Introduction to Geometric Intronization as a Security Technique

Qinghai Gao[†], Xiaowen Zhang[††], and Michael Anshel[†††]

[†] Dept. of Computer Science, Graduate Center / CUNY, New York, NY, USA
[††] Dept. of Computer Science, College of Staten Island / CUNY, Staten Island, NY, USA
[†††] Dept. of Computer Science, City College of New York / CUNY, New York, NY, USA

Summary

The intronization is defined as an encryption method, i.e., inserting introns into an exon sequence to obtain ciphertext. Intronization can be used for information hiding, password salting, virus morphing, and an intermediate step of other encryption or hash primitives. Geometric objects can be applied to guide the intronization process. The existence of large number of introns in ciphertext could make frequency-based cryptanalysis difficult.

Key words:

Geometry, Biometrics, Intronization, Exon, Security, MER.

1. Introduction

According to Shannon [1], a strong cipher should have good diffusion and confusion property. To achieve these properties some techniques would be applied during encryption such as substitution, permutation / transposition, combination, fractionation, etc. Cryptographic hash functions (e.g., SHA-1 and MD5) have good diffusion and confusion properties. However, they can't be used for matching biometrics due to the fuzzy measurement of biometrics.

In Biology, the genes of eukaryotes have introns that separate exons [2]. A pre-mRNA transcript is made directly from a gene, then the introns are sliced out, and exons are joined sequentially to form mRNA, which becomes the template of a protein sequence. In a eukaryotic cell, only less than 10% of the entire DNA sequence is directly used for protein coding. That is to say the majorities of DNA are introns, which were once called junk DNA. Modern biologists believe that introns play important roles. However, finding the exact roles of introns is an ongoing research problem. From the security point of view, the existence of introns or non-coding regions in DNA may be helpful to the survivability of organisms. It is commonly agreed that more advanced organisms have more introns in their DNA.

Inspired by introns we define Intronization [3] as a process of inserting non-coding symbols (introns) into plaintext (exons) to obtain ciphertext (mixture of introns and exons). Even though we borrowed the term "Intronization" from Biology, however, our method of introducing introns into a sequence is different from what happens in nature [4, 5]. Different methods can be used to insert introns in a plaintext. The randomized intronization is introduced in [3]. In this paper we propose to use geometric objects to intronize plaintexts. The rest of the paper is organized as follows. Section 2 introduces intronization with modified cubes. Section 3 proposes methods to control message expansion rates, which is followed by qualitative security analysis in Section 4. Finally, Section 5 concludes the paper with conclusions and future research.

2. Geometric Intronization

Geometric objects can be selected or designed to act as the key(s) for the intronization technique. In this section we will use the following examples to illustrate how geometric intronization technique works.

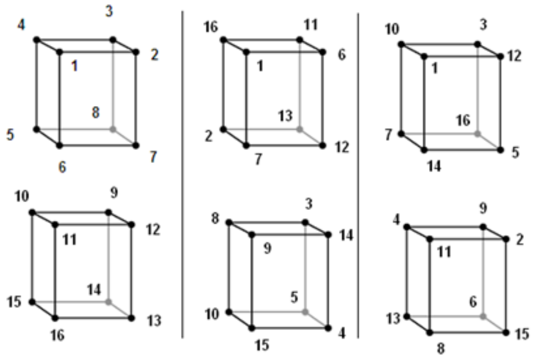


Fig. 1 Obtaining row I and II for Table 1(No intron)

Example 1

In this example we use a simple cube to wrap a sequence of length 16 without appointing any intron vertex. By wrapping the sequence along the vertices of the cube and then reading out vertically (refer to Fig. 1), we obtain the results given in Table 1, in which the sequence repeats after the third row. Therefore it is not secure.

Table 1 Simple cube, counter-clockwise (Top view) wrapping, 5 rounds *

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
I	1	6	11	16	2	7	12	13	3	8	9	14	4	5	10	15
II	1	7	9	15	6	12	14	4	11	13	3	5	16	2	8	10
III	1	14	11	8	12	5	2	15	3	16	9	6	10	7	4	13
IV	1	7	9	15	6	12	14	4	11	13	3	5	16	2	8	10
V	1	14	11	8	12	5	2	15	3	16	9	6	10	7	4	13

*Refer to Figure 1 for row I and II

Table 2 Body-centered cube, counter-clockwise (Top view), 4 rounds*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I	1	9	13	17	2	6	14	18	3	7	11	19	4	8	12	16	5	10	15	20
II	1	3	4	5	9	6	8	10	13	14	11	15	17	18	19	16	2	7	12	20
III	1	4	5	9	13	6	10	14	17	18	11	19	2	7	12	16	3	8	15	20
IV	1	9	13	17	2	6	18	7	3	8	11	15	4	10	19	16	5	14	12	20

*Refer to Figure 2 for row I and II. Introns are gray in color.

