오디오 신호에 적용된 Generalized Patchwork Algorithm 의 안전성

Security of Generalized Patchwork Algorithm for Audio Signal

김기섭*, 김형중, 아리나, 양재수
(Ki-Seob Kim, Hyoung-Joong Kim, Arina and Jae-Soo Yang)

Abstract: In this paper we present a cryptanalysis of the generalized patchwork algorithm under the assumption that the attacker possesses only a single copy of the watermarked audio. In the scheme, watermark is inserted by modifying randomly chosen DCT values in each block of the original audio. Towards the attack we first fit low degree polynomials (which minimize the mean square error) on the data available from each block of the watermarked content. Then we replace the corresponding DCT data of the attacked audio by the available data from the polynomials to construct an attacked audio. The technique nullifies the modification achieved during watermark embedding. Experimental results show that recovery of the watermark becomes difficult after the attack.

Keywords: Cryptanalysis, Digital Watermarking, Information Security, Multimedia Systems, Discrete Cosine Transform

I. 서론

지난 몇 년 동안 디지털 컨텐츠의 저작권 보호를 위하여 디지털 워터마킹의 기술이 빠르게 성장하고 있다. 이러한 기술은 이미지 뿐 아니라 오디오 신호를 근원으로 하고 있는 디지털 음악 분야에도 빠른 속도로 진행되고 있다. 현재 수많은 워터마크 기술들은 다양한 신호처리를 통한 공격에 강인함을 보이고 사용자의 신뢰를 얻기 위하여 복잡한 신호처리 기술을 사용하고 있다. 현재까지 알려진 대부분의 영상 워터마킹 기술은 회전, 일부 잘라내기, 기하학적인 변형 [12],[13]과 같은 공격을 이겨낼 수 있는 워터마킹 기술에 대하여 이야기 하고 있고 암호학적인 평가에 대하여는 언급하지 않고 있었다. 최근 가장 성공적인 공격의 방법으로 사용되는 것은 single watermarked copy 가 사용된다. 동일한 측면에서 우리는 generalized patchwork algorithm 의 single copy 공격에 대한 보안성에 대하여 논의한다.

일반적인 형태의 보이지 않는 워터마크의 기술은 원본신호인 I 에 워터마크 신호인 $s(i)$의 신호를 삽입하여 $I(i)$의 신호를 생성하게 된다. 우리는 소비자에게 디지털 컨텐츠를 판매할 때 i 번째 소비자에게 $s(i)$가 삽입된 $I(i)$의 디지털 컨텐츠를 판매함으로써 모든 구매자에게 다른 디지털 컨텐츠를 판매할 수 있다. 여기서 $s(i)$는 소비자의 정보를 포함하게 하여 혹시 발생할지도 모르는 법적인 문제에 대처할 수 있게 하고 있다.

일반적으로 암호학과 정보보호론의 기법은 일반적으로 공격자에게 잘 알려져 있다. 그러나 키 값이나 워터마크의 값은 알려지지 않는 것이 일반적이다. 그래서 워터마킹 기법을 아주 잘 알고 있는 암호학 전문가라 하더라도 비밀키를 알지 못하는 한 숨겨져 있는 워터마크 신호를 제거하기 어렵다. 이러한 기본 원리는 1883 년 전략가인 Kerckohoffs[8]에 의하여 처음으로 제시되었다. 최근 들어서는 Secure Digital Music Initiative(SDMI)에서 이것에 도전하고 있다.

워터마킹의 일반적인 과정에서 공격을 받은 미디어인 $I^{\#}$ 과 $I^{(i)}$ 에서 검출한 신호에는 서로간에 상호연관성에 아무런 의미를 가지고 있지 않은 특징을 지니고 있다. 이러한 신호가 만일 비트 스트림 신호일 경우에도 동일하게 적용될 수 있을 것이다. 이 신호를 사용하여 우리는 악의를 지니고 있는 구매자가 그것을 구매하여 악의적으로 사용을 할 경우에도 구별하기 매우 어려운 문제를 가지고 있게 된다. 또한 미디어 자체도 불법으로 생성된 미디어와 정당한 방법으로 제공된 미디어의 구별에 많은 어려움을 지닐 것이다.

