A REVIEW ON BIOLOGICAL INSPIRED COMPUTATION IN CRYPTOLOGY

Subariah Ibrahim\textsuperscript{1} \quad Mohd Aizaini Maarof\textsuperscript{2}

Department of Communication and Computer System  
Faculty of Computer Science and Information System  
Universiti Teknologi Malaysia, Skudai 81300, Johore, Malaysia

\textsuperscript{1}Tel: +60-07–557-6160 x 32386, Fax: +60-07–556-5044, E-mail: subariah@fsksm.utm.my  
\textsuperscript{2}Tel: +60-07–557-6160 x 32009, Fax: +60-07–556-5044, E-mail: maarofma@fsksm.utm.my

Abstract
Cryptology is a field that concerned with cryptography and cryptanalysis. Cryptography, which is a key technology in providing a secure transmission of information, is a study of designing strong cryptographic algorithms, while cryptanalysis is a study of breaking the cipher. Recently biological approaches provide inspiration in solving problems from various fields. This paper reviews major works in the application of biological inspired computational (BIC) paradigm in cryptology. The paper focuses on three BIC approaches, namely, genetic algorithm (GA), artificial neural network (ANN) and artificial immune system (AIS). The findings show that the research on applications of biological approaches in cryptology is minimal as compared to other fields. To date only ANN and GA have been used in cryptanalysis and design of cryptographic primitives and protocols. Based on similarities that AIS has with ANN and GA, this paper provides insights for potential application of AIS in cryptology for further research.

Keywords: Cryptography, Cryptanalysis, Artificial Neural Network, Genetic Algorithm and Artificial Immune System.

1. Introduction

Cryptography is a branch in cryptology that concerned with the construction of schemes that are resilient against malicious attacks. The schemes are designed in such a way as to ensure the security services of confidentiality, authenticity and integrity of messages. Another branch of cryptology is cryptanalysis that delves with the violation of security services to obtain messages. Major works in cryptology are based on number and information theory. The advances in software technology and systems will give more computational power for cryptanalyst to break the cipher. As new computational environment becomes more distributed, more diverse and more global, the transmission of information is becoming more vulnerable to adversary attacks. Thus making the design of cryptographic schemes that can counter new cryptanalysis techniques is becoming harder.

Biological systems can be regarded as sophisticated information processing systems and can provide inspiration for various ideas in engineering and technology. Modern artificial intelligence (AI) employs biological prototypes such as neural networks, genetic code and immune system for AI modeling. Artificial Neural Network (ANN), Genetic Algorithm (GA) and Artificial Immune System (AIS) are AI modeling for neural network, genetic code and immune system respectively. These AI models use the basic elements of the prototypes and their rules of interactions to define computational models
that can be applied to solve problems in various fields [1]. Over the past few decades there has been a growing interest in the use of biology as a source of inspiration for solving computational problems. This area of research is often referred to as Biologically Inspired Computing (BIC) [2].

Techniques taken from BIC, especially GA, ANN and AIS are steadily gaining ground in the area of cryptology and computer security. The goal of this paper is to review major works in the application of BIC in cryptology. Firstly, BIC computational models will be discussed, particularly GA, ANN and AIS. In the next section, the application of BIC in cryptology will be provided and the last section will conclude with some remarks.

2. Biological Inspired Computational Paradigm

Various aspects of biology have always been the inspiration in developing computational models and problem solving methods [3]. Primarily, the motivation of this field is to extract useful metaphors from natural biological systems, in order to create effective computational solutions to complex problems in a wide range of domain areas [2]. However, the actual models of computation used do not really correspond to what actually happens in biology, except at a fairly superficial level. BIC paradigm can be classified into: GA, ANN, AIS, DNA, Cellular Automata (CA) and ant colony. However, the focus of this paper is in GA, ANN and AIS. The more notable development are GA and ANN which has been applied to various fields whereas there have been relatively few applications in AIS [4]. The hybrid of GA, ANN and AIS can also be created in order to benefit the combining strengths of different paradigms [5][6].

