Sborník vědeckých prací Vysoké školy báňské - Technické univerzity Ostrava
číslo 2, rok 2006, ročník LII, řada strojní
článek č. 1533

Tatiana HODOROGEA[*], Mircea-Florin VAIDA[**]

BLOOD ANALYSIS AS BIOMETRIC SELECTION OF PUBLIC KEYS

KREVNÍ ANALÝZA JAKO BIOMETRICKÝ VÝBĚR VEŘEJNÉHO KLÍČE

Abstract

In this work we consider a technical process for protecting medical information and other data assets using a technique of deriving DNA public keys from blood analysis. A DNA encryption technique is further developed here in which a person's medical data is encrypted in DNA strands based on the central dogma of molecular biology. Protection is enhanced by using a patient's own blood mineral levels as a seed for selecting, transmitting, and recovering that person's public key.

Abstrakt

V příspěvku se zabýváme technickým procesem ochrany lékařských informací a údajů s využitím technicky odvození veřejného klíče založeného na DNA pomocí krevní analýzy. Technika DNA šifrování je v příspěvku dále rozvedena a je ukázáno, jak jsou lékařská data osoby šifrována pomocí DNA vláken založených na principech molekulární biologie. Ochrana je zvýšena využitím hodnot úrovně minerálů v krvi pacienta jako základu pro výběr, přenos a regeneraci veřejného klíče této osoby.

1 INTRODUCTION

The topic of information security is very broad – involving issues ranging from physical protection of computer infrastructures to maintaining the privacy of individuals. As our society progresses into the information age, everyone from citizens to business and governmental institutions is becoming aware of the urgent need to prevent the exploitation of personal data, including medical records. In this work we consider a technical process for protecting medical information and other data assets using a technique of DNA cryptography further developed here in which a person's data is protected using his own blood mineral levels as a seed for selecting, transmitting, and recovering that person's public key. As we know that the management of public keys remains a challenge, we will use as the public key each unique individual blood analysis.

2 CRYPTOGRAPHY

Common use of the term "computer security" refers to several very important aspects of any computer-related system: confidentiality, integrity, availability, and authentication, [GARFINKEL 2005]. Essential parts of what we may call *data security*, specifically confidentiality and authentication, are achieved using cryptography, which has a long and fascinating history. The most complete non-technical account of the subject is David Kahn's book, The Codebreakers, [KHAN

[*]PhD, Faculty of Electronics and Telecommunications, Technical University of Cluj-Napoca, Baritiu Street No.26, 3400, Cluj-Napoca, Romania, tel. (+40) 745 033977, e-mail thodorogea@yahoo.com
[**] Professor, Department Name, Faculty of Electronics and Telecommunications, Technical University of Cluj-Napoca, Baritiu Street No.26, 3400, Cluj-Napoca, Romania, tel (+40) 0364 111994. e-mail Mircea.Vaida@com.utcluj.ro

1967]. Public Key Cryptography is one set of cryptographic techniques for providing confidentiality, preventing data compromise (disclosure to unauthorized persons), detecting alteration of data and verifying its authenticity, [VAIDA 2004]. Recent research considers the use of the Human genome in cryptography. In 2000, the Junior Nobel Prize was awarded to a young Romanian-American student, Viviana Risca, for her work in DNA steganography.

2.1 DNA Steganography

A DNA encoded message is first camouflaged within the enormous complexity of human genomic DNA and then further concealed by confining this sample to a microdot. A prototypical 'secret message' DNA strand contains an encoded message flanked by polymerase chain reaction (PCR) primer sequences (Fig. 1).

Fig. 1 DNA coder, [4]

To encode data characters in DNA, triplets were used in a simple substitution cipher. Denatured human DNA provides a very complex background for concealing secret-message. Viviana knowing both the secret-message DNA, PCR primer sequences and the encryption key could readily amplify the DNA and then read and decode the message [TAYLOR 1999].We propose to encode the medical records of an individual in DNA data strand flanked by unique primer sequences, which we obtain in the process of deriving DNA Public Key from blood analysis. We then mix it among other decoy DNA strands that will together be sent to a receiver through a public channel.