Table 3 Body-centered cube, counterclockwise and then clockwise (Top view), 4 rounds

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I	1	7	11	17	2	6	12	16	3	9	13	19	4	8	14	18	5	10	15	20
II	1	12	13	5	7	6	19	18	11	3	4	15	17	16	8	10	2	9	14	20
III	1	15	17	7	19	6	14	9	4	13	5	8	2	10	18	3	12	11	16	20
IV	1	18	12	14	16	6	10	13	7	2	19	9	15	13	11	17	8	5	3	20

Example 2

We use a body-centered cube (BCC) to wrap a sequence by appointing the central vertex as intron position. The sequence is wrapped around the cube counterclockwise (Top view, Figure 2), and then read out vertically to form the new sequence. Since the center position of the cube is appointed as the intron position, 5, 10, 15 and 20 are all introns (gray-colored). With four rounds there are 7 exons and 13 introns, as given in Table 2.

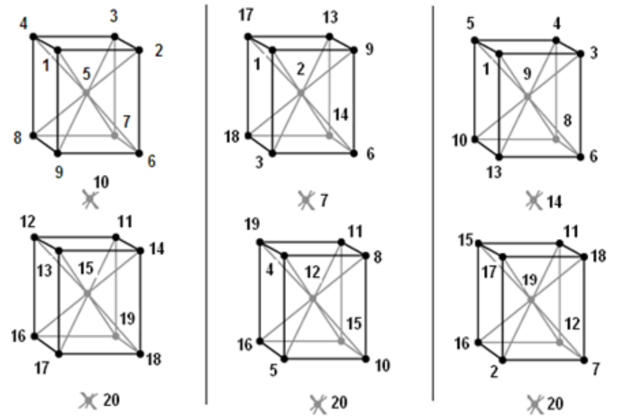


Fig. 2 Obtaining row I and II for Table 2

In Table 2, five positions, including 1, 6, 11, 16, and 20, never change values due to the one directional wrapping. If we apply a more sophisticated wrapping, e.g., wrapping counterclockwise and then clockwise alternatively (refer to Figure 3), the results are given in Table 3. We can see that with four rounds there are 5 exons and 15 introns (row IV). Since two exons (1 and 6) and one intron (20) never change their values during the four rounds of wrapping, an attacker can use the information to attack the system.

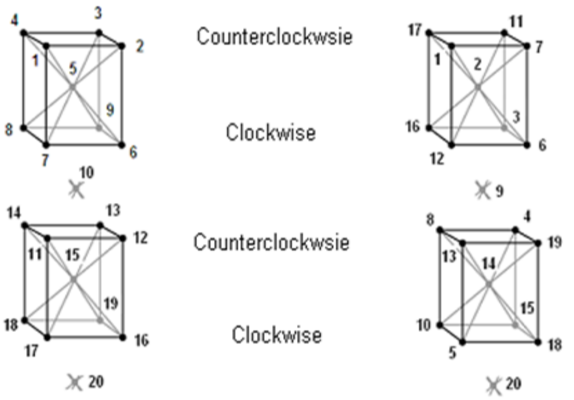


Fig. 3 Obtaining row I for Table 3

One way to overcome the above-mentioned problem is to

circularly rotate the sequence before wrapping it around the cube. Table 4 gives such an example. In Table 4, the row II and IV are the rotational results of the row I and III, respectively.

Table 4 Body-centered cube, rotating by 4 positions, 5 rounds (II and IV are rotations)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I	1	7	11	17	2	6	12	16	3	9	13	19	4	8	14	18	5	10	15	20
	5	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*1*	*2*	*3*	*4*
II	2	6	12	16	3	9	13	19	4	8	14	18	5	10	15	20	1	7	11	17
III	2	13	14	1	6	9	18	20	12	4	5	11	16	19	10	7	3	8	15	17
IV	6	9	18	20	12	4	5	11	16	19	10	7	3	8	15	17	2	13	14	1
V	6	5	10	2	9	4	7	17	18	16	3	14	20	11	8	13	12	19	15	1