2.1 Genetic Algorithm

GA mimics the evolutionary principles rooted in biological evolution inspired by neo-Darwinian theory [2]. Hence, the computational model is based on genetic and evolution. GA has been successfully applied to numerous applications in the field of search and optimization. The computer finds better and better solutions to a problem just as species evolve to better adapt to their environment [7]. GA was developed by Prof. J.Holland and his colleagues at the University of Michigan in 1965 [8]. It is an iterative procedure that consists of a constant-size population of individuals whereby the initial population is generated at random or heuristically. The population evolves by applying three basic operations, selection of solutions, mating of genes and occasional mutation [9]. These operations can be described as follows:

- Selection: Individuals are selected for survival and reproduction according to their fitness.
- Mating: New individuals are created by recombining and/or introducing genetic variation into the selected individuals.
- Mutation: Randomly perturbs a candidate solution for a better individual.

The basic cycle of GA is as shown in Figure 1. The cycle continues until a certain number of generations whereby the surviving individuals represent a solution to the problem (10).
2.2 **Artificial Neural Network**

ANN is a computational model of the brain. The model consists of a number of highly interconnected processors called neurons, which are analogous to the biological neurons in the brain. The neurons are connected by weighted links passing signals from one neuron to another. Each neuron receives several input signals through its connections and produces one output signal [11]. Figure 2 depicts a typical neuron. The manner the neurons are connected determines the flow of information in the network and defines the network model. Amongst the different network models are Multi-Layer Perceptron, Hopfield Networks and Kohonen Networks, each with distinct performance features [12].
2.3 Artificial Immune System

AIS which is inspired by the immune system represents the new and rapidly growing field of computer science. The approach is expected to give rise to powerful and robust information processing capabilities for complex problems. It possesses all the main properties of an AI system. AIS can learn new information processing, recall previously learned information and perform pattern recognition in highly decentralized fashion. Currently, AIS adopted three immunological principles [15][3]:

- Negative Selection
- Clonal Selection
- Immune Network Theory

The main role of the immune system is to provide defense mechanisms against foreign cells and to remove the malfunctioning cells. In order to perform these functions, the immune system needs to distinguish between good cells (known as self) and bad cells (non-self). The self elements that recognize any element of the self are eliminated. The process creates circulating self elements that react to non-self molecules only and thus is known as negative selection.

Clonal selection is a theory that explains how the immune system fights against non-self elements known as antigens. The system replicates the cells whose receptors can recognize and bind with an antigen. The replication is proportional to their degree of recognition. During the replication, individual cells undergoes a mutation so that it will be more adapted to the antigen recognition [5][15]. A basic algorithm of clonal selection is CLONALG [15][16]. The algorithm has some similarities with GA.

Immune network theory was proposed by Jerne in 1974 where she hypothesized that the immune cells and molecules are capable of recognizing each other in addition to recognizing invading antigens [5][15]. This endows in a network of communication between immune cells. When an immune cell recognizes an antigen or another immune cell, it is stimulated resulting in cell proliferation, cell activation and antibody secretion. However, when the cell is recognized by another immune cell, it is suppressed [15].

The computational model of AIS has been applied in various fields such as pattern recognition and classification, data analysis, computer security, robotics and optimization, information security and fault detection. In the field of information security, AIS is exploited in intrusion and detection system [17][18], virus detection [19] and steganography [20]. In fact, the analogy between computer security problems and biological processes was recognized as early as 1987, when the term “computer virus” was introduced by Adelman [21]. The role of the immune system may be considered analogous to that of computer security systems [22].

3.0 Review on Biological Intelligence Application in Cryptology

In this section, the application of BIC in cryptology will be presented. Literature review shows that GA, ANN, DNA, CA and ant colony have been exploited in cryptology. The review focuses mainly on works done using GA and ANN, however the works on DNA, CA and ant colony will be mentioned briefly for readers awareness.
3.1 Genetic Cryptology

Major works that involves GA focuses on cryptanalysis of cryptographic algorithms and design of cryptographic primitives. Most cryptanalytic research using GA was done on classical ciphers. An initial attempt was conducted by Spillman et al., whereby GA is exploited to cryptanalyse simple substitution ciphers [23]. Since known cryptanalytic attack for simple substitution ciphers employs frequency distribution of characters in the message [24], Spillman derived a cost or fitness function based on single-character and digram frequency distributions in his work. His attempt was fruitful as GA was proven to be highly successful in this cryptanalysis. He suggested the use of trigram frequency distribution and variations on crossover and mutation procedures as future research. Spillman continues his work and illustrated that GA can also be used in the cryptanalysts of public key cryptosystem, the knapsack ciphers [25]. The encryption scheme for knapsack ciphers is based on the NP-complete problem, which is a hard problem.

