랜덤 코돈 원형 부호 기반의 DNA 워터마킹

이석환[†], 권성근^{††}, 권기룡^{†††}

요 약

본 논문에서는 DNA 시퀀스의 불법 복제 및 변이 방지를 위한 DNA 워터마킹 기법을 제안한다. 제안한 DNA 워터마킹은 랜덤 맵핑 테이블에 의하여 코돈들을 랜덤 원형 각도로 수치화한 다음, 웨이블릿 국부 계수 최대치의 Lipschitz regularity 상수에 의하여 삽입 대상 코돈들을 탐색한다. 워터마크 삽입과정에서 DNA의 아미노산 코드가 변경되지 않도록 하기위하여 삼중 코돈들의 랜덤 코돈 원형 각도에 워터마크를 삽입한다. 삽입 대상 코돈들의 길이와 위치는 랜덤 맵핑 테이블에 의존하므로, 이 테이블을 알지 못할 경우, 워터마크 추출이 어렵다. 그리고 제안한 방법은 다양한 길이의 DNA 서열에 64개 코돈(종료, 개시 코돈 포함)들의 랜덤 맵핑 테이블을 적용함으로써 동일한 길이의 워터마크 키를 적용한다. 본 실험에서는 랜덤 맵핑 테이블과 삽입 위치의 높은 엔트로피를 통하여 워터마크의 보안성을 확인하였다. 또한 기존의 DNA-Crypt 워터마킹과의 유사한 용량 하에서 제안한 방법이 낮은 염기 변화율을 가지며, 포인트 변이, 삽입 및 삭제 변이에 대하여 낮은 에러율을 가지며, ROC 분석을 통하여 우수한 검출 능력을 가짐을 확인하였다.

DNA Watermarking Method based on Random Codon Circular Code

Suk-Hwan Lee[†], Seong-Geun Kwon^{††}, Ki-Ryong Kwon^{†††}

ABSTRACT

This paper proposes a DNA watermarking method for the privacy protection and the prevention of illegal copy. The proposed method allocates codons to random circular angles by using random mapping table and selects triplet codons for embedding target with the help of the Lipschitz regularity value of local modulus maxima of codon circular angles. Then the watermark is embedded into circular angles of triplet codons without changing the codes of amino acids in a DNA. The length and location of target triplet codons depend on the random mapping table for 64 codons that includes start and stop codons. This table is used as the watermark key and can be applied on any codon sequence regardless of the length of sequence. If this table is unknown, it is very difficult to detect the length and location of them for extracting the watermark. We evaluated our method and DNA-crypt watermarking of Heider method on the condition of similar capacity. From evaluation results, we verified that our method has lower base changing rate than DNA-crypt and has lower bit error rate on point mutation and insertions/deletions than DNA-crypt. Furthermore, we verified that the entropy of random mapping table and the location of triplet codons is high, meaning that the watermark security has high level.

Key words: DNA watermarking(DNA 워터마킹), Codon random mapping table(코돈 랜덤 맵핑 테이블), Amino acid code preservation(아미노산 부호 보존성), Mutation resistance(변이 견고성), Watermark security(워터마크 보안성)

※ 교신저자(Corresponding Author): 이석환, 주소: 부산광역시 남구 용호동 535번지 동명대학교(608-711), 전화: 051) 629-1285, FAX: 051) 629-1249, E-mail: skylee@tu.ac.kr

접수일: 2012년 10월 22일, 수정일: 2013년 1월 4일
완료일: 2013년 1월 26일

[†] 정회원, 동명대학교 정보보호학과

^{††} 종신회원, 경일대학교 전자공학과
(E-mail: sgkwon@kiu.ac.kr)

^{†††} 종신회원, 부경대학교 IT융합응용공학과
(E-mail: krkwon@pknu.ac.kr)

※ 본 연구는 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (KRF-2011-0023118)

1. 서 론

21세기 게놈 프로젝트의 완성, 유전자 정보를 이용한 의학 및 진단기술 발전, 및 생체인증기술 발달 등으로 인하여 새로운 유전자 정보에 대한 프라이버시 침해가 가능하게 되었다. 특히 DNA 안에 포함된 유전 코드(genetic code)는 '인간의 일기'라고 불리어지는 심오한 개인 정보로, 타인에게 노출될 경우 프라이버시를 포함한 인권 침해가 심각할 것이다. 최근 GMO(Genetically Modified Organism)의 암호 및 정보은닉을 위하여 DNA/RNA 서열인 A, G, T(U), C 문자열에 대한 암호화[1,2], 스테가노그래피[3-6] 및 워터마킹 기법 [7-14]들이 제안되어지고 있다.

DNA 암호화는 PCR(Polymerase Chain Reaction)과 DNA 칩 기반으로 생물학적인 DNA 암호 및 복호 방법을 제시한 것이나, 실제 구현의 어려움으로 인하여 아직까지 기존 암호 기술을 대체할 수 없다. DNA 스테가노그래피는 DNA에 정보를 은닉하는 기술로 DNA에 대용량 정보 저장, DNA 서명 및 인식에 매우 유용하다. 그러나 DNA 암호화 및 스테가노그래피는 실험 환경 또는 DNA 변이 등과 같은 경우에 DNA 또는 정보를 복원할 수 없으므로, 견고성이 요구되는 응용에서는 적합하지 못하다. 이와 같은 DNA 암호화 및 스테가노그래피의 문제점으로 DNA 워터마킹 기법이 제시되고 있다. DNA 기반 워터마크는 GMO의 저작권 보호 또는 야생형(wildtype) 유전체와 인공 유전체(artificial genome)와의 구별에 적용될 수 있다. 그러나 DNA 기반 워터마크는 부호 영역의 단백질 기능이 변경되지 않도록 삽입되어야 하며, 이는 기존 멀티미디어 워터마킹과의 가장 큰 차이점을 나타낸다. 기존 DNA 워터마킹 기법들을 살펴보면, 단순한 치환 기법 또는 유전 부호 기반의 심볼 및 코돈들의 비트 할당에 의한 워터마크 삽입과 in vivo 기반의 워터마크된 셀 구현에 초점을 맞추고 있다.

DNA는 길게 이어져 있는 뉴클레오티드(nucleotide)로 이루어져 있으며, 하나의 뉴클레오티드는 인(phosphate), 당(sugar), 염기(base)로 구성되어 있다. 그 중 DNA의 염기는 A (Adenine), G(Guanine), C(Cytosine), T(Thymine)의 네 종류가 있다. 여기서 코돈(codon)은 세 개의 염기서열로 유전 부호(Genetic code)에 의하여 하나의 아미노산으로 결정

된다. 즉, DNA의 코돈서열은 단백질 합성과정에서 유전 부호에 의하여 아미노산 서열로 번역된다. 부호 영역의 DNA 워터마킹은 코돈 서열에 워터마크를 삽입하는 것이다. 이 때, DNA 워터마킹은 다음과 같은 주요 요구 사항을 만족하여야 한다. 첫 번째는 아미노산 부호 보존성으로, 원본 DNA와 워터마크된 DNA가 전사되는 아미노산 서열들이 동일하여야 한다는 것이다. 이는 기존 멀티미디어 워터마킹과는 다른 DNA 워터마킹의 고유한 요구 사항이다. 두 번째로 워터마크는 DNA 삽입, 복제, 삭제 및 치환 등의 변이에 강인하여야 하며, DNA 정보 훼손에도 워터마크가 추출되어야 한다는 것이다. 네 번째로 제3자에 의한 워터마크 추출이 확률적으로 불가능하도록 DNA 워터마크의 보안성이 확보되어야 한다는 것이다. 마지막으로 충분한 정보가 삽입될 수 있도록 워터마크의 용량이 확보되어야 한다. 따라서 신호처리 관점에서 위와 같은 아미노산 부호 보존성, 강인성, 보안성, 및 용량성과 같은 평가 분석과 이들 조건을 만족하는 DNA 워터마킹 기법이 필요하다.

