eavesdropping is pervasive among animal predators. A predator, such as a large cat, stalks its prey before bounding on it. Stalking implies listening, or studying the prey’s behavior, and also trying to conceal the presence of the predator. Analogously, Eve in cryptography listens to the communication channel between Alice and Bob and endeavors to hide her action.

**Masquerading.** Masquerading was described as the attempt of Eve to present herself as Bob. Eve sends messages to Alice pretending that she is Bob. Eve achieves this by forfeiting Bob’s signature on a message, or more generally, Eve exhibits Bob’s characteristics.

Many animal and plant species have evolved to take the visual aspect of another species or of an inanimate object, in order to gain some advantage over their predators or prey. Such phenomena of deceit are called *mimicry* in biology[15], and were first described over a century ago by Henry Walter Bates [1] as he studied butterflies in the Amazon forest.
The biological concept of mimicry is very large and encompasses different types of signaling and behavior. In *aggressive mimicry*, a predator aims to hide under the characteristics of a harmless species or object. It is the "wolf wearing a sheep's skin". Consider the North American *Photuris* firefly [15]. The *Photuris* female attracts males of another firefly genus, namely *Photinus*, by emitting light flash patterns that mimic those emitted by *Photinus* females. When a *Photinus* male mistakenly approaches a *Photuris* female, he is eaten. Thus, *Photuris* females mimic the behavior of another genus, by sending out wrong signals.

*Defensive mimicry* aims to protect a species against its predator, while masquerading as a dangerous or unpalatable species. There are many cases of defensive mimicry. For example, there is a snake species called the false cobra, *Malpolon moilensis*. Its venom is mild compared to the Indian cobra, *Naja naja*. Nevertheless, the false cobra has a similar hood to threaten with. The duped enemy usually backs off at the false threat.

**Inserting a false message.** In this case Eve sends a false message to Alice. A strategy used by some birds, fish, or insects is *brood parasitism*. In this behavior individuals of a host species are manipulated to raise the young of another, called the brood parasite. For instance, females of the North American brown-headed cowbird *Molothrus ater* lay their eggs into the nests of a large number of other species. The parasitic young compete with their foster siblings for parental care. By begging for food more intensely and loudly [6], the parasites have an advantage in attracting the attention of the parents.

### 3.2 Eve at the molecular level

When we think of Eve in terms of cryptography, Eve has the full characteristics of a person. She has her own will, has intentional actions, is intelligent, cunning and shrewd.

Accidental damage of the communication channel between Alice and Bob is possible and has to be considered. Nevertheless, the treatment of accidental failures of the communication is rather a problem of technical reliability in a possibly adverse physical environment, not so much a problem of security. It is rather the "human" characteristics of Eve that bring us into the realm of cryptography. The enemy is an enemy that thinks and acts based on her will. Also, Eve understands the unencrypted content of a message as a human would do ... "Eve knows English".

When we deal with biological entities, especially at the suborganismal and subcellular levels, it is rather a stretching of the mind to consider an enemy with the proper characteristics of evil intentions and intelligence. To be able to keep the same setting as we are accustomed to in cryptography, the person of Eve has to be understood in a larger context. Eve would become a dummy person, responsible of any destructive action on the integrity and life of a cell. In this context, such attacks would not have a real intention behind them, but may be defined as attacks with a physical or chemical, or even biological cause. Such attacks can be repetitive, forcing the cell to develop methods to protect itself from them.
The most important message and information carrying molecules found in cells is DNA, which constitute the genome of an organism. Chemically, DNA is a linear polymer consisting of a phosphate-sugar backbone. To each sugar one of four nitrogenous bases, adenine, thymine, cytosine, and guanine, is attached. The linear sequence of bases within a DNA molecule allows for an enormous number of possible combinations. The sequence itself encodes the information carried by the molecule, and constitutes the genetic message. Within a cell, DNA is present as a double helix, consisting of two complementary strands, which means that if we know the sequence of one strand, we can deduce the sequence of the other. The information stored in DNA is used by cells in two ways. First, through a process called transcription, selected short stretches of sequence are copied into RNA, another type of information carrying molecule. RNA transcripts then go on to support all cellular functions. Second, before a cell can divide, each DNA molecule undergoes replication, by which two identical copies are created. The two copies are distributed to the two daughter cells during cell division.

