

논문 2015-52-7-7

연속적 차분 확장 기반 가역 DNA 워터마킹

(Consecutive Difference Expansion Based Reversible DNA Watermarking)

이 석 환*, 권 기 룡**

(Suk-Hwan Lee and Ki-Ryong Kwon[©])

요 약

대용량의 DNA 정보 저장, DNA 서열 저작권 보호를 위한 DNA 워터마킹, 및 비밀 통신을 위한 DNA 스테가노그래피에 대한 관심이 증대되면서, 원본 DNA 서열의 기능 유지와 복원이 가능한 가역성 DNA 워터마킹이 필요하다. 본 논문에서는 비부호영역 DNA 서열을 이용한 DE(Difference expansion) 기반 가역 DNA 워터마킹 기법을 제안한다. 가역 DNA 워터마킹에서는 생물학적 기능 변경이 없고, 문자 형태의 서열 내에 대용량의 데이터를 은닉하여야 하며, 원본 DNA 서열이 복원되어야 한다. 제안한 방법에서는 문자 서열을 십진수 형태의 수치계수로 변환한 다음, 인접 수치 계수 쌍의 DE 기반 다중비트 은닉 방법(DE-MBE, DE based multiple bits embedding)과 이전 은닉 수치계수를 예측으로 한 연속 DE 기반 다중비트 은닉 방법들(C-DE-MBE, consecutive DE based multiple bits embedding)에 의하여 워터마크가 은닉된다. 은닉 과정에서는 워터마크된 서열에 의하여 부호영역을 나타내는 허위 시작코돈 발생을 방지하기 위하여 비교 탐색을 수행한다. 실험 결과로부터 제안한 방법이 기존 방법에 비하여 높은 은닉 용량을 가지며, 허위 시작코돈이 발생되지 않으며, 기준 서열없이 원본 DNA 서열이 복원됨을 확인하였다.

Abstract

Of recent interests on high capacity DNA storage, DNA watermarking for DNA copyright protection, and DNA steganography for DNA secret communication are augmented, the reversible DNA watermarking is much needed both to embed the watermark without changing the functionality of organism and to perfectly recover the host DNA sequence. In this paper, we address two ways of DE based reversible DNA watermarking using noncoding DNA sequence. The reversible DNA watermarking should consider the string structure of a DNA sequence, the organism functionality, the perfect recovery, and the high embedding capacity. We convert the string sequence of four characters in noncoding region to the decimal coded values and embed the watermark bit into coded values by two ways; DE based multiple bits embedding (DE-MBE) using pairs of neighbor coded values and consecutive DE-MBE (C-DE-MBE). Two ways process the comparison searching to prevent the false start codon that produces false coding region. Experimental results verified that our ways have more high embedding capacity than conventional methods and produce no false start codon and recover perfectly the host sequence without the reference sequence. Especially C-DE-MBE can embed more high two times than DE-MBE.

Keywords : Reversible DNA watermarking, Difference expansion, DNA steganography, DNA storage, Bioinformatics

* 정회원, 동명대학교 정보보호학과

(Dept. of Information Security, Tongmyong University)

** 정회원, 부경대학교, IT융합응용공학과

(Dept. of IT Convergence and Application Eng., Pukyong National University)

© Corresponding Author(E-mail: krkwon@pknu.ac.kr)

※ 본 연구는 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원(NRF-2011-0023118, NRF-2014R1A1A4A01006663)과 2013년도 부산광역시 Brain Busan(BB21) 사업 지원을 받아 수행된 것임.

Received ; March 19, 2015 Revised ; June 5, 2015 Accepted ; June 27, 2015

I. 서 론

DNA 컴퓨팅 기술과 더불어 비유전적인 정보를 DNA 분자 매개물 내에 은닉하는 하는 기술로, 비밀 메시지 전송을 위한 DNA 스테가노그래피^[1~2], 장기간 대용량(Long-term Large-scale)의 유전체 메모리인 DNA 저장(DNA storage)^[3,4] 등이 제시되고 있다. 특히 변이에 강인한 DNA 워터마킹^[5-10]은 Heider 등이 제시한 DNA-Crypt 알고리즘 기반 이후 지금까지 연구되어지고 있다. DNA 워터마크는 GMO(genetically modified organisms)의 비인가 사용을 식별하거나 야생형 게놈(wild type genome)과 인공 게놈(artificially designed genome)을 구분하거나, 은닉된 태그의 PCR 증폭과 시퀀싱에 의한 추적과 에이전트, 고립(isolate), 변형(strain)을 식별하는 방법을 제공한다.

대부분의 DNA 스테가노그래피, DNA 저장 및 DNA 워터마킹 기법들은 워터마크된 또는 정보은닉된 DNA 서열로부터 원본 DNA 서열이 복원되지 않는 비가역성에 해당된다. 특히 멀티미디어 데이터와는 달리 DNA 서열들은 워터마크에 의하여 변이된 DNA 서열로부터 원본 DNA 서열이 복원이 되어야 하며, 생물학적 기능의 변경이 없어야 한다. 이를 해결하기 위하여 일부 연구자들에 의하여 가역 DNA 워터마킹에 대한 연구가 이루어져 왔다. 기존의 가역 DNA 워터마킹 방법들을 살펴보면, Chen 등^[11]은 비부호 DNA 서열의 네 문자열을 십진수로 변환한 후, 무손실 압축 및 DE(Difference Expansion) 기반 방법을 적용하였다. Huang 등^[12]은 낮은 염기 변화율을 위하여 최고 및 최저 히스토그램 값들의 쉬프팅을 이용한 방법을 적용하였으나, 매우 낮은 용량을 가진다. Shiu 등^[13]은 상보쌍 염기 치환 기반 데이터 은닉으로 워터마크 추출 및 복원시 참조되는 원본 서열이 필요하다. 기존 방법들은 원본 서열 길이를 유지하나, 허위 개시코돈 방지를 고려하지 않고, 비블라인드이거나, 매우 낮은 용량을 가진다. 부호 DNA 서열의 개시를 나타내는 'ATG'의 메티오닌(Methionine) 개시코돈(start codon)에 의하여 리보솜에서 단백질 번역이 시작된다. 비부호 DNA 서열이 워터마크에 의하여 개시코돈으로 변경될 경우, 단백질로 잘못 번역된다. 따라서 비부호 DNA 워터마킹에서는 허위 개시코돈(false start codon)이 발생되지 않도록 하여야 한다.

본 논문에서는 원본 서열 길이를 유지하면서 허위 개

시코돈 방지, 블라인드 추출 및 복원, 및 높은 은닉 용량을 가지는 DE 기반 비부호 DNA 서열의 가역 DNA 워터마킹 방법을 제안한다. DE 이론은 Tian 등^[14]의 가역 영상 워터마킹에서 처음 제안된 후, 은닉 용량 제어 및 향상을 위하여 PE(Prediction error expansion) 기반 워터마킹 방법^[15]이 제안되어져 왔다. 영상의 인접 화소 간의 높은 상관성과는 달리, DNA 서열 상에 인접한 염기서열 간의 상관성이 높지 않다. 또한 영상 화질과는 달리, 비부호 DNA 서열은 허위 개시코돈, 단백질 변경과 같은 생물학적 기능 변경이 조건에서 자유로이 변경이 가능하다. 따라서 제안한 방법에서는 DE 이론 기반으로 최대 확장 조건을 만족하는 다중비트를 은닉한다.