본 논문에서는 아미노산 부호 보존성, 변이 견고성, 및 워터마크 보안성을 가지는 cDNA(coding DNA) 서열 워터마킹 기법을 제안하며, 이를 in silico 기반으로 평가 분석한다. 제안한 방법의 주요 특징으로는 다음과 같다. 먼저 워터마킹 신호 처리 용이와 워터마크 보안성을 위하여 코돈 서열들은 랜덤 맵핑 테이블에 의하여 랜덤 원형 각도로 수치 변환된다. 랜덤 맵핑 테이블은 기존 유전자 부호 기반의 64개 코돈별로 맵핑으로 이루어지며, 이는 워터마크 키로 사용된다. 두 번째로는 랜덤 원형 각도들의 다해상도 국부 계수 최대치의 Lipschitz regularity 상수를 이용하여 삽입 대상 3중 코돈들을 선택한다. 랜덤 맵핑 테이블에 따라 다해상도 국부 계수 최대치는 달라지며, Lipschitz regularity 상수에 의한 삽입 대상 코돈들의 길이와 위치 또한 달라진다. 따라서 랜덤 맵핑 테이블을 모를 경우 워터마크 추출에 필요한 삽입 코돈들의 위치를 검출하기가 매우 어렵다. 세 번째로는 워터마크 비트에 따라 삽입 대상 코돈들의 랜덤 원형 각도 수치에 삽입된다. 이 때 아미노산 부호 보존을 위하여 아미노산의 원형 각도 수치 범위 내에 워터마크가 삽입되며, 변이 견고성을 위하여 대상 코돈의 중심 코돈과 인접 두 코돈의 원형 각도들의 원형 각도 차이에 각각 삽입된다.

in silico 기반의 평가 실험에서 제안한 방법이 아미노산 부호 보존성을 유지하며, BER 및 ROC 분석으로부터 기존 cDNA 방법에 비하여 치환, 삽입, 및 삭제 변이에 견고함을 확인하였으며 데이터 용량이 약 0.049정도 낮음을 확인하였다. 또한 엔트로피 계산을 통하여 워터마크 키로 사용되는 랜덤 맵핑 테이블의 보안성이 우수함을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 DNA 염기서열 구조, 유전자 부호, 및 DNA 워터마킹 기법들에 대하여 분석하고, 3장에서는 제안한 DNA 워터마킹 기법에 대하여 자세히 살펴보기로 한다. 4장에서는 in silico 기반 다양한 DNA들에 대한 평가 분석을 살펴보고 마지막으로 5장에서는 본 논문의 결론을 맺는다.

2. 기존 연구

DNA에 정보를 은닉하기 위한 연구는 1999년-2000년대 초 DNA 암호화 및 스테가노그래피를 시작으로 2000년대 후반 DNA 워터마킹까지 이어지고 있다. 그러나 이에 대한 연구는 일부 바이오 연구자들에 의하여 진행되고 있다.

Clelland 등[3]은 DNA 마이크로토드를 이용한 최초 DNA 스테가노그래피 기법을 제시하였고, in vitro에 의하여 구현하였다. 이 기법은 DNA 코드와의 간단한 치환 Cipher에 의하여 비밀 메시지를 삽입하나, 메시지 복호를 위하여 복호 함수와 시작/끝점인 Primer 정보가 필요하다. Leier 등[4]은 더미 DNA 시퀀스 내에 메시지 시퀀스를 이진 가감법에 의하여 결합하는 스테가노그래피 기법을 제안하였다. 그러나 이 기법 또한 메시지 복호를 위한 Primer 시퀀스 정보가 필요하며, 가감법으로 인하여 보안성이 취약하다. Shimanovsky 등[5]은 산술 부호와 공개키 기반 암호 기반으로 코돈 중복을 이용한 cDNA 데이터 삽입 기법을 제안하였다. 이 기법은 이진 정보를 0과 1 사이의 실수값으로 변환한 후, 연속된 코돈 시퀀스들을 실수값을 목표로 산술 부호화의 반복적인 분할 처리로 동일한 아미노산 시퀀스가 되도록 변경한다. Arita 등[6]은 치환 Cipher 기반 퇴행성 유전 부호(degeneracy genetic code)를 이용하여 짧은 서명을 B. Subtilis 박테리아의 cDNA에 삽입하였다. 그러나 이 기법은 메시지 복호시 원본 시퀀스가 필요한 단점

을 가진다.

ncDNA(non-coding DNA) 기반 워터마킹으로 Yachie 등[7]은 시퀀스 정렬 기반으로 B. Subtilis 박테리아의 ncDNA 시퀀스의 다중 위치에 데이터를 삽입하는 방법을 제안하였다. 또한 Heider 등[8]은 DNA-Crypt 알고리즘[9]을 in vivo 기반의 ncDNA 시퀀스에 적용하였으며, fRNA(functional RNA) 분자 또는 조절 서열(regulatory sequence)의 ncDNA인 효모균(yeast) 내 cytoplasmic RNA I과 Escherichia coli의 lac promoter를 이용하였다. 그러나 이들은 실험으로부터 워터마크된 RNA와 Promoter와 원본 간의 차이는 없었으나, 다른 RNA 또는 Promoter들이 이에 반응하므로, 조절 영역 내에 워터마크를 삽입하는 것은 적합하지 않다고 결론을 맺었다.