Another initial attempt conducted by Matthews [26] investigated the use of GA in cryptanalysis of transposition ciphers. In this work the fitness function is based on the message length, frequency distribution of digrams and trigrams tested for, the number of digrams and trigrams checked for and the likelihood of occurrence in successful deciphered messages.

An extensive research on classical cipher cryptanalysis was investigated by Clark in his PhD work [27] and Bagnall [28]. Clark’s cryptanalytic attack work covers a variety of classical ciphers that include simple substitution, transposition as well as poly-alphabetic ciphers. He proposed new attacks on these ciphers, which utilize simulated annealing and the tabu search. Existing attacks which make use of the genetic algorithm and simulated annealing are compared with the new simulated annealing and tabu search techniques. A comparison of search techniques of GA, simulated annealing and tabu search on cryptanalytic attack on simple substitutions was presented in a journal paper [10]. He showed that the tabu search outperformed the other techniques when used in the cryptanalysis of these ciphers. A parallel GA was proposed for attacking the poly-alphabetic substitution cipher by solving each of the key positions simultaneously. Through experimental evidence, he showed that parallel GA is highly efficient in solving poly-alphabetic ciphers even with a large period. On the other hand, Bagnall concentrates his work on cryptanalysis of Rotor machine [28]. He used a fitness function based on the phi test for non-randomness of text and showed that an unknown three rotor machines can be cryptanalysed with about 4000 letters of ciphertext.

The research using GA approach in cryptanalytic attack is gaining popular in the 2nd millennium. Further work on cryptanalytic attack on classical cipher is still taking place [29][30][31][32][33] while the work on modern ciphers emerges. Hernandez did the attack on two rounds TEA cipher [34] while Ali worked on RSA timing attack [35].

Another major application of GA in cryptology is in the design of Boolean functions and S-Boxes. The Boolean function is the linear combinations of S-box columns, which is an important cryptographic primitive for block and stream ciphers. The choice of Boolean functions used must be carefully considered when designing cryptosystems so that the cipher is not vulnerable to cryptanalytic attacks. Boolean functions for ciphers must be highly non-linear and balanced [36]. The conventional methods of Boolean function design are random generation and direct construction. By
using random generation, it is difficult to find functions with truly excellent properties due to the vast size of the search space while direct constructions may only meet certain design criteria. Millan enhanced his earlier work in designing balanced Boolean function using GA by combining it with the two-step hill climbing algorithm [37]. The hill climbing technique has improved the performance of GA in generating highly non-linear and balanced Boolean functions. Dimovski employs similar techniques but only tested for non-linearity criteria and compared it with random search which proved to be more powerful technique [38]. Finally, other application of GA in cryptology is the design of security protocols. Clark showed how simulated annealing and GA can be used to evolve efficient and provably correct protocols [39].

3.2 Neural Cryptology

ANN application in cryptology can be categorized in two sub-fields, that is cryptanalysis and key-exchange. Neural cryptanalysis work was conducted by Ramzan [40]. The aim of the research is to introduce new cryptanalytic techniques based on principles from machine learning, particularly ANN. He used Unix Crypt cryptosystems as a test bed and the results showed that ANN can accurately predict many of the plaintext bits with high probability even though the transfer function chosen for the network was rather naive. This proved that ANN can be trained to do cryptanalytic attack.

The work on neural key exchange is rather a new research area. The work in this area is performed by a research group from Institute for Theoretical Physics in Wurzburg, Germany and Minerva Center, Bar-Ilan University in Ramat-Gan, Israel [41][42]. Key exchange protocol is a cryptographic protocol that enables two users to exchange a key which will then be used for subsequent encryption of messages in a secure manner through public channel [24]. The first published protocol for this exchange is Diffie-Hellman key exchange whereby the security of the protocol is based on the difficulty of computing discrete log over a finite field GF(q)[43].

The neural key-exchange protocol does not employ number theory but is based on a synchronization of neural networks by mutual learning [41]. The architecture used is a two-layered perceptron, exemplified by a parity machine with \( K \) hidden units. The secret information of each entity is the initial values for the weights which are secret. Each network is then trained with the output of its partner. The work was extended to multilayer networks, parity machines [42]. When neural cryptography is combined with chaotic synchronization, the synchronization of neural networks accelerates [44]. A feedback mechanism is added to neural cryptography and is capable of enhancing the security of the system [45].