먼저 비부호 DNA 서열의 (A,T,G,C) 4-변수(quaternary) 문자열은 비트의 십진수 수치계수로 변환된 다음, DE 이론 기반으로 DNA 수치계수에 적용할 수 있는 DE-MBE(DE based Multiple Bit Embedding)과 CDE-MBE(Consecutive DE-MBE) 방법으로 다중비트를 은닉한다. DE-MBE 방법은 DE 이론에 따라 인접 수치계수 쌍의 차분 확장에 은닉할 수 있는 최대한의 다중 비트를 은닉한다. CDE-MBE 방법은 이전 은닉된 수치계수와 현재 수치계수 쌍의 차분 확장에 다중 비트를 은닉하는 것으로, DE-MBE보다 용량성을 확장하기 위한 것이다. 은닉 용량에 대한 비교 실험에서 제안한 DE-MBE 방법과 CDE-MBE 방법은 평균 0.119bpn과 0.235bpn이며, Chen 방법^[11]과 Huang 방법^[12]은 평균 0.109bpn과 0.026bpn으로 나타났다. 따라서 제안한 CDE-MBE 방법이 다른 방법에 비하여 은닉 용량이 제일 높음을 확인하였다.

본 논문의 구성은 다음과 같다. 먼저 II장에서는 가역 DNA 정보은닉 이론과 기존 가역 DNA 정보은닉 방법에 대하여 살펴본다. III장에서는 제안한 염기서열 수치화, 허위 개시코돈 방지에 대하여 살펴본 후, DE-MBE방법과 CDE-MBE 방법에 대하여 자세히 살펴본다. IV장에서는 제안한 방법과 기존 방법과의 비교 실험 분석 한 후, 마지막 V장에서 본 논문의 결론을 맺는다.

II. 가역 DNA 워터마킹 이론

1. 가역 비부호 DNA 워터마킹

비부호 DNA 서열에 대한 기본적인 가역 워터마킹

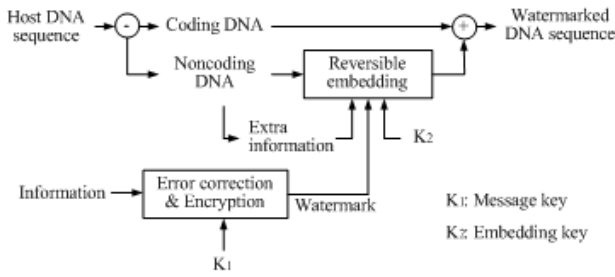


그림 1. 비부호 DNA 서열 가역 워터마킹 블록도
Fig. 1. Block diagram of reversible watermarking for noncoding DNA sequence.

기본적인 과정은 그림 1에서와 같다. 먼저 DNA 서열 상에서 분리된 비부호 DNA의 염기 서열들은 십진수 형태의 수치열로 변환된 후 오류 정정 또는 암호화된 워터마크가 가역성을 가지도록 수치열에 은닉된다. 그런 다음 워터마크된 수치열은 염기 서열로 변환된 후, 부호 DNA 서열에 연결하여 워터마크된 DNA 서열이 생성된다.

비부호 DNA는 영상 데이터의 화질과 같은 속성은 없으나, 고려해야 될 주요 특징이 있다.

1) 문자열의 수치 변환 : 염기 서열은 네 가지 문자 (A,T,C,G(or U))로 구성되므로, 영상 데이터에 비하여 매우 낮은 동적 범위를 가진다. 즉, 하나의 염기는 2비트의 정보를 가지며, 8비트의 영상 화소값에 비하여 매우 낮은 레벨을 가진다. 예를 들어, 4개의 연속된 염기 서열은 $44=256$ 레벨을 가지게 되므로, 8비트 영상 화소값과 같은 레벨을 가진다.

2) 허위 개시코돈 : 은닉 과정에 의하여 비부호 DNA의 일부 염기 서열이 부호 DNA 서열의 개시코돈 (Methionine, 'ATG')으로 변경될 가능성이 있다. 따라서 워터마크 은닉 과정에서 개시코돈으로 변경될 염기 서열들을 예측하여 제외하여야 한다.

3) 서열 길이 및 블라인드 : DNA 서열의 길이가 변경되지 않으면서, 참조 서열 또는 원본 서열이 없이 워터마크 추출 또는 DNA 서열 복원이 가능하여야 한다. 비부호 DNA 서열의 수치열이 주어졌을 때, 기존 영상 데이터에 사용되는 예측 차분 확장 (Prediction Error Expansion) 방법을 이용하여 수치별 1비트씩 은닉이 가능하다. 예를 들어, 임의의 수치 x 에 대한 예측 계수 \hat{x} 이고, 워터마크 비트 w 일 때, 은닉된 수치 x' 와 워터마크 추출 및 복원은

$$x' = \hat{x} + 2(x - \hat{x}) + w = 2x - \hat{x} + w, \quad (1)$$

$$w = x' - \hat{x} - 2 \lfloor \frac{x' - \hat{x}}{2} \rfloor, \quad x = \frac{1}{2}(x' + \hat{x} - w)$$

와 같다. 위의 방법은 각 수치에 1비트만 은닉이 가능하다. 그러나 영상 데이터와는 달리 DNA 서열의 수치열은 인접 수치들 간의 상관관계가 크지 않으므로, 오차가 적은 예측이 어렵다. 따라서 DNA 서열의 수치열에 적합한 대용량 가역 워터마킹 방법이 필요하다.

2. 기존 방법

Chen 등^[11]은 기존 가역 영상 워터마킹 방법에서 많이 사용되는 무손실 압축과 DE 기반 방법들을 각각 제안하였다. 이들은 먼저 "ATCG"의 염기 심볼들을 2비트 이진으로 변환한 다음, 염기 이진 서열을 $|w|$ 비트 단위의 워드로 십진수로 변환한다. 그리고 두 가지 방법에 의하여 비밀 메시지들을 십진수 서열에 은닉한다. 무손실 압축 기반 정보 은닉 방법에서는 십진수 서열을 산술부호화에 의하여 압축한 다음, 이진 비밀 메시지를 압축열의 마지막 단에 추가(append)한다. 이 방법은 $|w|$ 가 2비트일 때 bpn이 제일 높으며, 실험적으로 평균 0.75-0.81bpn를 가진다. 그러나 압축열에 비밀 메시지가 추가되므로, 압축 코딩 프로파일의 변경되어야 하며, 또한 압축열의 길이가 증가된다. DE 기반 정보 은닉 방법에서는 십진수 서열의 워드 쌍을 확장 집합 S1, 변경 집합 S2, 비변경 집합 S3으로 분류한 후 각 워드 쌍의 위치맵을 생성한다. 압축된 위치맵(map), S2에 속한 워드 쌍의 원본 LSB들(LSB(S2)), 및 비밀 메시지들을 압축열의 마지막 단에 추가한다. 이 방법은 Tian 등^[14]의 가역 영상 워터마킹 방법을 DNA 서열에 적용한 것으로, 은닉 데이터 용량이 매우 낮다.