cDNA 기반 워터마킹 기법을 살펴보면 다음과 같다. Heider 등[10]은 DNA-Crypt 알고리즘[9]을 이용하여 S. Cerevisiae 진핵 생물(eukaryotic organism) cDNA 상의 동일 코돈(codon)에 암호화된 정보를 삽입하였다. DNA-Crypt에서는 4개의 염기 {A, C, G, T}에 대한 2비트를 할당한 다음, 4중 중첩 코돈의 마지막 염기에 2비트씩 삽입한다. 여기서 이 알고리즘에서는 변이 정정을 위한 8/4 Hamming 부호와 n-time WDH 부호를 변이율, 서열 길이, 시간 안정성 변수를 입력으로 한 퍼지 제어를 통하여 적합한 변이 정정 부호를 선택한다. 그리고 AES, Blowfish, RSA 또는 OTP(One-Time pad)의 암호화 알고리즘에 의한 워터마크 이진 부호의 암호화가 가능하다. 그러나 암호화된 워터마크를 사용하더라도 삽입 위치가 공개되어 있으므로, 제3자에 의한 워터마크 제거가 용이하다. Shuhong 등[11]은 Arita의 유전 부호 기반으로 LSB(Least significant bit) 치환과 유사하게 코돈의 세 번째 위치의 심볼들을 비트에 따라 치환하였다. 이 기법은 동일한 아미노산을 가지는 코돈의 길이가 4일 경우에만 적용이 가능하다. Lee 등은 변이 강인성과 아미노산 부호 보존성을 가지기 위하여 코돈 원형 부호에 의한 워터마크 삽입[12]과 4/8 부-코돈 단위의 Lifting 기반 DWT 계수에 워터마크 삽입[13] 기법을 제안하였다. 그러나 이 기법들은 워터마크의 삽입 위치가 공개되므로, 보안성의 문제점을 가진다. 이외에도 Balado 등[14]은 ncDNA와 cDNA 시퀀스들 내에 변이에 견고한 최대 삽입 용량을 Shannon 이론에 기반으로 모델링하였다.

기존 cDNA 워터마킹 방법들은 치환, 삽입, 삭제 등의 변이에 대한 견고성과 보안성에 대하여 문제점을 가지고 있다. 따라서 본 논문에서는 이를 보완하기 위한 cDNA 워터마킹 기법을 제안한다.

3. 제안한 DNA 시퀀스 워터마킹

본 논문에서는 DNA 시퀀스의 워터마킹 기법을 제안하며, 다음과 같은 주요 특징을 가진다. 1) cDNA 서열에 이진 워터마크가 삽입된다. ncDNA 서열은 유전 체계의 주요한 기능을 조절하는데 포함되므로, ncDNA 보다 cDNA가 워터마크에 삽입 대상에 적합하다. 2) 워터마크된 코돈들이 아미노산 부호 보존성을 가지며, 워터마크는 포인트 변이 및 삽입/삭제 변이에 강인하다. 3) cDNA 서열의 길이에 상관없이 동일한 길이를 가지는 워터마크 키가 사용되며, 워터마크 키 및 삽입되는 코돈 위치의 보안성이 우수하다. 4) 워터마크 추출시 원본 시퀀스가 필요 없다. 제안한 워터마크 삽입 과정은 그림 1(a)에서와 같이 코돈 랜덤 맵핑 테이블에 의한 랜덤 원형 코돈 수치 변환, 다층 스케일 상의 Lipschitz regularity 기반 삽입 대상 코돈 선택, 코돈별 워터마크 삽입, 및 ncDNA와 워터마크된 cDNA와의 결합에 의한 워터

마크된 DNA 시퀀스 생성의 단계로 구성된다. 코돈 랜덤 맵핑 테이블은 워터마크 추출에 필요한 키로 저장된다. 워터마크는 그림 1(b)에서와 같이 삽입 과정과 유사한 과정으로 추출된다.

3.1 코돈 부호 변환

유전 부호는 코돈과 아미노산 동등성에 따라 구분되어 나타내어진다. 각 아미노산에 따라 코돈 길이가 1에서 6까지이다. 여기서 부호 영역을 시작하는 시작 코돈은 Met (ATG) (진핵 및 원핵 세포)이고, 부호 영역의 끝을 나타내는 종료 코돈은 Stp (TAA, TAG, TGA)이다. 시작 코돈 및 종료 코돈을 제외한 나머지 코돈들은 워터마크에 의하여 치환된다.

제안한 방법에서는 코돈을 6비트의 정수로 할당한 후, 이를 원형 형태의 각도로 변환한다[12]. 먼저 4개의 염기들은 $G=0, A=1, C=2, T=3$ 으로 놓은 다음, 각 코돈 $c=b_1b_2b_3$ 은 다항식 형태에 의하여 6비트 정수 c 로 $c=4^2 \times b_1 + 4^1 \times b_2 + 4^0 \times b_3$ 와 같이 할당한다. 그리고 정수 c 에 의하여 코돈을 원형 각도 $g(c)$ 로 $g(c)=(2\pi \times c)/64$ 와 같이 변환한다. 원형 각도 $g(c)$ 로부터 코돈 c 를 얻기 위한 역 변환에서는 먼저 c 를 구한 다음, 코돈의 세 염기를 $b_1 = \lfloor c/4^2 \rfloor$,

그림 1. 제안한 DNA 시퀀스의 워터마크 (a) 삽입 및 (b) 추출 과정

그림 2. (a) 아미노산 랜덤 정렬에 의하여 치환된 순회 원형 부호 (zero padding $M=0$ 일 때), (b) $M=3$ 때 워터마크 삽입 예; 워터마크 비트가 $w_i=1$ 일 때, $c_k^{(3)}=(ATC, CAG, GCC)$ 에 삽입

$b_2 = \lfloor c\%4^2/4 \rfloor$, $b_3 = (c\%4^2)\%4$ 와 같이 추출한다. 그리고 $b_1b_2b_3$ 로부터 코돈 c 가 구하여진다. 여기서 코돈의 정수 c 은 ASCII 코드 10진수 표현과 같이, 'G' 'A' 'T' 'C' 순에 의한 정수 할당으로 작은 정수값과 큰 정수값의 물리적인 의미는 없다. 코돈 부호 테이블 T 는 {G, A, T, C} 순서에 따라 달라지며, 총 24개의 테이블들 중 한 개가 사용될 수 있다.

코돈 원형 각도는 64개의 분해능을 가진다. 워터마크 삽입의 보안성 향상을 위하여 제안한 방법에서는 모든 아미노산들을 랜덤하게 나열된 후, 이들 아미노산의 중첩 코돈들은 랜덤하게 선택한 후, M 개 zero 단위로 값이 할당된다. 예를 들어, 랜덤하게 나열된 아미노산들이 $S = \{S_1, S_2, \dots, S_{20}\}$, $S_i = \{c_k | k \in [0, |S_i| - 1]\}$ for $i \in [1, 20]$ 라 할 때, i 번째 치환된 아미노산 S_i 의 중첩 코돈 부호는

$$c_k \rightarrow \hat{c}_k = \begin{cases} (l(k) + (\sum_{j=1}^{i-1} (M+1)|S_j| + M_0))\%R_i & \text{if } i > 1 \\ l(k) + M_0, & \text{if } i = 1 \end{cases}, \quad k \in [0, |S_i| - 1] \quad (1)$$

와 같다. 여기서 $l(k)$ 은 $[1, (M+1)|S_i|]$ 에서 랜덤하게 선택된 값이며, M_0 는 아미노산 코돈의 시작 부호로 $1 \leq M_0 \leq M$ 에서 랜덤하게 선택된 값이다. 이에 따라 기준 코돈 부호 테이블 T 의 코돈 부호 c 는 아미노산

랜덤 나열에 의하여 $\hat{c}$ 로 치환된다; $c \rightarrow \hat{c}$. 그리고 치환된 코돈 부호의 원형 각도 $g(c)$ 는 $\hat{c}$ 에 의하여 다시 계산되어진다.