2장에서는 Generalized Patchwork Algorithm에 대하여 기술할 것이며 3장에서는 제안한 기법의 실험 결과에 대하여 설명하였다. 마지막으로 4장에서는 본 논문의 결론에 대하여 이야기 하는 것으로 마무리 할 것이다.

* 책임저자(Corresponding Author)
김기섭, 김형중, 아리나: 강원대학교 정보보호대학원
양재수: 광운대학교
kskim@multimedia.kangwon.ac.kr, khj@kangwon.ac.kr,
arina@multimedia.kangwon.ac.kr

※ 본 연구는 고려대학교 ITRC의 지원을 받아 연구되었음.

II. Generalized Patchwork Algorithm

Generalized Pathcwork Algorithm을 설명하기 위하여 이미지 컨텐츠를 이용하였다. $M\times M$크기의 이미지를 $N\times N$크기로 DCT를 수행한다. 여기에서 비밀키 K, watermarking signal w 를 사용하고 있으며

$$w=[w_1,w_2,\cdots,w_t]\quad t=\frac{M\times M}{N\times N}$$

비트 패턴을 사용하고 있다.

1. Bit Embedding

DCT를 실행한 각 블록에 한 비트의 신호를 삽입하는 과정을 진행한다. 이 과정을 위하여 우리는 랜덤한 인덱스 값을 생성하기 위하여 키 K를 Seed값으로 사용하여 유사난수를 통하여 인덱스 값들을 생성한다. 이때 생성된 인덱스 값은 아래의 조건을 갖게 된다.

$$[Z_1,Z_2],\ 1\le Z_1<Z_2\le N\times N$$

이때 생성된 인덱스 값은 DCT블록에서 JPEG기법의 지그재그스캔의 순서와 같이 구성된다. 이렇게 생성된 인덱스 값은 I^0 와 I^1 의 2개의 집합으로 구성이 되며 각각의 집합은 $2n$개의 요소를 갖게 된다.

$$I^0=I^{0+}\cup I^{0-},\quad I^{0+}=\{I^0_1,\cdots,I^0_n\},\quad I^{0-}=\{I^0_{n+1},\cdots,I^0_{2n}\}$$

$$I^1 = I^{1+} \cup I^{1-},\quad I^{1+} = \{I_1^1, \cdots, I_1^1\},\quad I^{1-} = \{I_{n+1}^1, \cdots, I_{2n}^0\}$$

여기서 삽입하고자 하는 워터마크의 비트값이 0과 1일 때 우리는 I^0 와 I^1의 값을 사용하여 원하는 비트의 값을 삽입하게 되는 것이다.

이렇게 생성된 인덱스 값을 기준으로 하여 A^j와 B^j의 값을 생성하여 삽입하고자 하는 1과 0의 값에 따라 A^j와 B^j를 결정하고 결정된 값을 이용하여 워터마크의 값을 삽입하게 된다. 만일 0 비트를 삽입할 경우 진행되는 과정은 아래와 같이 진행된다.

$$\overline{A^0} = \frac{1}{2}\sum_{i=1}^{n} A_i^0,\quad \overline{B^0} = \frac{1}{2}\sum_{i=1}^{n} B_i^0$$

$$S_{A^0}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(A_i^0 - \overline{A^0}),\quad S_{B^0}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(B_i^0 - \overline{B^0})$$

$$A_i^{0*} = (1 + sign(S_{A0}^2 - S_{B0}^2)P1)A_i^0 + sign(\overline{A^0} - \overline{B_0})\sqrt{P_2}\frac{S_{E0}}{2}$$

$$B_i^{0*} = (1 + sign(S_{A0}^2 - S_{B0}^2)P1)B_i^0 - sign(\overline{A^0} - \overline{B_0})\sqrt{P_2}\frac{S_{E0}}{2}$$

$$S_{E0} = \sqrt{\frac{\sum_{i=1}^{n}(A_i^0 - \overline{A^0})^2 + \sum_{i=1}^{n}(B_i^0 - \overline{B^0})^2}{n(n-1)}}$$

4*n*개의 인덱스를 생성하고 실제 워터마크의 삽입 과정에서는 2*n*개의 요소값들이 변화된다.

