불법 유통 방지를 위한 핑거프린팅 코드에 관한 연구

이진흥*, 박지환*
*부경대학교 대학원 정보보호학과
e-mail: jhlee@shannon.pknu.ac.kr, jpark@pknu.ac.kr

A Study on Fingerprinting Code for Illegal Distribution Prevention

Jin-Heung Lee*, Ji-Hwan Park*
*Interdisciplinary Program of Information Security, Pukyong Nat'l Univ.

요 약

핑거프린팅은 컨텐츠 유통 시, 구매자의 정보를 컨텐츠에 삽입함으로써 불법 유통 행위된 컨텐츠에 대하여 불법 배포자를 추적할 수 있는 기법이다. 이 방법에는 서로 다른 구매자에 의한 핑거프린팅 코드를 제거하려는 공모 공격(collusion attacks)이 발생할 수 있다. 본 논문에서는 고차의 원시다항식을 이용하여 효율적이면서 공모 공격에 강인한 핑거프린팅 코드 구성 방법을 제안한다. 그리고, 제안된 방법을 오디오 데이터에 적용하여 공모 공격에 대하여 안전한 코드임을 보이고 있다.

1. 서론

디지털 워터마킹 기술은 멀티미디어 저작물을 보호하기 위하여 특정 형태의 워터마크 정보(저작권 정보, 로고, 일련번호 등)를 원 데이터에 감추고 추출하는 모든 기술적 방법을 말한다. 초기에는 원래의 멀티미디어 데이터 자체의 조작으로 워터마크 정보를 은닉시키는 방법이 개발되다 현재에는 많은 기술적 변환 방법을 이용하여 강인한 워터마킹 기술이 개발되고 있다[1].

핑거프린팅 기술은 워터마킹의 확장 기술로 컨텐츠의 상거래 시 소유자의 정보뿐만 아니라 구매자의 정보도 포함하는 핑거프린팅 정보를 컨텐츠에 삽입하여 불법 배포가 어느 구매자로부터 시작되었는지 추적할 수 있도록 해주는 기술이다. 핑거프린팅 기술은 고유한 저작권 정보를 삽입하여 소유권 분쟁의 증거로 활용하는 디지털 워터마킹과는 달리 복수개의 워터마크를 핑거프린팅 코드로 삽입해야 한다. 따라서, 다수의 사용자에 의한 공모 공격이 가능하게 된다.

이와 같은 공모 공격에 대한 대책의 하나로서 구매자의 코드를 유한 사영기하학을 기반으로 공모자가 d명일 때 강인하도록 코드를 설계한 d-detecting 코드가 있다. 그러나 이 방법에는 구매자의 수가 늘어나면 코드 길이가 기하급수적으로 증가하는 어려움을 가지게 된다. 따라서, 공모 공격에 대해 강인성을 유지하면서 짧은 코드 길이를 가지고, 확장이 용이한 코드 구성법이 요구된다.

본 논문에서는 새로운 핑거프린팅 코드 생성 방법을 제안한다. 제안하는 기법에서는 유한체의 원시다항식(*primitive polynomial*)을 이용하여 *trace*를 생성하고, 생성된 *trace*에 의해 순환 행렬을 구성하여 공모공격에 강인한 효율적인 핑거프린팅 코드를 생성하였다. 본 논문의 구성은 다음과 같다. 먼저, 2장에서 핑거프린팅 코드에 대한 기존의 방식에 대하여 설명하고, 핑거프린팅 코드의 공모 공격에 대하여 설명한다. 그리고, 제안한 핑거프린팅 코드 생성 방법에 대하여 3장에서 설명하고, 4장에서 디지털 오디오 신호를 이용하여 제안 방법의 안전성에 대한 실험 결과를 나타낸다. 끝으로 향후 연구 방향에 대하여 기술한다.

2. 관련 연구

핑거프린팅 기술은 워터마킹 기술을 바탕으로하여 공모공격에 강인한 프로토콜 및 코드 구성에 대하여 연구가 진행되어 왔다. 핑거프린팅 기술은 크게 대칭성과 익명성을 지원하는 암호학적 기법과 Malvar 등에 의해 제안된 이중 워터마킹/핑거프링팅 기법으로 구분된다[2,3]. 또한 삽입코드 자체를 공모공격이

본 연구는 한국과학재단 특정기초연구(R01-2002-000-00589-0) 지원에 의해 수행 되었음

불가능하도록 설계하는 공모공격에 안전한 코드 개발기법이 있다.