3 PUBLIC KEY INFRASTRUCTURE WITH DNA CRYPTOGRAPHY

Introducing DNA cryptography into the common PKI scenario, it is possible to follow the pattern of PKI, while also exploiting the inherent massively-parallel computing properties of DNA bonding to perform the encryption and decryption of the public and private keys, [GEHANI 1999]. The resulting encryption algorithm used in the transaction is much more complex than the one used by conventional encryption methods

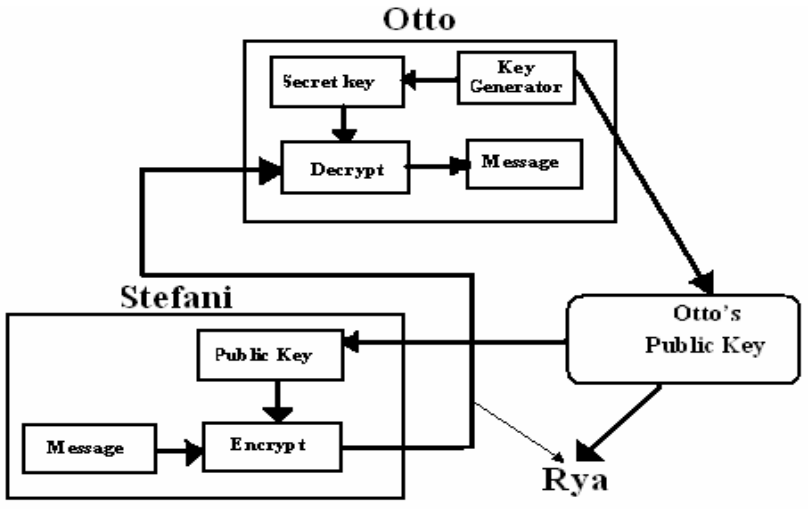

Fig. 2 Public Key Encryption

To put this into terms of the common Stefani and Otto description of secure data transmission and reception, they are basing their argument of DNA cryptography on Otto providing Stefani his public key which will constitute each unique blood analysis and Stefani will use it to send an encrypted message to him (Fig. 2). A potential eavesdropper, Rya will have a formidable amount of work to perform to attempt decryption of the transmission compared to Stefani or Otto, [HODOROGEA 2005]. Public key encryption splits the key up into a public key for encryption and a secret key for decryption. It's impractical to determine the secret key from the public key. Otto generates a pair of keys and tells everyone his public key, while only he knows his secret key. Anyone can use Otto's public key to send him an encrypted message, but only Otto knows the secret key to decrypt it. This asymmetric scheme allows Stefani and Otto to communicate in secret without having to physically meet, unlike symmetric key encryption methods, [CHUVAKIN 2004]. A prototypical 'secret message' DNA strand contains an encoded message flanked by primer sequences. The intended recipient identifies the secret data-carrying DNA strand using the program that associates the nucleotide sequence with the specific mineral levels from the particular individual's blood analysis. He then obtains the unique primer sequences that mark the beginning and the end of secret data DNA strand hidden among the decoy strands. In this last step, he uses the information conversion program and reads the medical record of the individual.

3.1 The Technique of Deriving DNA Public Keys from Blood Analysis

Starting from the idea that the DNA alphabet having 4 letters corresponding to the four nucleotides, A, C, G, T, a computer program generates a nucleotide sequence **S**, of length **L**, given a number n, by a random combination of DNA letters:

$$L=4^n \tag{1}$$

Then we associate a specific mineral **M**, with the corresponding nucleotide sequence S. As blood analyses are unique for each person, we then associate a specific mineral, M (like calcium), based on its concentration level **CL**, with the new nucleotide sequence $\mathbf{S}_1$. We derive S_1 from nucleotide sequence S, based on unique CL with value **V**. This value represents a unique concentration level of a certain M.

$$V= x.y_1y_2 \tag{2}$$

Where $x.y_1y_2$, is the number that represent the value V.