Table 5 Body-centered cube, rotating by 4 positions, 5 rounds (II and IV are rotations)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I	1	7	11	17	2	6	12	16	3	9	13	19	4	8	14	18	5	10	15	20
	5	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*1*	*2*	*3*	*4*
II	2	6	12	16	3	9	13	19	4	8	14	18	5	10	15	20	1	7	11	17
III	2	13	14	1	6	9	18	20	12	4	5	11	16	19	10	7	3	8	15	17
IV	6	9	18	20	12	4	5	11	16	19	10	7	3	8	15	17	2	13	14	1
V	6	5	10	2	9	4	7	17	18	16	3	14	20	11	8	13	12	19	15	1

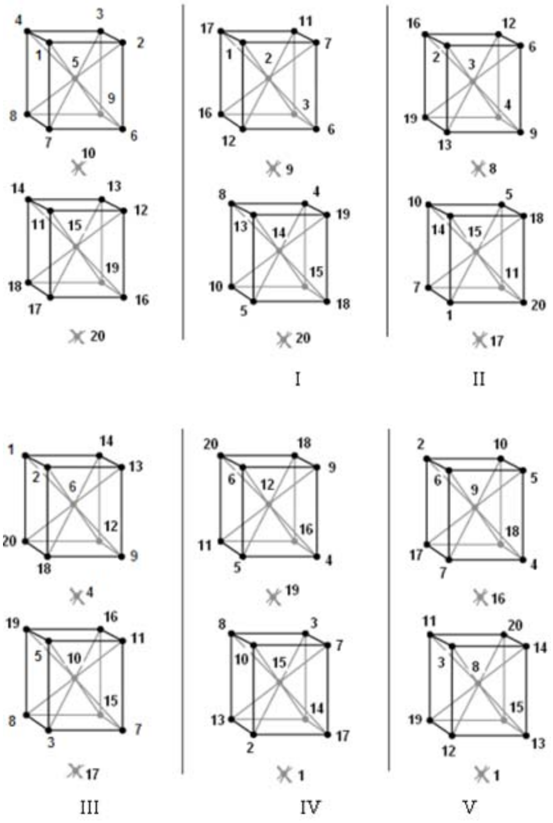


Fig. 4 Obtaining rows for Table 5

From these results we can see that one disadvantage of the intronization technique is that more rounds mean less positions can be used for plaintext symbols if all other factors are the same. For example, in the row III of Table 4 there are eight exons (13, 1, 18, 12, 11, 16, 19 and 7) and the 12 remaining positions are introns, thus Message Expansion Rate (MER)=20/8=2.25; in the row V there are only four exons (7, 18, 11, and 13) and 16 introns, thus the MER=20/4=5.

One advantage of the intronization technique is that the geometric object can be designed arbitrarily. For example, if we use a hexagon and select the numbered vertices in Figure 5 to wrap a sequence of length 20, the following sequence can be generated with one round (Note that some of the vertices in Figure 5 are skipped. Lay the left hexagon on top of the right one then read out the numbered vertices vertically):
1, 11, 12, 23, 2, 16, 22, 3, 9, 15, 21, 4, 14, 5, 10, 13, 20, 6, 18, 19, 7, 8, 17

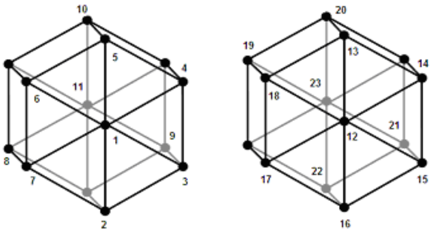


Fig. 5 Hexagon

Table 6 Multiple ciphertexts for one plaintext (X=A, T, C or G)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I	6	5	10	2	9	4	7	17	18	16	3	14	20	11	8	13	12	19	15	1
II	X	X	A	X	T	T	X	X	X	G	X	C	G	G	A	X	X	T	C	X

Also for a selected geometric object we can select the intron positions arbitrarily. For example, in Table 5 we choose vertex 1, 7 and 17 as intron positions and follow the same steps as those of Table 4, we will obtain a different sequence as given in Table 5. In the row V of Table 5 there are 10 exons and 10 introns.

Another advantage is that the values of introns can be chosen arbitrarily if there is no combination between exons and introns. Therefore, many ciphertexts can be generated for each plaintext. Table 6 gives such an example for a plaintext string ATTGCGGATC. Note each X can be A, C, G or T.