Huang 등^[12]은 낮은 염기 변화율을 가지는 히스토그램 기반 가역 DNA 워터마킹 방법을 제안하였다. 이 방법에서는 Chen의 방법과 같이 이진 염기서열을 2비트 단위로 십진수로 변환한 다음, 십진수 서열의 히스토그램을 구한다. 이 때 h 를 가장 높은 빈도수의 값, $L1$ 를 가장 낮은 빈도수의 값, $L2$ 를 두 번째 낮은 빈도수의 값이라 한다. 임의의 십진수 p_j 가 $L1$ 이면, p_j 를 $L2$ 로 변경하고, 위치맵을 1로 놓는다. p_j 가 $L2$ 이면, p_j 를 변경하지 않고, 위치맵을 0으로 놓는다. p_j 가 h 일 때, 은

닉 비트가 0이면, p_j 는 변경하지 않고, 은닉 비트가 1이면, p_j 는 L1으로 변경한다. 복원과 추출은 위치맵, h, L1, L2 값에 의하여 수행된다. 실험 상에서 $t=2$ 일 때, 0.024 bpn과 4.07-4.8%의 염기 변경율을 가지며, $t=3$ 일 때 0.011 bpn과 1.86-2.34%의 염기 변경율을 가진다. 이 방법은 염기 변경율은 낮으나, bpn이 매우 낮고, Chen의 방법과 같이 허위 개시코돈의 발생된다.

이 외에도 Shiu 등^[13]은 상보 쌍에 의한 치환 방법에 의하여 데이터를 은닉하는 것으로, 추출 및 복원을 위하여 참조(또는 원본) DNA 서열이 필요하다.

III. 제안한 가역 DNA 워터마킹

본 논문에서는 DE 이론 기반의 다중 비트 은닉이 가능한 가역 DNA 워터마킹 기법을 제안한다. 제안한 방법은 비부호 DNA 염기 서열에 가역 워터마크를 은닉하기 위하여 다음과 같은 특징을 가진다. 1) 4-문자 염기 서열의 신호처리 용이성 : 4-문자 염기 서열의 n 차 수치계수로 부호화하여 염기 서열에 대한 워터마크 은닉, 추출, 및 복원 과정이 용이하도록 한다. 2) 다중비트 은닉 : 높은 은닉 용량을 위하여 부호계수 쌍별 최대 확장 조건을 만족하는 비트수를 은닉한다. 3) 허위 개시코돈 방지 : 워터마크된 염기서열 상에 인접 염기들 간의 비교 탐색을 통하여 부호 영역의 개시코돈 생성을 방지한다. 4) 원본 DNA 서열의 길이와 생물학적 기능의 변경이 없으며, 블라인드 추출 및 복원이 가능하다.

본 장에서는 먼저 염기 서열의 수치 부호화에 대하여 살펴본 후, 이들 계수에 대한 DE 기반 다중 비트 은닉, 복원, 및 추출 과정에 대하여 각각 살펴보기로 한다.

1. 은닉 대상 비부호 DNA 서열 선택

그림 2와 같이 DNA 서열 D 내에 인트론 성분으로 구성된 비부호 DNA 영역 D^{nc} 과 엑손 성분으로 구성된 부호 DNA 영역 D^c 들이 존재한다. 제안한 방법에서는 n 차 염기 수치일 때, 영역 내의 염기 개수가 $10n$ 보

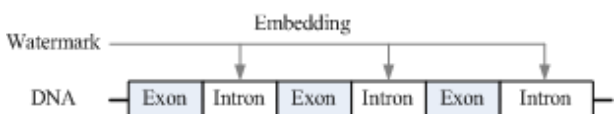


그림 2. DNA 서열 구조
Fig. 2. DNA sequence structure.

다 클 경우 이 영역을 은닉 대상 영역으로 선택한다. 즉, 선택된 모든 비부호 DNA 영역 $\Gamma^{nc}(n)$ 은

$$\Gamma^{nc}(n) = \{D_k = \{b_{ki} | i \in [1, |D_k|], |D_k| > 10n\}\} \quad (2)$$

와 같다. D_k 는 선택된 k 번째 비부호 DNA 영역이며, b_{ki} 는 D_k 영역의 i 번째 4-문자 염기이고, $|D_k|$ 는 D_k 의 염기 개수를 나타낸다. 차수 n 에 대한 은닉 대상 영역 비율 $R_{nc}(n)$ 은

$$R_{nc}(n) = \frac{|\Gamma^{nc}(n)|}{|D^{nc}|} \quad (3)$$

와 같이 전체 비부호 DNA 영역의 개수 $|D^{nc}|$ 와 은닉 대상 영역의 개수 $|\Gamma^{nc}(n)|$ 의 비율을 나타낸다. 각 영역마다 염기 개수가 다르므로, 차수 n 에 대한 은닉 대상 염기개수 비율 $R_{base}(n)$ 은

$$R_{base}(n) = \frac{\text{Number of bases} \in \Gamma^{nc}}{\text{Number of bases} \in D^{nc}} \quad (4)$$

와 같이 모든 영역 D^{nc} 내의 염기개수와 은닉대상 영역 $\Gamma^{nc}(n)$ 내의 염기개수의 비율을 나타낸다. 수치차수 n 은 다음 절에 자세히 설명하기로 한다.

다음 절부터 임의 비부호 DNA 영역 D_k^{nc} 상에 가역 워터마크를 은닉하는 방법에 대하여 살펴보기로 한다.

2. 4-문자 염기 수치화

일반적으로 4-문자 뉴클레오티드 염기는 $b = \{ 'A', 'T', 'C', 'G' \}$ 와 같이 4개 문자 중의 하나로 표현되며, 이를 4개의 십진수 또는 2비트의 이진수로 표현된다; $b = \{0,1,2,3\} \cdot 10 = \{00,01,10,11\} \cdot 2 \leftrightarrow b = \{ 'A', 'T', 'C', 'G' \}$.

워터마킹 신호 처리 용이성을 위하여 2비트 보다 큰 비트의 수치로 확장되어야 한다. 따라서 제안한 방법에서는 n 개의 연속적인 염기 부분서열 x 로 이루어진 $2n$ 비트의 수치 x 로

$$x = f(x) = b_1 \cdot 2^{2(n-1)} + \dots + b_{n-1} \cdot 2^2 + b_n \quad (5)$$

$$\text{where } x = (b_1, b_2, \dots, b_n), x \in [0, 2^{2n} - 1]$$

와 같이 부호한다. 수치 x 로부터 염기 부분서열의 각 염기들은

$$f^{-1}(x) = x \text{ where } b_j = (x \gg 2(n-j)) \% 3 \quad (6)$$

와 같이 쉽게 복원된다.

본 논문에서는 염기 부분서열의 염기 개수 n 를 수치 차수(numerical degree)로 부르기로 하며, n 차 염기 수치를 줄여 수치로 부르기로 한다.

가. 수치차수에 대한 허위 개시코돈

워터마크 은닉 과정에서 일부 염기 서열은 부호 DNA 서열의 개시코돈 ('ATG')으로 변경될 수 있다. 즉, 허위 개시코돈에 의하여 비부호 DNA 서열의 일부가 부호 DNA 서열로 변경된다. 허위 개시코돈은 수치 내 또는 수치 간에 발생할 수 있다.

1) 수치 내 (Intra codes) 경우 : 2보다 큰 수치차수 n 이 주어졌을 때, $n-2(n > 2)$ 개의 허위 개시코돈이 될 수치들이 $X^c = \{x_i^c | i \in [1, n-2], n > 2\}$ 가 발생된다. 이 때 수치차수 n 이 3보다 작을 때, 수치 내 허위 개시코돈은 발생되지 않는다. 워터마크 은닉 과정에서는 허위 개시코돈에 해당되는 부호 계수 X^c 가 발생되지 않도록 미리 예측하여야 한다.