$$\Rightarrow CL=x.y_1y_2 \tag{3}$$

In the next step, S is assigned to M based on the concentration level CL. The resultant new nucleotide sequence S_1 of length $\mathbf{L_1}$ will constitute the unique primer sequence of an individual. $S_1=L_1$

$$L_1=x*S+ (L-(y_1+y_2)) \tag{4}$$

$$\Rightarrow S_1=x*S+ (L-(y_1+y_2)) \tag{5}$$

This nucleotide sequence S_1, based on the medical results of a specific person, will constitute the unique primer sequence. Thus a person's data-carrying DNA strand will be flanked by primer sequences unique to that individual. We will make the association in such a way that from every most recent blood analysis result, a new unique primer sequence S_1 of length L_1 will be generated. As we know that the management of public keys remains a challenge, [GARFINKEL 2001], we will use each unique blood analysis as the basis for a public key. The medical results will then be of no use to an unauthorized person, and for an intruder it would prove extremely difficult to read and detect the DNA strand that contains the medical history of a person without knowing the specific unique primer sequences of that person.

4 THE ALGORITHM BY STEPS

Step 1: Stefani (the sender) provides Otto (the receiver) her public key which will constitute each unique blood analysis of the specific person.

Step 2: A secret DNA data strand contains 3 parts:

- Secret DNA data strand in the middle
- Unique primer sequences on each side S_1.

Step 3: Stefani uses the technique of deriving DNA public key from blood analysis.

3.1 In this process Stefani uses a program which associates to a specific mineral the nucleotide sequence based on the medical results of a specific person which will constitute the *unique primer sequences* S_1

Step 4: Using an information conversion program Stefani encodes the medical records of an individual in DNA data strand flanked by unique primer sequences S_1 and mixes it among other decoy DNA strands.

Comments: *Otto will use programs that perform the reverse processes as those performed by Stefani: simulating the transcription, splicing, and translation per the Central Dogma of Molecular Biology (CDMB).*

4.1 According to CDMB during the process of transcription Stefani cuts out the introns from the data-encoded DNA, resulting in Encryption key 1. *E1= starting and pattern codes of introns => C1=E1(P), where P is plaintext and C is the ciphertext.*

4.2 Stefani translates the resulted spliced form of the data => Encryption key 2, *E2=the codon amino acids mapping => C=E2(C1).*

4.3 Stefani obtains the data-encoded protein after the translation process.

4.4 Stefani sends to Otto through a public channel the keys E1 and E2.

Step 5: Stefani sends to Otto through a public channel the encoded protein form of the data.

Step 6: Otto uses the key E2 to recover the mRNA form of the data from protein form of the data. *Decryption key D1=E2 => P1=D1(C).*

Step 7: Otto recovers the DNA form of the data in the reverse order as Stefani encrypted it. *Decryption key D2=E1=> P=D2(P1).*

Step 8: Otto identifies the secret data-carrying DNA strand using the program that associates the nucleotide sequence based on the blood to a specific mineral analysis of the particular individual. He obtains the unique primer sequences S_1 that mark the beginning and the end of secret data DNA strand hidden among the decoy strands.

Step 9: In this last step, Otto uses the information conversion program and reads the medical record of the individual.

5 BIOLOGICAL PRINCIPLES

The central dogma of the molecular biology is illustrated in (Fig. 3). A DNA segment that constitutes a gene is read, starting from the promoter (starting position) of the DNA segment. The non-coding areas (intron) are removed according to certain tags remain coding areas (extron) are rejoined and caped. Then the sequence is transcribed into a single stranded sequence of mRNA (messenger RNA). The mRNA moves from the nucleus into the cytoplasm. In chromosomes, DNA acts as a template for the synthesis of RNA in a process called transcription. RNA Synthesis and Processing in the transcription and the splicing steps, introns are cut out, and extrons are kept to form mRNA, which will perform the translation work, [BOREM 2003].