Damage to DNA may be physical or chemical, i.e. lesions produced by endogenous and exogenous agents. Endogenous agents have the source inside the cell or organism, whereas exogenous agents originate in the environment. Even if the damage is physical or chemical in nature, it has biological consequences [5]. It affects the health of the cell and consequently of the organism and may fully inhibit the replication process of the DNA molecule. Some common physical agents are ultraviolet light and ionizing radiation (e.g. X-rays and gamma rays). Chemical agents that affect DNA sequence integrity are called mutagens. They have diverse modes of action, and many cause direct damage to DNA through specific chemical reactions. Others, however, interfere with replication, causing errors during copying of the sequence into newly synthesized DNA.

A remarkable class of agents that affect DNA integrity act by altering the sequence itself, without producing structural damage to the DNA molecule. Such agents reside at the boundary between living and nonliving matter, as they exhibit some, but not all characteristics of living organisms.

One such type of agent is represented by transposons, also called transposable elements or "jumping genes" [12]. Transposons are relatively short sequence segments found in the DNA of many species. They are mobile, in the sense that they can insert themselves into a new location within the DNA molecule through a mechanism of either "cut and paste" or "copy and paste". When a transposon moves, it alters the sequence of the DNA at the old and/or new location, which possibly has consequences on biological function and is a threat to genome integrity.

Certain viruses are sequence-altering agents as well. Viruses are biological entities that have a defined structural organization, have their own DNA or RNA, and are able to reproduce. However they are not composed of cells, and do not support their own metabolism, and thus lack some key characteristics of living things. Some types of viruses, such as the bacteriophage λ [4], or the HIV virus [11], insert their own sequence into the DNA of host cells. Through this
process they are able to alter the function of the host cell, hijacking the cellular machinery for their own purpose.

4 Electronic and Molecular Information Safety

The present section is dedicated to the actions that Alice and Bob may take to ensure or at least improve the reliability of their communication. Problems of protecting the message contents, correcting errors and mistakes, recovering as much as possible from the initial message are inherent to communication processes. The following is a non-exhaustive study of reliability or safety issues. The purpose is to parallel the two information holders: electronic and biological, in view of the fact that the problems faced by them are similar.

4.1 Protected Public Information

The first action that would be considered here is to protect the information from attempted change. Protecting information is rarely discussed explicitly in cryptography, but is nevertheless presupposed in several instances, such as public key cryptosystems.

Most commercially successful cryptosystems rely on public key cryptosystems, such as the Rivest-Shamir-Adleman (RSA) protocol [13]. Acceptable security levels are reached using “one-way” functions, functions that are easy to compute but difficult to invert. Such a system needs two keys: a public key and a private key. Bob encrypts a message with Alice’s public key, and then sends the message to Alice. Alice decrypts the message using her private key. Note that, the public key, as the name suggests, is visible and known to everyone, including Eve. Nevertheless, the message is unintelligible to anyone unless it is decrypted with the private key, which is known only to Alice. In order for the protocol to work, the public key is guaranteed to be protected, unchangeable. There is a consensus about the public key value. Eve is not allowed to change the public key value, or else Bob may not correctly encrypt the message. It is a strong requirement in public key cryptosystems, that public information can be protected from interference. This may not seem theoretically so obvious but works acceptably in practice. For example, if many copies of Alice’s public key exist, such as on the internet, and in several other public multi-media of a large audience, it can be assumed that Eve cannot control all public channels of communication.

Similar mechanisms of protecting information can be found at the molecular level, concerning the DNA molecule. The structure of the molecule itself has properties that offer protection in a possibly adverse chemical environment. In the DNA double helix, the phosphate sugar backbone faces the outside, while the information-carrying bases are hidden inside. Thus the bases themselves are protected from chemical attack [3].