제안한 방법에서는 T 와 랜덤 부호 테이블 $\hat{T}$ 에 대한 랜덤 맵핑 테이블 P 를

$$P = \hat{T} - T \quad (2)$$

와 같이 코돈 부호와 치환 부호와의 차이로 정의하며, 두 테이블은 워터마크 키로 저장한다. 여기서 PKI(public-key infrastructure) 프로토콜 상에서 코돈 부호 테이블 T 는 공용키(public key)로, 랜덤 맵핑 테이블 P 는 개인키로 사용될 수 있다. 그림 2(a)는 코돈 부호 테이블에 의하여 생성된 원형 부호와 $M=0$ 일 때 아미노산 랜덤 정렬에 의하여 치환된 원형 부호를 보여주고 있다. 이 그림으로부터 아미노산의 중첩 코돈 단위로 치환되므로 {S,R,L}의 연속적인 부호 할당이 가능하고, 랜덤 맵핑 테이블에 의하여 원래의 원형 부호로 역치환되며 워터마크의 보안성을 가지므로 워터마크 신호 처리에 용이함을 볼 수 있다.

3.2 워터마크 삽입

제안한 방법에서는 부호 영역 코돈 서열 C 에서

삽입 대상 코돈 그룹 $\mathbf{c}_k^{(3)} = (c_{k-1}, c_k, c_{k+1})$ 들을 선택한 후, 중심 코돈 c_k 과 좌우 인접 코돈들에 워터마크 1비트씩 $w_i = \{0, 1\}$ 삽입한다. 이 때 워터마크 비트 삽입은 랜덤 원형 각도에 의하여 이루어진다.

3.2.1 삽입 대상 중심 코돈 탐색

삽입 대상 코돈들은 랜덤 맵핑 테이블 $\mathbf{P}$ 에 의하여 치환된 원형 각도 서열의 특이점들을 이용하여 삽입 대상의 중심 코돈 $\mathbf{c}^* = f(g(\mathbf{C})|\mathbf{P})$ 을 찾는다. 이 때 특이점들은 Mallat 웨이블릿 기반[15]의 Lipschitz 정칙 상수에 의하여 검출된다. Mallat 이론에 따라, $g(c)$ 가 $j=0$ 에서 국부 최대 계수를 $|W_{2^0}^{LM}(g|\mathbf{P})| > 0$ 와 같이 가지며, 지역적으로 특이점을 가지는 신호와 분산 σ^2 의 가우시안 함수와의 컨볼루션으로 가정될 때, $g(c)|\mathbf{P}$ 의 Lipschitz 정칙 상수는 구간 $[a, b]$ 에서

$$|W_{2^j}^{LM}(g|\mathbf{P})| \leq K2^j s_0^{\alpha-1} \text{ where } s_0 = \sqrt{2^{2j} + \sigma^2} \text{ and } \{(|W_{2^j}^{LM}(g|\mathbf{P})|)_{1 \leq j \leq J} \forall c \in [a, b]\} \quad (3)$$

를 만족하는 α 집합의 상한치로 정의된다. 이 때, 세 변수 K, α, σ 들은 steepest decent 알고리즘에 의하여

$$\argmin_{K, \alpha, \sigma} \sum_{j=1}^J (\log_2 |W_{2^j}^{LM}(g|\mathbf{P})| - \log_2 K - j - \frac{\alpha-1}{2} \log_2 (2^{2j} + \sigma^2))^2 \quad (4)$$

를 최소화함으로써 구하여진다. 이 때 Lipschitz 정칙 상수가 $\alpha_k < th$ ($-1 < th \leq 1$)인 위치에 해당되는 $g(c_k^*)$ 와 좌우 인접 각도들을 삽입 대상 코돈 쌍으로

$$g(\mathbf{c}^*) = f(g(\mathbf{C})|\mathbf{P}) = (g(c_{k-1}), (g(c_k^*), g(c_{k+1}))|_{\alpha_k < th})_{|g(c_k) \in Z} \quad (5)$$

와 같이 결정된다.

3.2.2 워터마크 비트 삽입

제안한 방법에서는 워터마크 비트 w_i 를 삽입 대상

코돈 그룹 (Triplet) $\mathbf{c}_i^{(3)} = (c_{i-1}, c_i^*, c_{i+1})$ 의 원형 각도 $(g(c_{i-1})|\mathbf{P}, g(c_i^*)|\mathbf{P}, g(c_{i+1})|\mathbf{P})$ 에 각각 삽입한다. 랜덤 맵핑 테이블 $\mathbf{P}$ 에 의한 원형 각도들은

$$\begin{aligned} g(S_{k-1})|\mathbf{P} &= \{g(c_{k-1,1}), g(c_{k-1,2}), \dots, g(c_{k-1,|S_{k-1}|})\}|\mathbf{P} \\ g(S_k)|\mathbf{P} &= \{g(c_{k,1}), g(c_{k,2}), \dots, g(c_{k,|S_k|})\}|\mathbf{P} \\ g(S_{k+1})|\mathbf{P} &= \{g(c_{k+1,1}), g(c_{k+1,2}), \dots, g(c_{k+1,|S_{k+1}|})\}|\mathbf{P} \end{aligned} \quad (6)$$

와 같다. 중심 코돈의 원형 각도 $g(c_k^*)|\mathbf{P}$ 는 워터마크 비트 w_i 에 의하여

$$g'(c_k^*)|\mathbf{P} = \begin{cases} \max(g(S_k)|\mathbf{P}), & \text{if } w_i = 1 \\ \min(g(S_k)|\mathbf{P}), & \text{if } w_i = 0 \end{cases} \quad (7)$$

와 같이 변경된다. 랜덤 맵핑 테이블에 의한 코돈 정수 변환에서는 중첩 코돈들이 하나의 그룹으로 묶여지며, 이들의 수치는 그룹 내에서 연속적이다. 따라서 중첩 코돈 내의 최대 수치 및 최소 수치는 $\max(g(S_k)|\mathbf{P}) = g(c_{k,1})|\mathbf{P}$ 와 $\max(g(S_k)|\mathbf{P}) = g(c_{k,|S_k|})|\mathbf{P}$ 와 같다.