2) 수치 간 (Inter codes) 경우 : 이전 워터마크된 부호 계수 x'_{t-1} 의 염기 부분서열 x'_{t-1} 와 현재 처리된 계수 x'_t 의 염기 부분서열 x'_t 간에 허위 개시코돈이 발생할 수 있다. 그림 3에서와 같이 이전 워터마크된 염기 부분서열 x'_{t-1} 과 현재 워터마크된 염기 부분서열 x'_t 이 ($x'_{t-1}x'_t$) 일 때 (...A, TG...) 또는 (...AT, G...) 때 중간 부분에 허위 개시코돈이 발생된다. 워터마크 은닉 과정에서는 x'_{t-1} 의 마지막 또는 마지막 두 개의 염기와 x'_t 의 처음 또는 처음 두 개의 염기가 'ATG'에 해당되는지 항상 비교 검색하여야 한다.

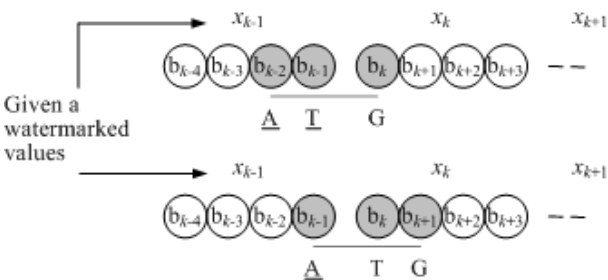


그림 3. 수치 간에 허위 개시코돈 발생 가능성 ($n=4$)
Fig. 3. Possibility of false start codon between base numerical values.

나. n 차 염기 수치 예측의 제한 조건

대용량 데이터 은닉을 위하여 오차가 적은 염기 수치 예측기가 필요하다. 그러나 가역 DNA 워터마킹에 사용되는 염기 수치 예측기는 “추출 및 복원 과정에서 예측기에 의한 부가 정보가 필요없고, 모든 DNA 서열 상에 (일관성 있는 예측 모델) 적용할 수 있어야 하며, 현재 수치와의 오차가 최소화되도록 예측”하여야 한다. 대부분의 인접한 염기 서열 (또는 수치 계수)들 간에는 상관도가 적다. 그림 4는 수치차수 n 이 4일 때, 인접 수치들의 차분에 대한 분포를 보여준다. 이 분포를 살펴보면, 수치 차분은 균일 분포에 근접한 분산이 매우 큰 가우시안 분포를 가진다. 즉, 균일 분포에 가까운 수치 차분 분포 상에서 최소 오차를 가지는 수치 예측은 매우 어렵다. 따라서 DNA의 염기 수치 예측기 설계는 매우 힘들다. 따라서 본 논문에서는 예측기가 없고, 다중 비트 은닉이 가능한 차분 확장 기반 가역 워터마킹 기법을 제안한다.

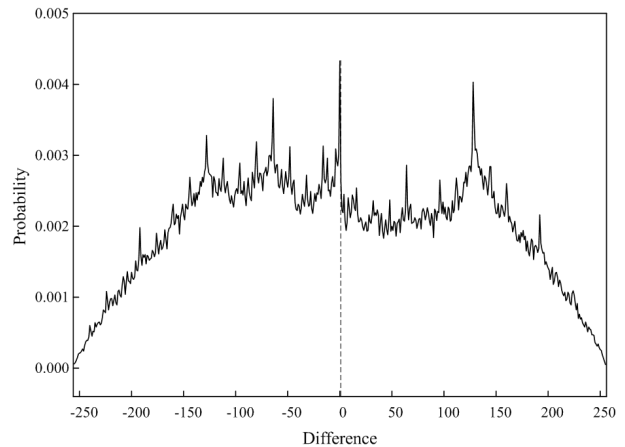


그림 4. 비부호 DNA 서열 계수 상에 수치차수가 $n=4$ (8bit 계수)일 때 인접 수치 차분 히스토그램
Fig. 4. Histogram of numerical value difference with numerical order $n=4$ in noncoding DNA numerical sequence.

3. DE 기반 다중 비트 은닉 방법

N_D 개의 염기들로 구성된 비부호 DNA 영역 D_k^{nc} 과 수치차수 n 이 주어졌을 때, D_k^{nc} 은 $N = \lfloor N_D/n \rfloor$ 개의 수치로 구성된 수치서열 X 로 부호된다. 본 장에서는 먼저 다중 비트 은닉을 위한 확장 조건에 대하여 살펴본 후, 수치서열 X 내에 두 가지의 DE 기반 다중 비

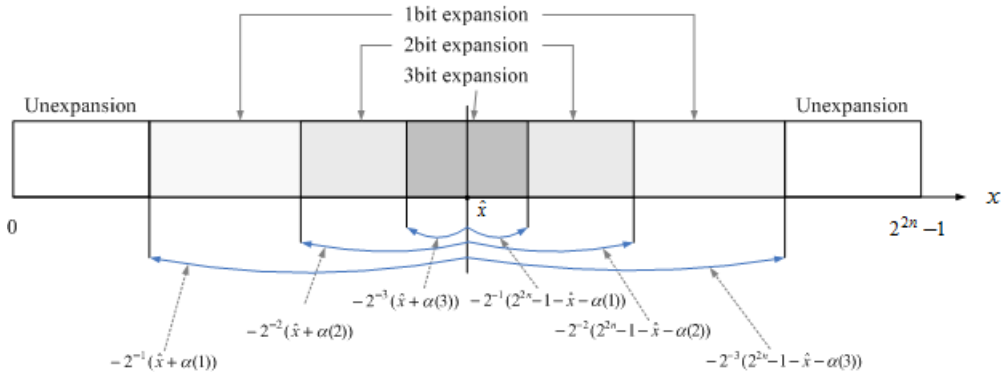


그림 5. 임의의 기준 수치 \hat{x} 상에서 비트별 확장 가능한 x 의 범위

Fig. 5. Expandible range of numerical value x for bits on any reference numerical value \hat{x} .

트 은닉 방법에 대하여 살펴보기로 한다.

가. 다중 비트 은닉위한 확장 조건

인접 화소 간의 상관관계가 높은 영상 데이터에서는 대상 화소(또는 계수)에 1비트 이상의 데이터를 은닉할 경우, 차분 (d)를 비트 수의 배수만큼 확장하여야 하므로, 영상 화질이 열화된다. 즉, 화질 열화에 따라 1비트 이상의 데이터 은닉이 어렵다. 이와는 달리 화질에 대한 조건이 없는 비부호 DNA 서열에는 허위 개시코돈 제한을 제외하고, 유효 범위 내에 이동이 자유롭다. 따라서 대상 수치 쌍에 대한 차분(d)는 확장 조건에 따라 k 배 확장이 가능하므로 k 비트가 은닉될 수 있다.