One problem with geometric wrapping is that there may not be enough randomness in the output. To approach the problem, we can use pseudo random sequence to guide the wrapping (refer to Figure 6). Assume we wrap a sequence vertex-by-vertex with the simple cube. After we reach vertex 4, there will be four different choices for the vertex 5. With this method we need another sequence to specify which vertex to select at each level. And this level-wised vertex selection sequence can be generated with PRNG.

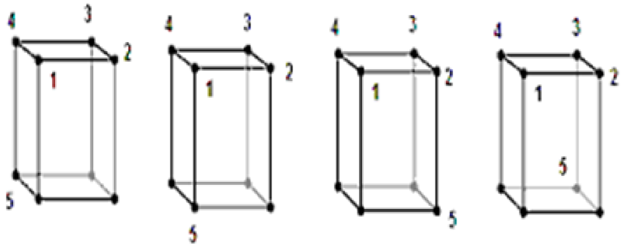


Fig. 6 Wrapping guided by pseudo random sequence

3. Control the Message Expansion Rate

The security of the intronization method depends on MER (i.e., the length of ciphertext divided by length of plaintext.), which is an undesirable results in terms of storage and processing. An ideal way of using intronization would be maximizing the security with limited message expansion rate.

A few methods can be used to control MER, including **Exon Elimination** and **Intron Compression/Removal**.

Exon Elimination

So far, the ciphertexts obtained in Section 2 are mixtures of introns and exons. If exons contain fuzzy bits, arithmetic operations like XOR between exons will unavoidably propagate errors. Therefore, we need to avoid combining two exons.

Unlike exons, introns do not contain any fuzzy bits. Therefore, XORing an exon with an intron will not introduce new error bits. Based on this observation, we introduced a technique of dissipating exon into intron, and called it **Exon Elimination**. Table 7 gives such an example. The row III is obtained by removing the blank cells in row II and shift all the letters to the left. With Exon Elimination, an exon set can be completely hidden in an intron set even though the diffusion property [1] can be limited.

Intron Compression/Removal

In Table 8, we apply Intron Compression to row I, for example, by XORing continuous introns, to generate the row II. The row III is obtained by left packing. The non-invertibility is enhanced because it is difficult to map from the row III or the row II back to the row I.

In Table 9, the row I is the original intronized biometric template. Assuming we choose to remove the introns whose lengths are equal to or greater than a number, e.g., 3, we obtained the result given in the row II. Right shifting gives the row III.

4. Security Analysis of Intronization

Shannon [1] listed five criteria to estimate the value of a proposed secrecy system:

- *Amount of secrecy*: the less the better
- *Size of key*: smaller is better
- *Complexity of enciphering and deciphering operations*: simpler is better
- *Propagation of errors*: less is better
- *Expansion of message*: less is better

As Shannon [1] pointed out, it is very difficult to achieve good results for all five criteria. Table 10 gives a comparison of different cryptographic techniques.

From Table 10, we can see that the main advantage of the intronization technique is its zero-error propagation. Its main disadvantage is message expansion.

Table 7 Exon Elimination - Dissipating exon into intron(A=10, C=00, G=11, T=01, operation XOR)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
I	A	C	G	A	T	A	G	G	C	T	T	G	A	G	T	C	C	A	G	C
II	T	T		C			A	C	A		A					T	C	A	G	C
III	T	T	C	A	C	A	A	T	C	A	G	C								

* The calculations for the 3rd row:

1⊕3=A⊕G=T, 2⊕5=C⊕T=T, 4⊕6=A⊕A=C, 7⊕10=G⊕T=A, 8⊕12=G⊕G=C,
9⊕13=C⊕A=A, 11⊕14=T⊕G=A, 15⊕16=T⊕C=T

Table 8 Illustration of Intron Compression*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	6	5	10	2	9	4	7	17	18	16	3	14	20	11	8	13	12	19	15	1
I	1	0	1	1	0	0	1	1	0	1	1	1	0	1	1	1	0	0	1	1
II		1		1		1		0		1		1		1		1		0		
III	1	1	1	0	1	1	1	1	0											

* Ciphertexts are binary and introns are gray in color.

Table 9 Illustration of Intron Removal*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	6	5	10	2	9	4	7	17	18	16	3	14	20	11	8	13	12	19	15	1
I	1	0	1	1	1	0	1	0	0	0	0	0	1	0	1	1	1	0	1	0
II	1	0	1	1	1	0				0	0	0	1	0	1	1	1	0	1	0
III				1	0	1	1	1	0	0	0	0	1	0	1	1	1	0	1	0

* Ciphertexts are binary and introns are gray in color.