좌우 코돈의 원형 각도들은 워터마크 비트 w_i 에 의하여 기준점과의 두 방향에 따른 최대 및 최소 원형 각도 거리로 치환된다. 먼저 그림 2(b)의 순회 원형 코드 상에서 왼쪽 인접 코돈의 $g(S_{k-1})|\mathbf{P}$ 에는 시계 방향으로 기준점 R_k 와의 거리 $d \cup$ 와 오른쪽 인접 코돈의 $g(S_{k+1})|\mathbf{P}$ 에는 반시계 방향으로 기준점 R_k 와의 거리 $d \cup$ 가

$$d \cup (R_k, g(c_{k-1,j})|\mathbf{P}) = |g(c_{k-1,j})|\mathbf{P} - R_k|, \forall j \in [1, |S_{k-1}|] \quad (8)$$

$$d \cup (R_k, g(c_{k+1,j})|\mathbf{P}) = 2\pi - |g(c_{k+1,j})|\mathbf{P} - R_k|, \forall j \in [1, |S_{k+1}|] \quad (9)$$

와 같이 정의된다. 여기서 기준점 R_k 는 중심 코돈에 대한 중첩 코돈들의 평균 원형 각도 $R_k = \frac{1}{|S_k|} \sum_{j=1}^{|S_k|} (g(c_{k,j})|\mathbf{P})$ 이다. 왼쪽 인접 코돈의 원형 각도 $g(c_{k-1})|\mathbf{P}$ 는 워터마크 비트 w_i 에 의하여 시계 방향으로 기준점과의 최대

표 1. 실험에 사용된 DNA 시퀀스 (NCBI, <http://www.ncbi.nlm.nih.gov/>)

GenBank accession number	Gene	Organism	CDS 염기수	Codon 수	$H(I N)$
NM_000520	HEXA	Homo sapiens	1590bp	530	293.92
NM_001145	ANG	Homo sapiens	444bp	148	78.84
NM_001179490	CCT2	Saccharomyces cerevisiae S288c	1584bp	526	289.15
JQ670900	bp1	Bacillus subtilis	1407bp	469	257.21

Identities = 401/446 (90%), Positives = 415/446 (93%)		Identities = 148/148 (100%)	
001	ATGGTGATGGGCTGGGCGTTTGTGTTGGTCTTCGTGCTGGGCTGACCCACCGA	001	MVMGLGVLLLVFVLGLGLTPPTLAQDNSRYTHFLTQHYDAKPQGRDDRYCESIMRRRLTSPCK
001	ATGGTGATGGGCTGGGCGTTTGTGTTGGTCTTCGTGCTGGGCTGACCCACCTA	001	MVMGLGVLLLVFVLGLGLTPPTLAQDNSRYTHFLTQHYDAKPQGRDDRYCESIMRRRLTSPCK
065	CCCTGGCTCAGGATAAATCCAGGTACACACACTTCCTGACCCAGCACTATGATGCCAAACACA	065	DINTFIHGNKRSIKAICENKNGNPHRENLRISKSSFQVTTCKLHGGSPWPPCQYRATAGFRNVV
065	CGCTGGCTCAGGATAAATCCAGGTACACACACTTCCTGACCCAGCACTATGATGCCAAACACA	065	DINTFIHGNKRSIKAICENKNGNPHRENLRISKSSFQVTTCKLHGGSPWPPCQYRATAGFRNVV
129	GGGCCGGGATGACAGATCTGTGAAAGCATCATGAGGAGACGGGGCTGACCTCACCC-TGCAA	129	VACENGLPVHLDQSFRRP*
129	GGGTCGTGATGACCGTTATTGTGAAAGCATTATGAGGAGAGGGGTTTAAACCA-GCCCTTGTAA	129	VACENGLPVHLDQSFRRP*
(a)		(b)	
Identities = 388/444 (87%), Positives = 415/444 (93%)		Identities = 148/148 (100%)	
001	ATGGTGATGGGCTGGGCGTTTGTGTTGGTCTTCGTGCTGGGCTGACCCACCGA	001	MVMGLGVLLLVFVLGLGLTPPTLAQDNSRYTHFLTQHYDAKPQGRDDRYCESIMRRRLTSPCK
001	ATGGTAATGGGACTAGGAGTATTGTGTTAGTATTCTGACTAGGACTTGGCTTACTCTCCCAA	001	MVMGLGVLLLVFVLGLGLTPPTLAQDNSRYTHFLTQHYDAKPQGRDDRYCESIMRRRLTSPCK
065	CCCTGGCTCAGGATAAATCCAGGTACACACACTTCCTGACCCAGCACTATGATGCCAAACACA	065	DINTFIHGNKRSIKAICENKNGNPHRENLRISKSSFQVTTCKLHGGSPWPPCQYRATAGFRNVV
065	CAGTACAGGATAAATCCAGGTACACCCACTTCCTGACCCAGCACTATGATGCCAAACCCAA	065	DINTFIHGNKRSIKAICENKNGNPHRENLRISKSSFQVTTCKLHGGSPWPPCQYRATAGFRNVV
129	GGGCCGGGATGACAGATCTGTGAAAGCATCATGAGGAGACGGGGCTGACCTCACCCTGCAAA	129	VACENGLPVHLDQSFRRP*
129	GGGTCGTGATGACAGTACTGTGAAAGCATCATGAGGAGTCGTGGGCTGACCTGCGCGTGCAAA	129	VACENGLPVHLDQSFRRP*
(c)		(d)	

그림 3. ANG(NM_001145)에 대하여 제안한 방법의 (a) 워터마크된 염기 서열과 (b) 아미노산 서열 및 Heider 방법의 (c) 워터마크된 염기 서열과 (d) 아미노산 서열

또는 최소 거리에 있는 원형 각도로

$$g'(c_{k-1})|P = \begin{cases} \max(d \cup (R_k, g(S_{k-1})|P)), & \text{if } w_k = 1 \\ \min(d \cup (R_k, g(S_{k-1})|P)), & \text{if } w_k = 0 \end{cases} \quad (10)$$

와 치환되고, 오른쪽 인접 코돈의 원형 각도 $g(c_{k+1})|P$ 는 반시계 방향으로 기준점과의 최대 및 최소 거리에 있는 원형 각도로

$$g'(c_{k+1}^*)|P = \begin{cases} \max(d \cup (R_k, g(S_{k+1})|P)), & \text{if } w_k = 1 \\ \min(d \cup (R_k, g(S_{k+1})|P)), & \text{if } w_k = 0 \end{cases} \quad (11)$$

와 같이 치환된다. 위의 워터마크된 원형 각도 $(g'(c_{k-1})|P, g'(c_k^*)|P, g'(c_{k+1})|P)$ 는 치환된 코돈 부호 $\hat{c}'$ 와 랜덤 맵핑 테이블 P 로부터 워터마크된 코돈 부호 $\hat{c}'$ 로 변환된다.

$$\hat{c}' = \hat{c}' + P(\hat{c}') \quad \text{where}$$

$$\hat{c}' = \begin{cases} \frac{64}{2\pi}(g'(c)|P), & \text{if } g'(c)|P \geq 0 \\ \frac{64}{2\pi}(2\pi + (g'(c)|P)), & \text{if } g'(c)|P < 0 \end{cases} \quad (12)$$

그리고 최종 워터마크된 코돈 $\hat{c}_k = (b_1 b_2 b_3)$ 는 $b_1 = \lfloor \hat{c}'/4^2 \rfloor$, $b_2 = \lfloor \hat{c}'\%4^2/4 \rfloor$, $b_3 = (\hat{c}'\%4^2)\%4$ 에 의하여 생성된다. 이와 같은 방법에 의하여 선택된 모든 삽입 대상 코돈 집합들에 워터마크 비트가 각각 삽입된다.