워터마크의 k 비트 $\{w_j\}_1^k$ 가 주어졌을 때, 대상 수치에 대한 비트 은닉은

$$x' = \hat{x} + d_w = \hat{x} + 2^k d + \text{sgn}(d) \sum_{j=0}^{k-1} 2^j w_{j+1} \quad (7)$$

where $d = x - \hat{x}$

와 같다. 은닉된 수치 x' 와 비트 수 k 가 주어졌을 때, 워터마크 추출 및 수치 복원은

$$w_j = ((x' - \hat{x}) \gg (j-1)) \% 2 \quad (8)$$

$$x = \hat{x} + d = \hat{x} + (x' - \hat{x}) \gg k \quad (9)$$

이며, 수치차수 n 일 때, k 배 확장 위한 차분 d 조건은

$$0 \leq \hat{x} + 2^k d + \text{sgn}(d) \sum_{j=0}^{k-1} 2^j w_{j+1} \leq 2^{2n} - 1 \quad (10)$$

와 같다. 위 식을 살펴보면, 차분 d 의 확장 조건은 워터마크의 k 비트 $\{w_j\}_1^k$ 와 기준 계수 \hat{x} 에 의하여 결정된다. $\{w_j\}_1^k$, \hat{x} 가 주어졌을 때, 차분 d 가 k 배 확장하기 위하여 수치 x 는 다음과 같은 조건을 만족하여야 한다.

$$x \in [\max(0, \lceil \hat{x} + 2^{-k}(-\hat{x} - \alpha(k)) \rceil), \min(2^{2n} - 1, \lfloor \hat{x} + 2^{-k}(2^{2n} - 1 - \hat{x} - \alpha(k)) \rfloor)] \quad (11)$$

$$\text{where } \alpha(k) = \text{sgn}(d) \sum_{j=0}^{k-1} 2^j w_{j+1}$$

위 조건을 이용하여, 수치 x 에 은닉 가능한 워터마크 비트수를 판단할 수 있다.

그림 5는 임의의 기준 수치 \hat{x} 상에 은닉 가능한 비트수에 따른 수치 x 의 범위를 나타낸다. 그래프 상에서 1비트 이상 확장 가능 영역은 비확장 가능 영역과 동일한 면적을 가지며, 비트수가 증가할수록 확장 가능 영역은 기하급수적으로 좁아진다. 워터마크 비트 $\{w_j\}_1^k$ 에 따른 k 배 확장 가능한 x 의 영역 $R_k(\{w_j\}_1^k)$ 는

$$R_k(\{w_j\}_1^k) = [\max(0, \lceil \hat{x} + 2^{-k}(-\hat{x} - \alpha(k)) \rceil), \min(2^{2n} - 1, \lfloor \hat{x} + 2^{-k}(2^{2n} - 1 - \hat{x} - \alpha(k)) \rfloor)] \quad (12)$$

$$\text{where } R_1 \supset R_2 \cdots \supset R_{n-1}$$

와 같다.

나. DE-MBE 방법

(1) 부호 과정

인접 수치 쌍 $\{(x_i, x_{i+1})\}$ 들 중, 임의 수치 쌍

(x_i, x_{i+1}) 에 대한 차분 d_i 는 $d_i = x_i - x_{i+1}$, ($\hat{x}_i = x_{i+1}$)이다. 이 때 확장 조건에 따라 d_i 의 은닉 비트수 k_i 가 결정된 후, x_i 에 순차적으로 워터마크의 k_i 비트 $\{w_j\}_{j=1}^{k_i}$ 가 은닉된다.

$$x'_i = x_{i+1} + 2^{k_i}d_i + \alpha(k_i) \quad (13)$$

$$\text{where } \alpha(k_i) = \text{sgn}(d_i) \sum_{j=0}^{k_i-1} 2^j w_{j+1}$$

은닉된 계수 x'_i 의 역기서열 $x'_i = f^{-1}(x'_i)$ 과 이전 계수 x_{i-1} 의 역기서열 $x_{i-1} = f^{-1}(x_{i-1})$ 이 주어졌을 때, 수치 계수 내 및 수치 계수 간 개시코돈 “ATG”이 발생할 경우, x_i 에 대한 은닉 비트수 k_i 를 하나 감소한 후 위 식에 의하여 k_i 비트를 은닉한다. k_i 가 0이 될 때까지 반복 수행한다. 위의 방법에 의하여 가역 워터마크 된 영역 D_k^{nc} 를 얻는다.

수치 쌍에 은닉 비트수 정보 K 는 워터마크된 비부호 DNA 영역 D_k^{nc} 에 포함되어 전송되어야 한다. 이 때 K 추출위한 다른 외부 정보없이 허위 개시코돈이 발생되지 않아야 한다. 제안한 방법에서는 D_k^{nc} 의 4-문자 염기 이진수의 LSB 비트 B 와 이진 비트수 정보 K 의 부가 데이터를 산술 부호화(arithmetic coding)에 의하여 무손실 압축된 비트열 C 을 생성한다.

수치 쌍 최대 은닉 비트수가 수치 차수 n 으로 설정될 때, K 는 $\lfloor N/2 \rfloor \lceil \log_2 n \rceil$ 비트로 표현되며, 부가 데이터 비트는 $\lfloor N/2 \rfloor \lceil \log_2 n \rceil + N_D$ 이고, B 는 N_D 비트이다. 압축 비트 c_j 는 D_k^{nc} 내에 4-문자 염기의 이진수 b'_i 의 LSB에 차례로 은닉된다.

$$b''_i = b'_i \oplus c_j = (b_i \gg 1) \ll 1 + c_j \quad (14)$$

$$\text{if } b''_{i-2} \neq 'A' \text{ and } b''_{i-1} \neq 'T'$$

이 때, 이전 은닉된 두 개의 염기 (b'_{i-2}, b'_{i-1})가 “AT”일 경우, 은닉 과정을 생략한다. 최종적으로 압축 비트열 C 을 포함하는 영역 D_k^{nc} 내에 “AT” 염기열은 다음 염기에 압축 비트가 포함되지 않음을 직접적으로 나타내는 마커로 수행된다. 압축열 C 의 길이는 압축 알고리즘에 의하여 결정되나, 본 논문에서는 일반적인 무손실 압축 알고리즘인 산술 부호화를 사용하기로 한다.

D_k^{nc} 내에 치환 가능한 비트량 Ψ_k 는 치환 과정에서 생략되는 염기의 개수에 의하여 결정된다. 압축열 C 의 길이는 항상 $C < \Psi_k \leq N_D$ 이 되어야 한다. 압축열 C 길이가 $\rho(\lfloor N/2 \rfloor \lceil \log_2 n \rceil + N_D)$ ($\rho < 1$) 일 때, 압축율 ρ 는

$$\rho < \frac{N_D}{\lfloor N/2 \rfloor \lceil \log_2 n \rceil + N_D}, N = \lfloor \frac{N_D}{n} \rfloor \quad (15)$$

이어야 한다.

(2) 복호 과정

복호 과정에서는 전송된 비부호 DNA 영역 D_k^{nc} 가 주어졌을 때, “AT” 다음에 오는 염기를 제외한 모든 염기들의 LSB 비트들로부터 압축 비트열 C 가 얻는다. 그리고 C 로부터 수치 쌍 은닉 비트수 정보 K 와 원본 염기 이진수의 LSB 비트 B 를 쉽게 얻어진다. D_k^{nc} 내의 염기 이진수의 LSB에 B 를 치환하여 가역 워터마크 된 영역 D_k^{nc} 가 얻어진다. D_k^{nc} 내에 n 차 염기 수치열로부터 수치 쌍 은닉 비트수 정보 K 를 이용하여 워터마크가 추출되고 원본 영역 D_k^{nc} 이 복원된다.

예를 들어 수치 쌍 (x'_i, x_{i+1}) 의 은닉 비트수가 $k_i > 0$ 일 때, 은닉된 수치 x'_i 상에 워터마크의 k_i 비트 $\{w_j\}_{j=1}^{k_i}$ 가

$$w_j = ((x'_i - x_{i+1}) \gg (j-1)) \% 2 \quad (16)$$

$$\text{for } j = 0, \dots, k_i - 1$$

와 같이 추출된다. 그리고 원본 수치 x_i 은 차분의 k_i 비트 쉬프트에 의하여

$$x_i = x_{i+1} + ((x'_i - x_{i+1}) \gg k_i) \quad (17)$$

와 같이 복원된다.