Table 10 Comparison of intronization with other cryptographic techniques

Cipher	AES	RSA	OTP	Vigenere	Intronization	
					PRNG	Geometric
Amount of secrecy	key	Private key	Same as message	Key	Seed	Object, Intron positions
Size of key	128, 192, 256	512, 1024, 2048	Variable	Variable	Variable	Variable
Operation complexity	Complex	Complex	Simple	Simple	Complex	Complex
Propagation of errors	Yes	Yes	No	No	No	No
Expansion of message	No	No	No	No	Yes	Yes

Shannon [1] proposed these criteria about 60 years ago, when computers were slow, and had very limited processing power, memory and storage. However, modern computers have far superior processing capability, memory and storage, the disadvantage of message expansion should be able to be offset at least partly by the enhancement of security.

The intronization technique is secure against ciphertext-only attack (COA). It is relatively vulnerable to known-plaintext attack (KPA). However, it is not easy to launch KPA for the following two reasons. First, exact plaintext of a biometric template is not easy to obtain by stealing biometric image. Second, Exon Elimination could make ciphertext secure. Our effort has been focused on developing intronization into a secure method against COA without using Shannon's diffusion property. The intronization technique can also be applied to enhance the security of substitution ciphers against the common frequency analysis attack.

As Shamir stated in his 2002 Turing Award lecture [6], the three laws of security are:

- Absolutely secure systems do not exist
- Cryptography is typically bypassed, not penetrated
- To halve your vulnerability, you have to double your expenditure

Intronization, an information security technique developed by nature in billions of years, should have been used more widely.

5. Conclusions and Future Research

Inspired by the Central Dogma of Biology, we exemplify the geometric intronization as a security technique, which can be used for information hiding, password salting, virus morphing, or as an intermediate processing step for other conventional encryption or hash algorithms. Among other factors the security of intronization also depends on message expansion rate. It seems that larger the MER, more secure the system is. However, insertion of introns increases the demand for storage and processing. Therefore, Exon Elimination, Intron compression and Intron Removal are proposed to reduce MER.

In the future we would like to do more extensive tests, to develop more advanced intronization techniques and better mechanisms to control and balance the MER, and further to investigate the relationship between security and the MER.

Acknowledgments

The authors would like to thank Professor Ted Brown for his invaluable support for our research.

References

[1] Claude E. Shannon. Communication Theory of Secrecy Systems. *Bell System Technical Journal, Vol.28-4, page 656-715, 1949.*

[2] Annie S. Wu, Robert K. Lindsay. A Survey of Intron Research in Genetics. Proceedings of the 4th International Conference on Parallel Problem Solving from Nature, *Lecture Notes In Computer Science; Vol. 1141, p101-110 (1996).*

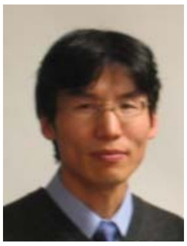
[3] Qinghai Gao. *Secure Biometrics*. PhD thesis (2008), The City University of New York.

[4] M. Irimi et al. Origin of introns by 'intronization' of exonic sequences. *Trends in Genetics 2008 Aug;24(8):378:81*

[5] P Senapathy. Introns and the origin of protein-coding genes. *Science 1995 268:1366-136.*

[6] Adi Shamir, Turing Lecture on *Cryptology: A Status Report* (2002).

Dr. Qinghai Gao currently works in the IT department of Linear Lighting Corporation in New York. His present research interests include Biometrics, Biological Information System, Cryptography, Polymorphic virus and Network Security. He received a PhD in Computer Science from the City University of New York in December 2007.



Dr. Xiaowen Zhang is a faculty member of the College of Staten Island (CSI). Prior to joining CSI, he worked in both academia as a research fellow and lecturer, and industry as a software and electronic engineer. His research interests include information security, cryptography, RFID, quantum computing, and wireless communications. He received a PhD in Computer Science from the City University of New York (CUNY), New York, USA in 2007, and a PhD in Electrical Engineering from Northern Jiaotong University, Beijing, China in 1999.

Dr. Michael Anshel has taught at the City College of New York (CUNY) since 1968. He has been a member of the CUNY Doctoral Faculty since 1973, teaching in the Engineering, Computer Science and Mathematics programs. He has mentored over forty doctoral dissertations. Dr. Anshel is co-inventor of three patents in cryptography and has published numerous articles in Mathematics and Cryptography. Dr. Anshel is a member of the AMS, MAA, ACM, IEEE, IACR. Dr. Anshel holds a Ph.D. in Mathematics from Adelphi University in Garden City, New York.