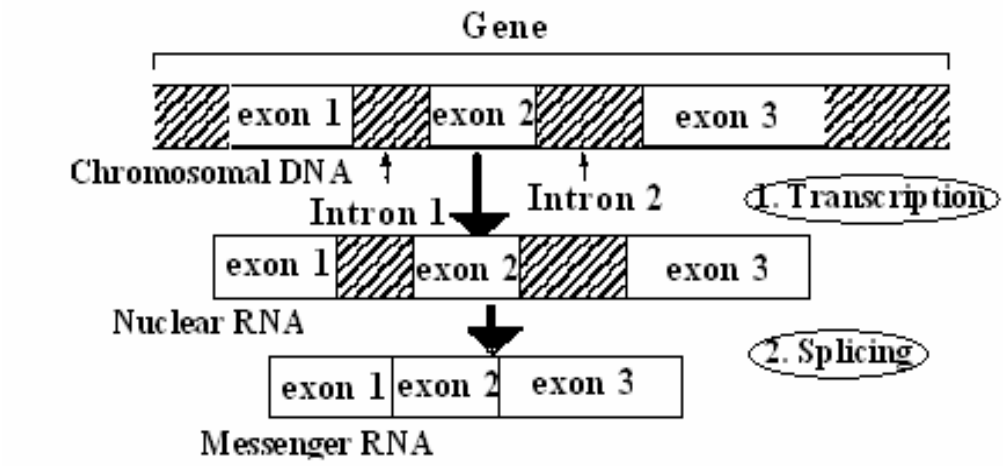

Fig. 3 Central Dogma of Molecular Biology

In the translation process, codons are translated into the amino acids according to the genetic code. The DNA form of information is scanned by the Sender to find the locations of the introns; which he then records. He cuts out the introns according to the specified pattern, so that the DNA form of data is translated into its mRNA form, which then translates into protein form of data according to the genetic code table (61 codons to 20 amino acids).

6 CONCLUSIONS

The DNA form of information, *D,* has length *n*, and is composed of *k* introns having average length *m,* thus, the mRNA form of information, *D'* has length: *n-k*m*, [KANG]. Since one codon (consisting of 3 nucleic acids) generally can be translated into one amino acid, the protein form of information *D''* has length: *(n-k*m)/3.* If Rya, an eavesdropper, can listen to Otto and Stefani's communication and tries brute force attack, it would be a very expensive computational problem for her. Such a brute force attack method could be impossible for Rya.

As an additional layer of security, we propose to associate each mineral with a corresponding mineral, which we will call here, a synergetic mineral pair – for example, Ca-Fe. Blood analysis may prove to be a secure and cost effective biometric method of selecting public keys for use in DNA encryption techniques.

REFERENCES

[1] BOREM Aluízio Fabricio, R.Santos, 2003 *Understanding Biotechnology*, Publisher: Prentice Hall PTRPub Date: January 17, 2003.

[2] GEHANI Ashish, La Bean, Thomas H. Reif, JohnH, 1999 *DNA-Based Cryptography*, Department of Computer Science, Duke University. June 1999,

[3] CHUVAKIN Anton, Cyrus Peikari, 2004, *Security Warrior*, Publisher: O'Reilly, Pub Date: January 2004

[4] HODOROGEA Tatiana, Mircea-Florin Vaida, 2005, *Alternate Cryptography Techniques*, ICCC 2005.

[5] GARFINKEL Simson, 2001, *Web Security, Privacy & Commerce*, 2nd Edition, Publisher, O'Reilly, November 2001

[6] GARFINKEL Simson, Lorrie Faith Cranor, 2005, *Security and Usability*, Publisher, O'Reilly, August 2005

[7] KAHN D., 1967, *The Codebrakers*, McMillan, New York, 1967

[8] KANG Ning, A *Pseudo DNA Cryptography Method*, Independent Research Study Project for CS5231

[9] TAYLOR Clelland Catherine, Viviana Risca, Carter Bancroft, 1999, Hiding Messages in DNA Micodots. *Nature Magazine* Vol.. 399, June 10, 1999.

[10] VAIDA Mircea-Florin, 2004, Information Society Development and Human Evolution, In *ICCC 2004*, Baile Felix, May 27-29, 2004, pp. 414-420

[11] VAIDA Mircea-Florin, 2004, Security and Java, In *Conference at the Université de Savoie, France*, supported by Shuffle project, May 2004

[12] VAIDA Mircea-Florin 2005, Teaching Computers As a Human Spiritual Evolution, In *4th IASTED International Conference on Web-Based Education*, Grindelwald, Switzerland, pp. 667-672

Reviewer: doc. Ing. Radim Farana, CSc., VŠB – Technical University of Ostrava