3.3 워터마크 추출

전송 또는 의심스러운 DNA 시퀀스로부터의 워터마크 추출 과정은 삽입 과정에서 저장된 랜덤 맵핑 테이블 P 와 중심 코돈 인덱스 I 에 의하여 그림 1(b)에서와 같이 수행된다. 추출 과정에서는 cDNA 시퀀스들의 모든 코돈 $\hat{C}$ 들을 랜덤 맵핑 테이블 P 에 의하여 원형 각도 $g(\hat{C})|P$ 로 변환한 후, DWT hard 임계화 된 원형 각도 서열의 국부 최대치에 해당되는 중심 코돈들 $\hat{c}^* = f(g(\hat{C})|P)$ 를 탐색한다. 그리고 삽입 과정에서 저장된 중심 코돈의 인덱스 I 와 탐색된 중심 코돈의 인덱스 $\hat{I}$ 와의 동기화한 다음, 동기화된 인덱스 I 로부터 삽입 대상 코돈 그룹들을 구한다. 삽입 및 삭제 변이에 의하여 시퀀스들이 쉬프트되므로, I 와 $\hat{I}$ 를 차례로 매칭하면서 불일치한 코돈부터 한 코돈 씩 쉬프트함으로써 동기화가 수행된다.

4. 실험 결과

본 실험에서는 표 1에서와 같이 NCBI에서 제공하는 DNA 서열을 사용하여 제안한 방법과 Heider [9,10]의 DNA-Crypt 기반 워터마킹 방법을 비교 평가하였다. Heider의 DNA-Crypt 실험에서는 오류 정정을 위한 WDH(5) 부호화가 사용되었으며, 4중,

그림 4. (a) 제안한 방법에서 문턱치 th 에 대한 워터마크 용량과 (b) Heider 방법의 $WDH(n)$ 부호화에서 n 에 대한 워터마크 용량

6중 중첩 코돈에 대하여 2비트씩 워터마크를 삽입하였다. 본 실험은 Matlab Bioinformatics toolbox 기반으로 수행되었다.

4.1 아미노산 부호 보존성

cDNA 워터마킹은 의도적인 침묵 변이에 해당되므로, 워터마크에 의하여 염기가 변하더라도 아미노산의 부호 변화는 없다. 그림 3은 원본 ANG 시퀀스와 제안한 방법 및 Heider 방법에 의하여 워터마크된 ANG 시퀀스들을 보여주고 있다. 이 그림을 살펴보면, 제안한 방법과 Heider 방법에 의하여 워터마크된 ANG 염기 서열들은 원본 염기 서열에 비하여 10%와 13% 정도 차이가 나지만, 두 방법 모두 ANG 아미노산 서열들은 원본 서열과 일치함을 볼 수 있다. 따라서 두 방법 모두 아미노산 보존성이 유지됨을 알 수 있다.

4.2 용량성

제안한 방법의 워터마크 용량은 랜덤 맵핑 테이블 $\mathbf{P}$ 에 의한 첫 번째 스케일의 국부 계수 최대치 $W_{2^0}^{LM}(g|\mathbf{P})$ 와 Lipschitz regularity α 의 문턱치 th ($-1 < th \leq 1$)에 의하여 결정된다. 중심 코돈의 개수가 $|C^{(3)}|$ 이고 전체 코돈 개수가 $|C| = N$ 일 때, 제안한 방법의 워터마크 용량 C 는

$$C = \frac{|C^{(3)}|}{N-2} \leq \frac{1}{3} \text{ [bit/codon]} \quad (13)$$

와 같다. 여기서 $N-2$ 는 시작 및 종결 코돈을 제외한

코돈 개수를 나타낸다. Heider의 방법은 워터마크 비트를 $WDH(n)$ 에 의하여 부호화한 다음, 코돈 길이 $|S|$ 가 4 이상인 아미노산의 코돈들에 각각 2비트씩 차례로 삽입한다. Heider 방법의 워터마크 용량 C 는

$$C = \lfloor \frac{1}{n} \times \frac{2|Z|}{N-2} \rfloor \text{ [bit/codon]} \quad (14)$$

와 같다. 여기서 $|Z|$ 는 아미노산 $\{G, A, V, T, R, P, L, S\}$ 으로 번역되는 코돈의 개수를 나타내고, n 은 반복 횟수를 나타낸다. 반복 횟수 n 은 제안한 방법의 워터마크 용량과 유사한 용량이 되도록 설정하였다.

그림 4(a)는 제안한 방법의 Lipschitz regularity α 의 문턱치 th 에 대한 워터마크 용량 C 과 그림 4(b)는 Heider 방법의 반복 횟수 n 에 대한 워터마크 용량 C 를 보여주고 있다. 제안한 방법에서는 $th=1$ 일 때 최대 용량을 가진다. 본 실험에서는 $th=0$ 이고 $\alpha < 0$ 보다 작은 국부 계수 최대치를 중심 코돈으로 선택하였으며, 이 때 워터마크 용량은 평균 0.137이다. Heider 방법에서는 반복 횟수가 클수록 변이에 대한 강인성을 가지나, 워터마크 용량이 작아진다. 본 실험에서는 Heider 방법의 강인성과 제안한 방법과 유사한 용량을 가지도록 $n=5$ 로 하였다. 이 때 워터마크 용량은 평균 0.088으로 제안한 방법보다 약 0.049정도 낮다.

4.2 변이 강인성

변이 실험에서는 워터마크된 코돈 서열 상에 포인트 변이, 삽입 및 삭제 변이를 임의의 위치에 동시적으로 발생시켰다. 전체 염기 개수에 대한 변이된 염

그림 5. 제안한 방법 및 Heider 방법에 의하여 동시 변이(포인트 변이, 삽입/삭제 변이)된 (a) HEXA(NM_000520)와 (b) ANG(NM_001145)에서 추출된 위터마크의 BER

기 개수 비율이 r 일 때, 포인트 변이는 $2r$, 삽입 변이는 r , 삭제 변이는 r 의 비율로 동시에 발생되며, r 은 1-20까지 가변되었다. 예를 들어, r 이 5일 때, 5%의 삽입 및 삭제 변이와 10%의 포인트 변이가 동시에 발생된다. 삽입 및 삭제 변이 시 해독틀 복원을 위하여 본 실험에서는 위터마크된 코돈 서열 또는 공개된 코돈 서열을 기준으로 변이된 서열을 Smith-Waterman 알고리즘에 의하여 재배열한 다음, 위터마크를 추출하였다.

실험에 사용된 평가 척도로는 추출된 위터마크 $\hat{\mathbf{w}}$ 의 BER(bit error rate), $e(\mathbf{w}, \hat{\mathbf{w}})$ 이다. 그림 5는 위터마크된 DNA 서열에 동시 변이를 랜덤하게 5,000번 발생하여 추출된 BER들의 평균을 보여주고 있다. 이 결과를 살펴보면, 모든 동시 변이율에서 제안한 방법이 기존 방법에 비하여 BER이 최소 0.003에서 최대 0.14 정도 낮음을 알 수 있다. 일반적으로 $e(\mathbf{w}, \hat{\mathbf{w}}) < 0.2$ 일 경우, 두 이진 위터마크가 유사하다고 판단된다. $e(\mathbf{w}, \hat{\mathbf{w}}) = 0.2$ 일 때의 동시 변이율 r 을 최대 저항 세기 $r_{\max}$ 라 할 때, 제안한 방법은 기존 방법에 비하여 $r_{\max}$ 가 최소 18에서 최대 65 정도 높음을 알 수 있었다.