3.3 Other Biological Approaches in Cryptology

Other BIC approaches that are applied in cryptology are DNA, CA and ant colony. The basic element of CA is a biological cell which has a fixed set of states and a fixed position within the node of a spatial grid. It uses a simple mathematical abstraction of the principles of interaction between biological cells. Toffoli and Margolus showed that CA leads to the definition of a computing medium whose computational power is equivalent to the Turing machine [1]. CA has been applied in block and stream ciphers and these ciphers are implemented in hardware [46]. In fact the application of CA in cryptography came as early as 1980’s [47].
Cryptography based on one-time-pads was constructed using DNA due to its ultra-scale computation and ultra compact storage, thus only a small amount of DNA is needed for huge one-time-pads [48]. Finally, Bafghi performed a differential cryptanalysis on Serpent using ant colony and claimed that it can be used for any block cipher [49]. Ant colony algorithms are multi-agent systems where the behaviour of each single agent, the ants, is inspired by the behaviour of real ants [15]. Table 1 summarizes the work done on applications of BIC in cryptology.

Table 1: Biological approach applications in cryptology

<table>
<thead>
<tr>
<th>Biological Approach</th>
<th>Cryptology Sub-Field</th>
<th>Research Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>Cryptanalysis</td>
<td>Simple Substitution, Transposition, Polyalphabetic, Rotor Machine, Knapsack</td>
</tr>
<tr>
<td></td>
<td>Modern Cipher</td>
<td>TEA, RSA</td>
</tr>
<tr>
<td>GA</td>
<td>Cryptographic Primitive</td>
<td>Boolean Function and S-Box</td>
</tr>
<tr>
<td>GA</td>
<td>Cryptographic Protocols</td>
<td>Protocol Design</td>
</tr>
<tr>
<td>NN</td>
<td>Cryptanalysis</td>
<td>Unix Crypt</td>
</tr>
<tr>
<td>NN</td>
<td>Cryptographic Protocols</td>
<td>Key Exchange</td>
</tr>
<tr>
<td>DNA</td>
<td>Cryptographic Algorithm</td>
<td>Encryption</td>
</tr>
<tr>
<td>CA</td>
<td>Cryptographic Algorithm</td>
<td>Encryption for Stream and Block Cipher</td>
</tr>
<tr>
<td>Ant Colony</td>
<td>Cryptanalysis</td>
<td>Encryption for Block Cipher</td>
</tr>
</tbody>
</table>

4. Immune Cryptology

Analysis from the literature shows that AIS has some similarities with GA and ANN [12][15][50]. GA can be used for solving hard problems [9], thus sharing the same major characteristic with cryptology [39]. In cryptography, GA is used to heuristically design cryptographically strong balanced Boolean function, cryptographic protocols and techniques of cryptanalysis of several cryptographic algorithms [26][28][29][30][38][51]. The clonal selection algorithm of AIS computational model is similar to GA [15][16]. Cziko noted that the clonal selection can be interpreted as a Charles Darwin’s law of evolution, with the three major principles of repertoire diversity, genetic variation and natural selection [16]. Hence, it is in our opinion that this algorithm may be applied in the same cryptology subfield where GA was applied.

As discussed in section 3.2, ANN has been exploited to do key exchange protocol and cryptanalysis. ANN is a computational tool originally designed to model the human brain aiming at producing an “intelligent behaviour”. Castro and Zuben performed theoretical and empirical comparisons between ANN and aiNET, the artificial immne
network model and found that there are many similarities between them [52]. Immune network models have been used for the improvement of ANN models [5][53]. Castro and Zuben applied an immunological approach to initialize feed-forward neural network weights [53]. Tarakanov, in his paper [54] showed the principal possibility of using Formal Immune Network (FIN) for encryption but to date no work has been pursued. In order to prove that FIN is suitable, a rigorous mathematical model based on FIN that can be used for encryption need to be explored.

The discussion in this section suggests that application of AIS in cryptology is open to further investigations and innovations. A hybrid of BIC paradigms may also be explored to enhance the performance of the existing work.

5. Conclusion

This paper summarizes the work done in cryptology that employed BIC paradigm. A survey of literature shows that only a few research works exploits BIC in cryptology. However, there is a growing interest from the computer security community towards GA and ANN approach in cryptology in the second millennium. The potential utilization of ANN and GA has been shown useful in the design and analysis of cryptographic primitives, cryptographic protocols and cryptanalysis. However no work has been done in cryptology using the new emerging field of AIS. Since AIS shares some similarities with ANN and GA, thus it may have potential application in cryptology. A hybrid of AIS, ANN and GA for the application in cryptology may also be tackled for future research.

References