Considering the idea that public information can be protected by keeping it in many copies, cells and organisms have several methods to provide the same kind of redundancy. Bacterial cells often contain several replicas of their
DNA. In organisms where DNA is organized into multiple chromosomes and packaged into a nucleus, chromosomes typically exist in pairs in each cell, such that there are two copies of each DNA sequence available at all times. If one of the copies is altered and loses a certain gene function, the other copy can often times compensate. Tissues such as skin or bone contain large numbers of cells, thousands to millions, and thus the same number of copies of the full DNA complement of a single cell. If the information in any single cell is damaged, the cell is destroyed by the immune system and replaced through cell division from a normal cell.

A way of protecting information is to keep it in different formats, or on different hardware supports. Backups are very usual for humanly manipulated information. An interesting analogy is represented by the two strands that make up the DNA double helix. They contain the same information, but in two complementary formats. At any particular location along the DNA, only one of the strands is biologically functional. The other serves as the "backup". If either of the strands is damaged, it can be reconstructed from the complementary strand. DNA repair mechanisms exist for various situations [10].

Another strategy for protecting information is to identify and "quarantine" changes to the original message. Cells have mechanisms by which they can inactivate defined regions of their DNA. Inserted foreign sequences, such as transposons or viral DNA, can be identified and silenced. One such mechanism found in animals and plants involves a chemical modification, i.e. methylation, of cytosine in the DNA molecule. Transposon sequences are specifically identified and methylated, thus preventing them to jump and insert themselves into new locations [9].

4.2 Error Correcting or Repair Mechanisms

If binary information is transmitted over an unreliable channel, the message may reach its destination in a corrupted form. If the message has been partially altered by faulty transmission, the correct message needs to be reconstructed. The field of error correcting codes [7] aims to develop encoding techniques that allow for errors to happen during transmission, while preserving the full content of the message. Shannon proved [14] that at a rate below the capacity of the communication channel, the message can be sent with arbitrarily high accuracy. Probably the best known error code is the Huffman code. Error codes protect the message in that some \( n \) bits of information are encoded into \( m \) (\( m > n \)) bits to be sent across the channel.

Similarly, DNA information needs to be corrected when damaged. Molecular mechanisms that deal with reconstructing the DNA molecule after a damage are called repair mechanisms [10]. Some repair mechanisms deal with specific damage types, others work more generally.

For example, ultraviolet light of type UV-C and UV-B produces a specific damage on DNA chains for which a specific repair mechanism exists [8]. The range of UV-C and UV-B radiation is from 180 to 320 nm, and includes the DNA absorption maximum at 260 nm. When two adjacent thymine bases absorb
a photon, they bond covalently forming a dimer. This results in a structural distortion of the DNA double helix that physically prevents replication. Through a process called photoreactivation, the covalent bond in the thymine dimer is reversed, and the DNA strand is directly repaired. Photoreactivation occurs in most organisms (but not in humans), and requires the action of the enzyme photolyase, which depends on the presence of light in the range of 313-475 nm.

Another example of a DNA repair mechanism is nucleotide excision repair [2]. Faults in the pairing of complementary strands are detected by the presence of distortions in the DNA structure. A stretch of one strand around a distorted area is cut off from the double helix. Subsequently, the enzyme DNA polymerase fills out the missing part according to its complementary strand. In the end, DNA ligase seals the nicks and thus completes the sugar-phosphate backbone of the repaired strand.

5 Conclusion

We have shown that some rules that apply to secure communication among humans are directly translatable to rules in biology, at different hierarchical levels of organization. As computer science and biology have developed separately, meaning in their own scientific community, we observe that the language that describes similar concepts, naturally differs from the computer science community to the biology community.

This paper explores basic cryptographic needs for molecular biology, such as protecting information and correcting errors. At the higher level of interacting organisms, the eavesdropper has a higher level personality, including the capability of masquerading.

With the advent of quantum explanations to biological processes, we may well look forward to find mechanisms of quantum cryptography imbued in the processes of life.

References