(3) 용량 분석

임의의 비부호 DNA 서열 D_k^{nc} 에 은닉되는 워터마크 비트 N_W^{DE} 는 수치 쌍 은닉 비트수 $K = \{k_i\}_1^{\lfloor N/2 \rfloor}$ 의 합에 해당되며, 뉴클레오티드 염기 당 비트수 bpn(bit per base) bpn_W^{DE} 은

$$bpn_W^{DE} = \frac{N_W^{DE}}{N_D} = \frac{1}{N_D} \sum_{i=1}^{\lfloor N/2 \rfloor} k_i \quad [\text{bit/base}] \quad (18)$$

와 같다. N_D 는 D_k^{nc} 의 염기 개수를 나타낸다.

다. C-DE-MBE 방법

(1) 부호 과정

DE-MBE 방법에서 수치 쌍은 (은닉될 수치, 참조 수치)로 구성된다. 비부호 DNA 서열의 수치 쌍 차분은 그림 4와 같이 균일 분포에 가까우므로, 은닉된 수치와의 차분은 균일 분포에 가깝다. C-DE-MBE 방법에서는 용량 향상을 위하여 참조 수치로 이전 은닉된 수치를 사용한다. 현재 수치와 이전 은닉된 수치 쌍이 (x_i, x'_{i-1}) 일 때 두 수치의 차분은

$$d_i = x_i - x'_{i-1}, \hat{x}_i = x'_{i-1}, x'_0 = 0 \quad (19)$$

와 같다. 따라서 d_i 의 확장 조건에 따라 은닉될 비트수 k_i 가 결정된 다음, x_i 에 k_i 비트 $\{w_j\}_{j=1}^{k_i}$ 가

$$x'_i = x'_{i-1} + 2^{k_i} d_i + \alpha(k_i) \quad (20)$$

$$\text{where } \alpha(k_i) = \text{sgn}(d_i) \sum_{j=0}^{k_i-1} 2^j w_{j+1}$$

와 같이 은닉된다. x_i 와 x'_{i-1} 의 염기 부분서열 x'_i 와 x'_{i-1} 이 개시코돈을 포함할 경우, 비트수 k_i 를 하나 감소한 다음 $k_i=0$ 일 때까지 위의 과정을 반복한다. 위의 방법에 의하여 가역 워터마크된 영역 D_k^{nc} 가 얻어진다.

C-DE-MBE 방법은 DE-MBE 방법과는 달리 허위 개시코돈을 포함할 수치를 제외한 모든 수치에 비트가 은닉된다. 따라서 워터마크 추출 및 수치 복원에 필요한 부가 데이터로는 수치 은닉 비트수 $K = \{k_i\}_1^N$ 와 D_k^{nc} 의 4-문자 염기 이진수의 LSB 비트 B 로 총 $N \lceil \log_2 n \rceil + N_D$ 비트의 길이를 가진다. 부가 데이터의 무손실 압축 비트열 C 은 DE-MBE 방법과 동일하게 D_k^{nc} 내에 4-문자 염기의 이진수 b'_i 의 LSB에 차례로 은닉되어, 최종 압축 비트열 C 가 포함된 비부호 DNA 영역 D_k^{nc} 가 얻어진다. 이 때, 이전 은닉된 두 개

의 염기 (b'_{i-2}, b'_{i-1})가 “AT”일 경우, 은닉 과정을 생략한다. C-DE-MBE 방법에서 압축 비트열 C 가 D_k^{nc} 내에 은닉되기 위하여 무손실 압축율 ρ 는

$$\rho < \frac{N_D}{N \lceil \log_2 n \rceil + N_D} \quad (21)$$

이어야 한다.

(2) 복호 과정

C-DE-MBE 방법의 복호 과정은 DE-MBE 복호 과정과 동일하게, D_k^{nc} 로부터 압축 비트열 C 의 수치 은닉 비트수 K 와 원본 염기 수치 LSB B 를 얻은 다음, B 에 의하여 D_k^{nc} 를 얻는다. 그리고 D_k^{nc} 로부터 K 에 의하여 워터마크 추출과 원본 서열 D_k^{nc} 을 복원한다. 예를 들어, 수치 쌍 (x'_i, x'_{i-1}) 과 비트수 $k_i > 0$ 가 주어졌을 때, x'_i 에 은닉된 k_i 비트 $\{w_j\}_{j=1}^{k_i}$ 추출과 원본 수치 x_i 의 복원은 다음에 의하여 구하여진다.

$$w_j = ((x'_i - x'_{i-1}) \gg (j-1)) \% 2 \quad (22)$$

for $j = 1, \dots, k_i$

$$x_i = x'_{i-1} + ((x'_i - x'_{i-1}) \gg k_i) \quad (23)$$

(3) 용량 분석

임의의 비부호 DNA 서열 D_k^{nc} 상에 은닉되는 데이터 비트 N_W^{CDE} 은 모든 수치 상에 은닉된 비트수 $K = \{k_i\}_1^N$ 의 합이며, bpn (bit per base) bpn_W^{CDE} 은

$$bpn_W^{CDE} = \frac{N_W^{CDE}}{N_D} = \frac{1}{N_D} \sum_{i=1}^N k_i \quad [\text{bit/base}] \quad (24)$$

와 같다. C-DE-MBE 방법은 허위 개시코돈을 포함할 수치를 제외한 모든 수치에 대하여 데이터가 은닉되므로, DE-MBE 방법에 비하여 은닉 데이터 용량이 2배 정도 높다.

IV. 실험 결과

본 실험에서는 제안한 방법과 Chen 방법^[11], Huang

방법^[12]의 용량 및 변이에 대한 가역성에 대하여 비교 분석하였다.

1. 실험 환경

실험에서 사용된 서열은 NCBI GenBank에서 제공된 것이며, 타입, 접근번호, 염기 개수, 비부호 DNA 영역 개수 정보는 표 1에 나타내었다. 비부호 DNA 영역들은 다양한 염기 개수를 가지며, 매우 작은 염기 수로 구성된 영역들은 데이터 은닉 대상에서 제외된다.

그림 6은 수치차수 n 이 2에서 10일 때 은닉 대상 영

표 1. NCBI에서 제공되는 테스트 DNA 서열
Table 1. Test DNA sequences in NCBI.

Type	Access No.	Total Bases	비부호 영역 개수	염기 개수
Archaea	AE017199	490,885	289	38,932
Bacterium	CP000108	2,572,079	1,770	301,761
Bacterium	CP000247	4,938,920	3,850	570,214
Bacterium	CP000672.1	1,887,192	1,444	466,266
Bacterium	AF012886.2	6,756	7	2,058
Bacterium	AE014075.1	5,231,428	3,767	631,026
Bacterium	CP000473.1	9,965,640	6,224	962,527
Eukaryota	nm_000520	2,437	2	847
Eukaryota	NC_001709.1	19,517	11	8,347
Eukaryota	NC_006033	1,195,132	533	393,739
Eukaryota	AL161582.2	198,669	302	137,622
Eukaryota	AL161595.2	198,151	215	126,917
Eukaryote	NC_006047	2,007,515	1,099	516,557
Moss	AP005672.1	122,890	99	51,916
Plant	NC_025652.1	141,255	68	91,971
Virus	AY653733.1	1,181,404	883	155,805

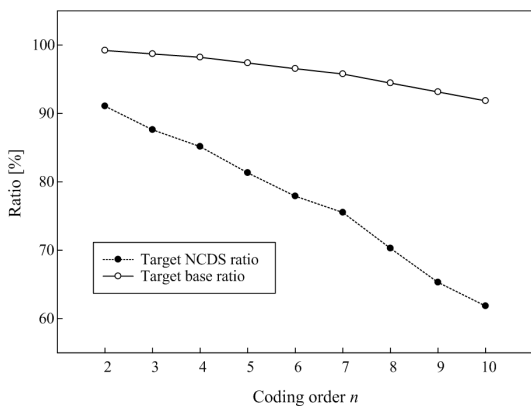


그림 6. 수치차수 n 에 대한 은닉 대상 비부호 DNA 영역 개수 비율 R_{NCD} 과 염기 개수 비율 R_{base}

Fig. 6. Ratio R_{NCD} of target noncoding DNA region numbers and ratio R_{base} of base numbers for numerical order n .