본 실험에서는 위터마크 검출 성능을 판단하기 위하여 동시 변이율 10%와 20%일 때 FP(False positive probability) P_{FP} 와 TP(True positive probability), P_{TP} 에 의한 ROC(Receiver operating characteristic) 곡선을 계산하였다. 검출 문턱치 T_1, T_2 에 대한 P_{FP} 와 P_{TP} 는 다음과 같이 정의된다.

$$P_{TP}(T_1) = \Pr[e(\mathbf{W}', \hat{\mathbf{W}}) < T_1 \mid \mathbf{W}' \simeq \hat{\mathbf{W}}, \mathbf{W}' \neq \mathbf{W}],$$

(Detect H_1 when H_0 is true) (15)

$$P_{TP}(T_2) = \Pr[e(\mathbf{W}, \hat{\mathbf{W}}) < T_1 \mid \mathbf{W} \simeq \hat{\mathbf{W}}],$$

(Detect H_0 when H_0 is true) (16)

여기서 T_1, T_2 는 $0 < T_1, T_2 < 1$ 으로 가변된다. 각 코돈 서열에 대한 ROC 곡선은 그림 6에서와 같다. 이 결과를 살펴보면, 동시 변이율이 높을수록 제안한 방법이 기존 방법에 비하여 위터마크 검출 성능이 우수함을 볼 수 있다.

4.3 보안성 분석

DNA 위터마킹 기법과 DNA 코돈 서열이 공개되더라도, 위터마크의 추출이 어려워야 한다. 그러나 기존 DNA 위터마킹 방법에서는 위터마크의 보안성에 대한 언급과 대응이 없으므로, 제3자에 의하여 쉽게 위터마크 추출이 가능하다. 제안한 방법에서는 공개되어 있는 코돈 서열 내에 랜덤 맵핑 테이블 $\mathbf{P}$ 에 의한 부호 변환과 Lipschitz regularity에 의한 삽입 그룹 선택함으로써 코돈 부호 및 위터마크 삽입 위치의 예측이 어렵다. 특히 하나의 코돈 서열이라도 $\mathbf{P}$ 에 의하여 삽입 코돈의 개수와 위터마크의 길이가 달라진다. 즉, 위터마크 키인 $\mathbf{P}$ 가 없을 경우, 제3자는 삽입 위치와 개수를 정확히 알아야 위터마크 추출이 가능하다. 또한 기존 생물학적 DNA 암호화에서와 같이 PCR 프라이머 쌍 또는 특정 프로브를 암호키로 삽입된 서열 위치를 암호함으로써 생물학적 DNA 위

터마킹으로 확장될 수 있다.

랜덤 맵핑 테이블 $\mathbf{P}$ 의 경우의 수 $N(\mathbf{P})$ 는 수식 (4)에 의하여

$$N(\mathbf{P}) = (20!) (\prod_{i=1}^{20} (M+1) |S_i|) M_0 \quad (17)$$

와 같이 아미노산 랜덤 정렬의 개수, 아미노산 내에 중첩 코돈의 랜덤 치환의 개수 및 시작 코돈 부호 개수 $M_0 = M+1$ 에 의하여 계산되어진다. 따라서 $N(\mathbf{P})$ 는

$$\lceil \log_2 N(\mathbf{P}) \rceil = \lceil \log_2 (20! (\prod_{i=1}^{20} (M+1) |S_i|) (M+1)) \rceil \geq 90 + \lceil 21 \log_2 M \rceil \quad (18)$$

의 비트수로 표현될 수 있다.

워터마크의 보안성은 $\mathbf{P}$ 의 비트 길이 $\lceil \log_2 N(\mathbf{P}) \rceil$ 에 의하여 결정되며, $\lceil \log_2 N(\mathbf{P}) \rceil$ 는 위 식에 따라 zero padding의 개수 M 에 의하여 결정된다. 원하는 키의 길이가 N_K^* 일 때, M 은

$$M = \lfloor 2^{\frac{N_K^* - 90}{21}} \rfloor \quad (19)$$

으로 설정된다. 만약 256비트의 대칭키와 동일한 보안성을 유지하고자 하면, $M=239$ 이다.

랜덤 맵핑 테이블 공간 $\mathbf{X} = \{P_1, P_2, \dots, P_{N(\mathbf{P})}\}$ 에서 임의의 랜덤 맵핑 테이블 P_i 의 pmf(probability mass function)는 $\Pr_X(P_i) = 1/N(\mathbf{P})$ 이다. 따라서 $\mathbf{P}$ 의 uncertainty를 측정하기 위한 $\mathbf{X}$ 의 엔트로피는

$$H(\mathbf{X}) = - \sum_{i=1}^{N(\mathbf{P})} \Pr_X(P_i) \log_2 \Pr_X(P_i) = \log_2 N(\mathbf{P}),$$

$$(H(\mathbf{X}) = 20! \prod_{i=1}^{20} |S_i|, \text{ for } M=0) \quad (20)$$

와 같다. zero padding의 개수 M 에 대한 엔트로피

그림 7. zero padding의 개수 M 에 대한 랜덤 맵핑 테이블 공간 $\mathbf{X}$ 의 엔트로피 $H(\mathbf{X})$

$H(\mathbf{X})$ 는 그림 7에서와 같다. $H(\mathbf{X})$ 가 클수록 랜덤 양이 크며 키 값을 예측하기 위하여 많은 시도($\alpha^{H(\mathbf{X})}$, $\alpha > 1$)가 필요하다.

워터마크 키인 랜덤 맵핑 테이블 $\mathbf{P}$ 를 알지 못할 경우, DNA 코돈 서열 $\mathbf{C}$ 내에 삽입 개수와 삽입 위치를 알아야 한다. 랜덤 맵핑 테이블 $\mathbf{P}$ 에 의하여 랜덤 수치 서열에서 웨이블릿 기반 Lipschitz regularity 상수에 의한 삽입 개수 및 위치는 $\mathbf{P}$ 와 확률적으로 독립이다.