2. Bit Extraction

워터마크를 검출하는 과정에서 우리는 원본 미디어를 필요로 하지 않고 워터마크가 삽입된 미디어만을 이용하여 우리가 원하는 워터마크의 신호를 복원하여 확인할 수 있다. 이 과정에서 인덱스 값을 생성하기 위한 시드값인 *K*만을 필요로 하여 진다.

비트를 검출하기 위한 기본 식은 아래와 같이 구성된다.

$$T_0^Q = \beta\max\{\frac{S_{A^*}^2}{S_{B^*}^2}, \frac{S_{B^*}^2}{S_{A^*}^2}\} + (1-\beta)\frac{(\overline{A^0} - \overline{B^0})^2}{S_{E0}^2}$$

$$T_1^Q = \beta\max\{\frac{S_{A^1}^2}{S_{B^1}^2}, \frac{S_{B^1}^2}{S_{A^1}^2}\} + (1-\beta)\frac{(\overline{A^1} - \overline{B^1})^2}{S_{E1}^2}$$

$$P_1 = 0 \quad and \quad P_2 > 0, \quad then \quad \beta = 0,$$

$$P_1 > 0 \quad and \quad P_2 = 0, \quad then \quad \beta = 1,$$

이 경우 $T_0^Q > T_1^Q$이면 우리가 검출하게 되는 워터마크 비트의 값은 0을 검출하게 된다. 그 이외의 경우에는 검출되는 비트는 1이라는 값을 결과로 검출하게 된다.

이러한 검출의 방법을 통하여 삽입된 워터마크를 검출하게 되며 또한 원본의 미디어가 필요 없는 형태의 워터마크 기법을 진행하게 되는 것이다

III. Generalized Patchwork Algorithm 의 평가 실험

본 실험을 진행하기 위하여 총 5개의 오디오를 대상으로 실험을 진행하였다. 5개의 오디오는 모든 오디오의 특성을 살펴보기 위하여 5개의 각기 다른 장르의 음악을 선정하였으며 Audio 1은 PopMusic, Audio 2는 조용한 음악 Audio 3는 클래식 음악 Audio 4는 악기의 연주음악 그리고 마지막으로 Audio 5는 자연의 소리를 녹음한 음악을 사용하였다. 이렇게 5가지 형태의 음악을 사용하여 테스트를 진행함으로써 모든 오디오의 특성에 대한 실험은 아니지만 우리가 가질 수 있는 대부분의 특성을 가질 수 있는 음악을 대상으로 하여 실험했다.

아래의 그림은 5개의 음악 파일의 각각의 파형을 그림으로 나타낸 것이다.

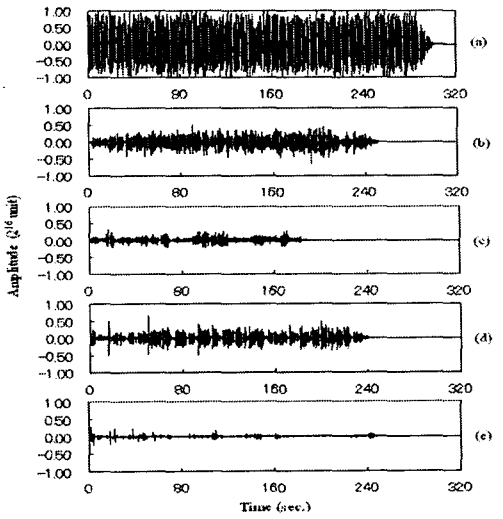

그림 1. 테스트 오디오 파형

아래의 실험에서는 워터마킹 인코딩 과정을 여러 번 반복하여 삽입하였을 때 Generalized Patchwork Algorithm이 지니는 특성에 대하여 실험하였으며 또한 커브피팅을 진행한 결과에 대한 실험도 진행하였다. 첫번 째 반복 삽입을 통한 실험은 5개의 음악샘플을 대상으로 실험한 결과 중 Audio 2와 Audio 5에 해당하는 데이터를 표로 제시하였다. 각각의 값은 반복적으로 각기 다른 키값을 가지고 워터마크를 삽입하고 이때 삽입된 워터마크를 검출했을 때 나타나는 특징을 표로 보여주고 있다.