역 비율 $R_{NCD}(n)$ 와 염기개수 비율 $R_{base}(n)$ 을 보여 준다. n 이 증가할수록 $R_{NCD}(n)$ 은 비례적으로 줄어드나, $R_{base}(n)$ 은 92% 이상을 유지한다. 본 실험에서는 수치차수 n 을 2에서 10까지 가변하였으며, 워터마크 bpn bpn_W 에 최적인 수치차수를 구한 다음, 이 수치차수에 대하여 각 방법들을 비교하였다.

2. 수치차수에 대한 용량 분석

그림 7(a)는 표 1의 모든 테스트 DNA 서열에 대하여 수치차수 n 에 대한 워터마크 데이터 bpn (bpn_W)와 염기변화율 평균치를 보여준다. DE-MBE, C-DE-MBE 방법은 $n=3$ 일 때 bpn_W 과 bpn_E 이 제일 높으며, CDE의 bpn_W 이 DE의 bpn_W 보다 2배 정도 높게 나타났다.

본 실험에서는 표 1의 테스트 DNA 서열에 대한 결과를 자세히 살펴보기 위하여, 용량효율이 제일 높은

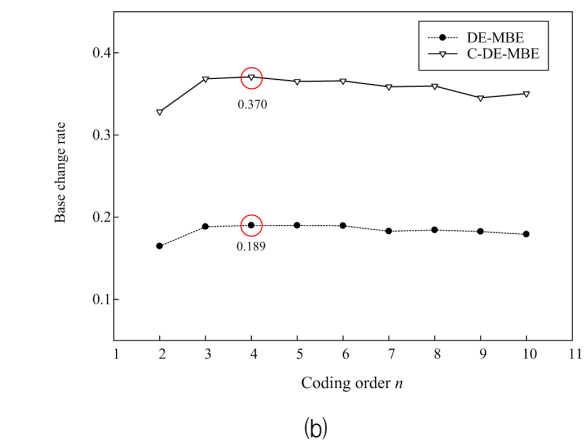
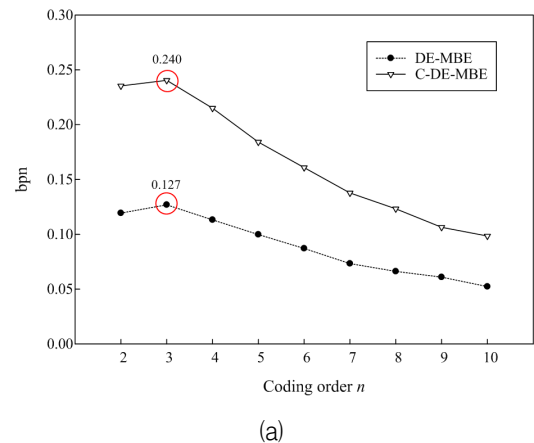


그림 7. 수치차수 n 에 대한 (a) 워터마크 데이터 bpn bpn_W , (b) 염기 변화율

Fig. 7. (a) Watermark data bpn bpn_W and (b) base change ratio for numerical order n .

표 2. 테스트 DNA 서열에 대한 은닉 워터마크 데이터 bpn과 염기 변화율
Table 2. Watermark data bpn and base change ratio of test DNA sequence.

Type	Access No.	워터마크 데이터 bpn				염기 변화율			
		제안 방법		Chen ($ w =2$)	Huang ($t=2$)	제안 방법		Chen ($ w =2$)	Huang ($t=2$)
		DE-MBE ($n=3$)	C-DE-MBE ($n=3$)			DE-MBE ($n=3$)	C-DE-MBE ($n=3$)		
Archaea	AE017199	0.118	0.237	0.098	0.034	0.185	0.351	0.154	0.060
Bacterium	CP000108	0.115	0.227	0.114	0.015	0.181	0.360	0.203	0.021
Bacterium	CP000247	0.118	0.231	0.096	0.021	0.179	0.364	0.126	0.032
Bacterium	CP000672.1	0.114	0.232	0.098	0.044	0.181	0.368	0.176	0.079
Bacterium	AF012886.2	0.118	0.226	0.111	0.022	0.186	0.342	0.175	0.035
Bacterium	AE014075.1	0.118	0.230	0.115	0.044	0.178	0.363	0.151	0.067
Bacterium	CP000473.1	0.115	0.229	0.115	0.037	0.181	0.362	0.181	0.066
Eukaryota	nm_000520	0.125	0.238	0.099	0.028	0.205	0.412	0.142	0.051
Eukaryota	NC_001709.1	0.136	0.262	0.127	0.022	0.201	0.381	0.188	0.037
Eukaryota	NC_006033	0.119	0.235	0.117	0.025	0.190	0.373	0.163	0.045
Eukaryota	AL161582.2	0.120	0.234	0.115	0.046	0.191	0.369	0.183	0.082
Eukaryota	AL161595.2	0.120	0.234	0.097	0.003	0.192	0.375	0.155	0.004
Eukaryote	NC_006047	0.116	0.233	0.115	0.016	0.189	0.370	0.210	0.023
Moss	AP005672.1	0.123	0.242	0.109	0.012	0.192	0.368	0.192	0.019
Plant	NC_025652.1	0.114	0.234	0.116	0.037	0.187	0.364	0.191	0.061
Virus	AY653733.1	0.120	0.243	0.101	0.015	0.195	0.372	0.144	0.021
평균		0.119	0.235	0.109	0.026	0.189	0.370	0.171	0.044

수치차수 n 일 때 (DE-MBE, C-DE-MBE는 $n=3$), bpn_W 을 표 2에 나타내었다. 이 때 동등한 성능 평가를 위하여 Chen 방법($|w|=2$)과 Huang 방법($t=2$)에서는 bpn_W 이 제일 높을 때의 변수를 사용하였다. 표 2의 결과를 살펴보면, 모든 테스트 DNA 서열에 대하여 C-DE-MBE 방법이 워터마크 데이터 bpn bpn_W 이 다른 방법에 비하여 높음을 볼 수 있다. 평균적으로 제안한 C-DE-MBE 방법의 bpn_W 은 Chen 방법의 2.16배, Huang 방법의 9.03배 높으며, 제안한 DE-MBE 방법의 bpn_W 은 Chen 방법의 1.09배, Huang 방법의 4.57배 높게 나타났다.