코돈 개수가 $|S|$ 일 때, 삽입되는 중심 코돈 개수 n 은 평균 워터마크 용량에 따라 $\lfloor 0.05|S| \rfloor$, $\lfloor |S|/3 \rfloor = [a, b]$ 범위에 가우시안 확률 분포 $N(m, \sigma^2)$ ($m=0.137$, $\sigma=0.05$)를 가진다. 따라서 삽입 개수 n 의 정규화된 이산 확률 $\Pr_N(n)$ 은

$$\Pr_N(n) = N(n; m, \sigma^2) / \sum_{n=a}^b N(n; m, \sigma^2), \text{ for } n \in [a, b] \quad (21)$$

이다. 삽입 개수가 n 일 때, 코돈 서열에서 n 개의 코돈이 선택될 확률 $\Pr_I(I|n)$ 은

$$\Pr_I(I|n) = 1 / \binom{|S|}{n} = \frac{n! (|S| - n)!}{|S|!} \quad (22)$$

이다. 따라서 삽입 개수에 의한 삽입 위치 검출 확률 $\Pr(I|N)$ 에 의하여 조건 엔트로피 $H(I|N)$ 는

$$H(I|N) = \sum_{n \in N} \Pr_N(n) H(I|N=n) = \sum_{n \in N} \Pr_N(n) (\log_2 \binom{|S|}{n}) \quad (23)$$

이다. DNA 서열에 대한 조건 엔트로피 $H(I|N)$ 는 매우 높으며, 코돈 길이가 길수록 $H(I|N)$ 가 높아진다. 실험으로부터 $H(I|N)$ 가 코돈 서열 길이 N 에 대한 1차 선형에 가까움을 알 수 있었다.

5. 결 론

DNA 워터마킹의 주요한 조건으로는 변이의 강인성, 아미노산 부호 보존성 및 워터마크의 보안성이다. 본 논문에서는 위의 세 조건을 만족하는 코돈 랜덤 원형 각도 기반의 DNA 워터마킹 기법을 제안하였다. 실험에서는 제안한 방법과 기존의 WDH 기반의 DNA-Crypt 워터마킹과 변이 강인성을 비교 평가하였다. 평가 결과로부터 제안한 방법이 기존 방법에 비하여 최소 0.003에서 최대 0.14 정도 BER이 낮았으며, 통계적 검출이 보다 우수함을 확인하였다. 본 논문에서 제안한 방법은 기존 멀티미디어 워터마

킹[16],[17]과 결합하여 새로운 유전자 기반의 워터마킹 이론이 제시될 수 있을 것이다.

참 고 문 헌

- [1] T. Kazuo, O. Akimitsu, and S. Isao, "Public-Key Systems using DNA as a One-Way Function for Key Distribution," *BioSystems*, Vol. 81, No. 1, pp. 25-29, 2005.
- [2] A. Gehani, T. LaBean, and J. Reif, "DNA-based Cryptography," *Aspects of Molecular Computing, Lecture Notes in Computer Science*, Vol. 2950, pp. 34-50, 2004.
- [3] C.T. Clelland, V. Risca, and C. Bancroft, "Hiding Messages in DNA Microdots," *Nature*, Vol. 399, No. 6736, pp. 533-534, June 1999.
- [4] A. Leier, C. Richter, W. Banzhaf, and H. Rauhe, "Cryptography with DNA Binary Strands," *Biosystems*, Vol. 57, Issue 1, pp. 13-22, 2000.
- [5] B. Shimanovsky, J. Feng, and M. Potkonjak, "Hiding Data in DNA," *Proc. of the 5th Intl. Workshop in Information Hiding*, pp. 373-386, 2002.
- [6] M. Arita, "Writing Information into DNA," *Molecular Computing, Lecture Notes in Computer Science*, Vol. 2950, pp. 23-35, 2004.
- [7] N. Yachie, K. Sekiyama, J. Sugahara, Y. Ohashi, and M. Tomita, "Alignment-based Approach for Durable Data Storage Into Living Organisms," *Biotechnol. Prog.* Vol. 23, No. 2, pp. 501-505, 2007.
- [8] D. Heider and A. Barnekow, "DNA Watermarks in Non-Coding Regulatory Sequences," *BMC Bioinformatics*, Vol. 2, No. 125, doi:10.1186/1756-0500-2-125, 2009.
- [9] D. Heider and A. Barnekow, "DNA-Based Watermarks using the DNA-Crypt Algorithm," *BMC Bioinformatics*, Vol. 8, No. 176, doi:10.1186/1471-2105-8-176, 2007.
- [10] D. Heider and A. Barnekow, "DNA Watermarks - A Proof of Concept," *BMC Bioinformatics*, Vol. 9, No. 40, doi:10.1186/1471-2199-9-40, 2008.
- [11] J. Shuhong and R. Goutte, "Code for Encryption Hiding Data Into Genomic DNA of Living Organisms," *International Conference on Signal Processing (ICSP)*, pp. 2166-2169, 2008.
- [12] S.H. Lee and K.R. Kwon, "Robust DNA Watermarking Based on Coding DNA Sequence," *The Institute of Electronics Engineers of Korea*, Vol. 49-CI, No. 2, pp. 123-133, 2012.
- [13] S.H. Lee, S.G. Kwon, and K.R. Kwon, "A Robust DNA Watermarking in Lifting Based 1D DWT Domain," *The Institute of Electronics Engineers of Korea*, Vol. 49, No. 10, pp. 91-101, 2012.
- [14] F. Balado, "On the Embedding Capacity of DNA Strands Under Insertion, Deletion and Substitution Mutations," *SPIE Media Forensics and Security XII*, Vol. 7541, 2010.
- [15] S. Mallat and S. Zhong, "Characterization of Signals from Multiscale Edges," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 14, Issue 7, pp. 710-732, 1992.
- [16] B.J. Jang, S.H. Lee, and K.R. Kwon, "Active Video Watermarking Technique for Infectious Information Hiding System," *Journal of Korea Multimedia Society*, Vol. 15, No. 8, pp. 1017-1030, 2012.
- [17] W.J. Kim, T.Y. Seung, S.H. Lee, and K.R. Kwon, "An Information Security Scheme Based on Video Watermarking and Encryption for H.264 Scalable," *Journal of Korea Multimedia Society*, Vol. 15, No. 3, pp. 299-311, 2012.

이 석 환

1999년 경북대학교 전자공학과
졸업(공학사)
2001년 경북대학교 전자공학과
졸업(공학석사)
2004년 경북대학교 전자공학과
졸업(공학박사)

2005년~현재 동명대학교 정보보호학과 부교수
관심분야: 워터마킹, DRM, 영상신호처리

권 성 군

1996년 경북대학교 전자공학과
졸업 (공학사)
1998년 경북대학교 전자공학과
졸업 (공학석사)
2002년 경북대학교 전자공학과
졸업 (공학박사)

2002년~현재 삼성전자 무선통신 사업부 연구원
관심분야: 영상처리, 영상통신, 정보보호

권 기 룡

1986년 경북대학교 전자공학과
학사 졸업(공학사)
1990년 경북대학교 전자공학과
석사 졸업(공학석사)
1994년 경북대학교 전자공학과
박사 졸업(공학박사)

2000년~2001년 Univ. of Minnesota, Post-Doc.
1996년~2005년 부산외국어대학교 디지털정보공학부
부교수
2006년~현재 부경대학교 전자컴퓨터정보통신공학부
교수
2008년~현재 한국멀티미디어학회 국제담당부회장
관심분야: 멀티미디어 정보보호, 영상처리, 웨이브릿 변환