두 번째 실험을 워터마키가 삽입된 오디오 컨텐츠에서 워터마크가 커브피팅 공격을 진행하였을 때 얼마나 많은 데이터가 추출되며 또한 오디오의 품질은 어떻게 변화되는 지를 실험하였다.

3.1 워터마크 신호의 반복 삽입을 통한 공격

오디오 신호에 워터마크의 삽입을 키의 변경을 통하여 진행하였다. 총 30회의 워터마킹 삽입을 진행하였을 때 워커마킹의 검출된 검출율과 원신호와의 SNR일 비교하여 작성하였다.

워터마크를 검출했을 때 검출된 워터마크의 개수를 나타내고 있다. 원본 오디오에는 각각 50개의 워터마크를 삽입하였다. 두 번째 표가 나타내고 있는 부분은 원본 신호와

공격을 진행하였을 때 만들어진 오디오 신호와의 SNR을 비교한 도표이다.

표 1. Audio 2의 실험 결과 (조용한 음악)

공격횟수	검출수	SNR
1	50	31.9519
5	50	31.9519
10	50	22.4064
15	50	18.3381
20	50	15.3993
25	50	13.7464
30	50	12.4112

표 2. Audio 5의 실험 결과 (자연의 소리)

공격횟수	검출수	SNR
1	48	17.8503
5	47	15.4375
10	49	13.2722
15	49	11.2775
20	49	9.5824
25	49	7.9759
30	50	6.6496

실험의 결과에서 확인할 수 있듯 박복적인 복사를 통한 공격에서는 Generalized Patchwork Algorithm의 결과는 워터마크의 검출에 있어서는 거의 모든 워터마크 신호를 검출해 낼 수 있었다. 반면 원본 신호와의 차이에서는 공격이 많아 질수록 워터마크가 삽입된 오디오의 신호가 원본신호와 많은 차이를 가지고 있는 것을 확인할 수 있었다. 이것은 Generalized Patchwork Algorithm이 박복적인 삽입의 공격을 통한 공격범에서는 충분히 좋은 결과를 보여주고 있는 것을 확인 할 수 있다.

3.2 Curve-Fitting 공격을 진행한 결과

커브 피팅 공격방법에 대하여 설명을 하면 워터마크가 삽입된 오디오 컨텐츠를 4410 샘플의 윈도우 크기로 분리하여 각 샘플들의 집합을 정렬한다. 이때 정렬된 샘플들의 값을 여러가지 형태의 함수들을 사용하여 커브피팅을 진행하였다. 이번 실험에서 사용한 함수는 3차방정식과 4차 방정식 그리고 3차 지수함수와 4차 지수함수를 사용하여 실험을 진행하였다. 적용되는 함수의 형태는 아래와 같다.