3. 염기 변화율 및 아미노산 변화율

원본 은닉 대상 서열 기준으로 워터마크가 은닉된 서열과의 염기 변화율 $e(n)$ 은

$$e(n) = \frac{1}{|\Gamma^{nc}(n)| |D_k^{nc}|} \sum_{k=1}^{|\Gamma^{nc}(n)|} \sum_{i=1}^{|D_k^{nc}|} e_{ki} \quad (25)$$

$$\text{where } e_{ki} = \begin{cases} 1, & \text{if } b_{ik} \neq b'_{ik} \\ 0, & \text{if } b_{ik} = b'_{ik} \end{cases}$$

와 같이 데이터에 의하여 변경된 모든 염기들의 비율을

나타낸다. 그림 7(b)는 수치차수 n 에 대한 염기 변화율 $e(n)$ 을 보여준다. $e(n)$ 은 워터마크 데이터 bpn bpn_W 에 대하여 의존적이며, DE-MBE와 C-DE-MBE 방법의 $e(n)$ 은 n 이 3이상일 때부터 0.189와 0.370에 근접하게 나타난다. 제안한 방법에 비하여 상대적으로 낮은 bpn_W 을 가지는 Chen의 방법과 Huang의 방법의 $e(n)$ 은 0.171과 0.044로 DE-MBE와 C-DE-MBE 방법보다 낮게 나타났다. 특히 Huang 방법은 낮은 염기 변화율을 가지는 가역성 워터마크 방법으로 다른 방법에 비하여 낮은 bpn_W 과 낮은 $e(n)$ 을 가짐을 확인하였다.

비부호 DNA 서열은 Junk DNA으로 가정할 때, 염기 변화율이 높더라도 부호 DNA 서열에 영향을 미치지 않는다. 또한 제안한 방법들은 부호 DNA 서열의 개시 코돈으로 변경되지 않으므로, 아미노산 서열에 전혀 영향을 미치지 않는다. 따라서 아미노산 변화율은 0이다. 만약 비부호 DNA 서열에서 발현 인자 조절과 같은 염기 서열이 알려졌을 경우, 허위 개시코돈 방지와 같은 방법으로 이들 서열이 변경되지 않도록 한다.

V. 결론

본 논문에서는 유전체 기능을 보존하면서 워터마크

를 은닉 및 추출과 원본 DNA 서열을 복원할 수 있는 비부호 영역 가역 DNA 워터마킹 방법을 제안하였다. 제안한 방법은 4-문자 염기 서열을 n 차 수치로 부호화 다음, DE-MBE 방법 또는 C-DE-MBE 방법에 의하여 다중 비트를 은닉한다. 이 때, 높은 은닉 용량을 위하여 수치 쌍별 최대 확장 조건을 만족하는 비트수가 은닉된다. 또한 원본 DNA 서열의 길이와 생물학적 기능의 변경이 없으며, 블라인드 추출 및 복원이 가능하다. 실험 결과로부터 제안한 DE-MBE와 C-DE-MBE 방법들은 참조 서열없이 모두 워터마크 추출과 원본 DNA 서열의 복원되며, C-DE-MBE 방법의 bpn이 DE-MBE 방법의 2배, Chen 방법의 2.16배, Huang 방법의 9.03배 높음을 확인하였다.

제안한 가역 DNA 워터마킹은 DNA를 이용한 정보 저장과 전송, 암호화 등에서 많이 활용될 것이며, 원본 DNA 복원이 가능하고 유전 기능의 손상없이 외부 정보들을 무제한으로 은닉 및 추출하는 응용에 필요할 것이다. 향후에는 자연적 또는 인위적 DNA 변이 (또는 열화)에 강인한 가역 DNA 워터마킹에 대하여 연구하고자 한다.

REFERENCES

- [1] C. T. Clelland, V. Risca, C. Bancroft, "Hiding messages in DNA microdots," *Nature*, vol. 399, pp. 533-534, June 1999.
- [2] D. Tulpan, C. Regoui, G. Durand, L. Belliveau, and S. Léger, "HyDEn: A hybrid steganocryptographic approach for data encryption using randomized error-correcting DNA codes," *BioMed Research International*, Article ID 634832, 2013.
- [3] J. Cox, "Long-term data storage in DNA," *TRENDS in Biotechnology*, vol. 19, no. 7, pp. 247-250, July 2001.
- [4] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E.M. LeProust, B. Sipos, E. Birney, "Towards practical high-capacity, low-maintenance information storage in synthesized DNA," *Nature*, vol. 494, pp. 77-80, Feb. 2013.
- [5] D. Heider and A. Barnekow, "DNA Watermarks - A proof of concept," *BMC Bioinformatics*, vol. 9, no. 40, April 2008.
- [6] M. Liss, D. Daubert, K. Brunner, K. Kliche, U. Hammes, A. Leihner, and R. Wagner, "Embedding Permanent Watermarks in Synthetic Genes," *PLOS ONE*, vol. 7, issue 8, e42465, doi/10.1371/journal.pone.0042465, Aug. 2012.
- [7] S.-H. Lee, "DNA sequence watermarking based on random circular angle," *Digital Signal Processing*, vol. 25, pp. 173-189, Feb. 2014.
- [8] S.-H. Lee, "DWT based coding DNA watermarking for DNA copyright protection," *Information Sciences*, vol. 273, pp. 263-286, July 2014.
- [9] S.-H. Lee, K.-R. Kwon, and S.-G. Kwon, "A Robust DNA Watermarking in Lifting Based 1D DWT Domain," *Journal of The Institute of Electronics Engineers of Korea*, vol. 49, no. 10, pp. 91-101, Oct. 2012.
- [10] S.-H. Lee, S.-G. Kwon, and K.-R. Kwon, "Robust DNA Watermarking based on Coding DNA Sequence," *Journal of The Institute of Electronics Engineers of Korea*, vol. 49, no. 2, pp. 123-133, March 2012.
- [11] T. Chen, "A Novel Biology-Based Reversible Data Hiding Fusion Scheme," *Frontiers in Algorithmics, Lecture Notes in Computer Science*, Vol. 4613, pp 84-95, 2007.
- [12] Y.-H. Huang, C.-C. Chang, and C.-Y. Wu, "A DNA-based data hiding technique with low modification rates," *Multimedia Tools and Applications*, vol. 70, issue 3, pp 1439-1451, June 2014.
- [13] H.J. Shiu, K.L. Ng, J.F. Fang, R.C.T. Lee, and C.H. Huan, "Data hiding methods based upon DNA sequences," *Information Sciences*, vol. 180, issue 11, pp. 2196-2208, June 2010.
- [14] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, issue 8, pp. 890-896, Aug. 2003.
- [15] D. Coltuc, "Improved Embedding for Prediction-Based Reversible Watermarking," *IEEE Trans. on Information Forensics and Security*, vol. 6, no. 3, pp. 873-882, Sept. 2011.

 저 자 소 개



이 석 환(정회원)

1999년 경북대학교 전자공학과
학사 졸업.

2001년 경북대학교 전자공학과
석사 졸업.

2004년 경북대학교 전자공학과
박사 졸업.

2005년~현재 동명대학교 정보보호학과 부교수
<주관심분야 : 영상신호처리, 콘텐츠보안, 3D그래픽스>



권 기 룡(정회원)

1986년 경북대학교 전자공학과
학사 졸업.

1990년 경북대학교 전자공학과
석사 졸업.

1994년 경북대학교 전자공학과
박사 졸업

2000년~2001년 Univ. of Minnesota, Post-Doc

1996년~2006년 부산외국어대학교 컴퓨터전자공
학과 부교수

2006년~현재 부경대학교 IT융합응용공학과 교수
<주관심분야 : 통신, 컴퓨터, 신호처리, 반도체>