$a_1x^3 + a_2x^2 + a_3x + a_4$ 3차 방정식

$a_1x^4 + a_2x^3 + a_3x^2 + a_4x + a_4$ 4차 방정식

각각의 경우 Matlab에 존재하는 lsqcurvefit함수를 사용하여 커브 피팅을 진행하였다. 실험의 결과에서 동일한 함수를 사용하더라도 lsqcurvefit함수에서 사용하고 있는 초기값에 따라 서로 다른 결과를 보여주고 있다. 실험의 결과에서 3차와 4차의 지수함수 형태에서는 삽입된 50개의 워터마크 신호를 모두 검출해내는 특징을 보여주고 있으며 3차와 4차 방정식을 통한 커브피팅을 진행하였을 때에는 초기값과 함수의 차수에 따라 검출률에 많은 차이를 보이고 있었다. 예를 들어 3차 방정식을 사용할 경우 초기 값을 [1, 1, 1, 1]을 사용하여 커브피팅을 진행하면 50개의 삽입된 워터마크를 모두 검출하고 있다. 이때 SNR의 값은 4.7836의 값을 가지고 있다. 반면 초기 값을 [0, 0, 0, 0]을 사용하여 계산을 하면 이때는 50개의 워터마크 가운데 13개만을 검출하여 많은 워터마크 신호를 검출하지 못하는 특징을 보여주고 있다. 이때 SNR의 값은 5.1432를 보이고 있다. 아래의 표는 오디오 2와 오디오 3를 대상으로 적용한 방정식에 따른 워터마크 삽입개수와 SNR의 값을 나타내고 있다.

표 3. Audio 2의 실험 결과

방정식	초기값	검출수	SNR
3차방정식	[1 1 1 1]	50	4.7836
4차방정식	[0 0 0 0]	13	5.1432
4차방정식	[0 0 0 0 0]	5	5.1355

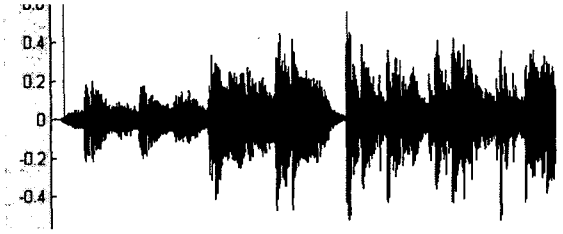

그림 2. 워터마크가 삽입된 Audio 2 (공격전)

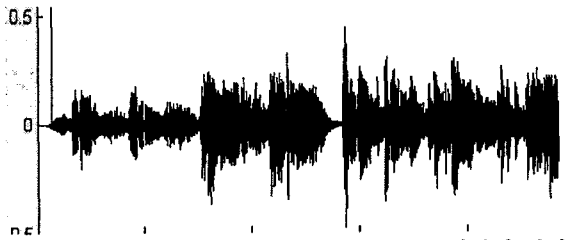

그림 3. Audio 2를 4차 방정식을 통하여 커브피팅한 결과

표 4. Audio 3의 실험 결과

방정식	초기값	검출수	SNR
3차방정식	[1 1 1 1]	50	3.5762
4차방정식	[0 0 0 0]	42	4.4808
4차방정식	[0 0 0 0 0]	20	4.4754

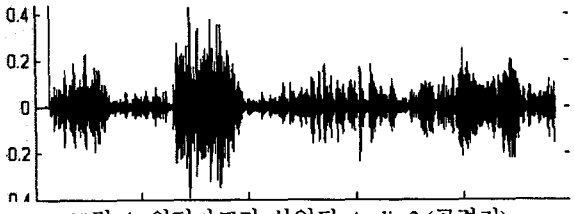

그림 4. 워터마크가 삽입된 Audio 3 (공격전)

그림 5. Audio 3를 3차 방정식을 통해 커브피팅한 결과

일반적으로 오디오 신호는 이미지에 비해 사람의 청각에 매우 민감한 반응을 보이고 있다. 위의 그림에서 보듯이 공격을 진행한 결과와 공격 진행전의 파형을 비교해 봤을 때 조금의 차이가 느껴진다. 그러나 실제 음악을 들었을 경우 조금의 차이는 많은 차이를 갖고 있다는 것을 확인 할 수 있다.

위의 표에서 보았을 때 Generalized Patchwork Algorithm의 경우 커브피팅공격을 진행하였을 경우 워터마크가 많이 사라지는 것을 확인할 수 있었다. SNR의 값을 검출 수와 같이 비교하여 보면 원본 신호와의 차이가 많이 발생하는 것을 확인할 수 있을 것이다. 실제 공격후의 음악을 청취할 경우 음악 자체에 많은 잡음과 손실이 가해진 것을 확인할 수 있다.

IV. 결론

본 실험의 결과를 통하여 기존에 제안된 Generalized Patchwork Algorithm의 성능에 대한 평가를 진행하였으며 그 결과 오디오 파일에서 워터마크 신호를 중복하여 삽입하는 방법을 사용하였을 경우 워터마크의 손실은 없는 것으로 확인되었으며 반면에 커브피팅 공격을 하였을 경우에는 삽입된 워터마크 신호가 많이 손상되는 것을 확인할 수 있었다. 그러나 커브피팅 공격을 진행하였을 경우 사용한 오디오 컨텐츠가 많이 손상되는 것을 확인할 수 있다.

Generalized Patchwork Algorithm의 경우 통계적인 방법을 사용하여 워터마크를 삽입하여 다양한 공격을 진행하였을 경우 강인성을 보이고 있었다. 그러나 커브피팅이라는 방법을 사용하였을 경우에는 워터마크가 많은 손상을 입어 훼손되는 문제가 존재하고 있다.

앞으로는 커브피팅 기법의 수정과 최적화된 초기값검출을 통하여 오디오신호의 훼손을 최대한 줄여 원본신호에 영향을 적게 끼치는 형태의 공격 방법에 대한 실험의 진행을 지속하여야 할 것이다.

참고문헌

[1] R. J. Anderson and F. A. P. Petitcolas, "On the limits of steganography," *IEEE Journal of Selected Areas in Communications*, vol. 16, no. 4, pp. 474-481, May 1998.

[2] J. Boeuf and J. P. Stern, "An analysis of one of the SDMI candidates,", *Lecture Notes in Computer Science*, vol 2137, pp. 395-410, 2001.

[3] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673-1687, 1997.

[4] T. K. Das, S. Maitra, and J. Mitra, "Cryptanalysis of optimal differential energy watermarking (DEW) and a modified robust scheme," *IEEE Transactions on Signal Processing*, PART II, vol. 53, no. 2, pp. 768-775, February 2005.

[5] T. K. Das and S. Maitra, "Cryptanalysis of correlation based watermarking schemes using single watermarked copy," *IEEE Signal Processing Letters*, vol. 11, no. 4, pp. 446-449, April 2004.

[6] F. Ergun, J. Kilian and R. Kumar, "A note on the limits of collusion-resistant watermarks," *Lecture Notes in Computer Science,*, vol. 1592, pp. 140-149, 1999.

[7] T. Kalker, J. P. M. G. Linnartz, and M. v. Dijk, "Watermark estimation through detector analysis," *Proceedings of the International Conference on Image Processing*, 1998.

[8] A. Kerckhoffs, "La cryptographie militaire," *Journal des Sciences Militaires*, 9th series, IX, pp. 5-38, January 1883, and pp. 161-191, February 1883.

[9] D. Kirovski and H. S. Malvar, "Embedding and detecting spread-spectrum watermarks under the estimation attack," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002.

[10] D. Kirovski and F. A. P. Petitcolas, "Replacement attack on arbitrary watermarking systems," *ACM Workshop on Digital Rights Management*, 2002.

[11] M. K. Mihcak, R. Venkatesan, M. Kesal, "Cryptanalysis of discrete-sequence spread spectrum watermarks," *Lecture Notes in Computer Science*, vol. 2578, pp. 226-246, 2003.

[12] F. A. P. Petitcolas, R. J. Anderson, M. G. Kuhn, and D. Aucsmith, "Attacks on copyright marking systems," *Lecture Notes in Computer Science*, vol. 1525, pp. 218-238, 1998.

[13] F. A. P. Petitcolas and R. J. Anderson, "Evaluation of copyright marking systems," *IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, vol. 1, pp. 574-579, June 1999.

[14] I.-K. Yeo and H. J. Kim, "Generalized patchwork algorithm for image watermarking," *Multimedia Systems*, vol. 9, pp. 261-265, 2003.

[15] I.-K. Yeo and H. J. Kim, "Modified patchwork algorithm: A novel audio watermarking scheme," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 381-386, July 2003.