2.1 대칭형과 비대칭형 핑거프린팅

대칭형 핑거프린팅은 초기의 연구기법으로 판매자에 의해 구매자를 식별하고, 판매된 리스트를 입력하여 핑거프린팅을 수행하여 핑거프린팅된 컨텐츠와 구매 코드를 생성한다. 이 경우, 컨텐츠의 불법 유통시 판매자는 불법 유통된 컨텐츠와 원본, 그리고 구매자의 기록에 의해 복사본의 원 구매자를 찾게된다. 하지만 이 방법에는 판매자와 구매자가 핑거프린팅된 컨텐츠에 접근 가능하므로 불법 유통된 컨텐츠의 배포자를 구매자인지 판매자인지 판단하기 어려운 문제점이 있다.

비대칭형 핑거프린팅 기술은 대칭형 핑거프린팅의 문제점을 해결하기 위해 제안된 방식이다. 판매자와 구매자 사이의 프로토콜에서 핑거프린팅된 컨텐츠에 대한 접근은 구매자만이 가능하게 하고, 불법 유통된 컨텐츠의 배포자 추적은 신뢰 센터를 통하여 이루어진다.

2.2 공모공격에 강인한 핑거프린팅 코드 생성

핑거프린팅된 컨텐츠간의 차이점을 이용한 공모공격은 적은 수의 컨텐츠로 핑거프린팅 정보를 제거할 수 있다. 따라서 이러한 공모 공격에 강인한 핑거프린팅 코드를 삽입해야 한다. 일반적으로 핑거프린팅 코드는 구매자마다 연관성이 없는 랜덤 순열을 사용하여 어느 정도의 공모 공격에 대한 강인성을 제공한다. 그러나 공모자가 많아질수록 필요한 코드의 수가 기하급수적으로 증가한다는 문제점을 가진다. 이러한 문제점을 해결하기 위해 구매자마다 다른 위치에서 공통된 부분을 가지는 코드에 대한 설계 및 연구가 진행되고 있다.

Boneh와 Shaw는 같은 위치에 동일한 마크가 삽입되지 않았을 경우에는 검출 가능하다는 삽입가정(marking assumption)을 바탕으로 공모공격에 강인한 핑거프린팅 코드를 제안하였다[4]. 마크는 다른 삽입 위치를 갖는 제한된 수의 코드 조합이고 핑거프린팅 코드는 이 마크의 집합이다.

Dittmann은 사영 평면(projective plane)을 바탕으로 공모 공격에 강인한 핑거프린팅 코드를 생성하고 이것을 이미지에 적용하였다[5]. 이 방법은 3명의 배포자중 2명이 서로 공모하였을 때 공모자를 추출 가능하다. 또한, 유한 사영 기하학을 기반으로 하여 공모자가 d명일 때 모든 공모자를 검출할 수 있는 d-detecting 코드를 제안하였다. 그러나, 이 방법에서는 구매자의 증가에 따른 사영평면 구성의 어려움으로 구매자 수에 제한을 받는 문제점을 가지고 있다. 또한, Trappe와 Wu는 BIBD를 이용하여 멀티미디어 데이터를 위한 AAC(Anti-Collusion Code)를 이용하여 공모자를 검출할 수 있는 알고리즘을 제안하였다[6].

이러한 방법들은 2,3명의 사용자들이 공모하여 만든 컨텐츠로부터 최소 한 명의 공모자를 추출하기 위한 방법이다. 이러한 코드의 단점은 구매자가 많아질수록 필요로 하는 코드의 길이가 기하급수적으로 증가하기 때문에 구매자가 많은 인터넷 환경에서는 적용이 매우 어렵게 된다. 또한 코드의 형태도 단순하여 공격자에 의해 쉽게 예측 가능한 문제점도 가지고 있다.

3. 새로운 핑거프린팅 코드의 구성법

디지털 핑거프린팅 코드는 일반적으로 다음과 같은 항목들을 만족해야 한다. 먼저, 합법적인 사용자와 불법적으로 유통한 사용자를 정확하게 검출하여야 하고, 다른 시스템 내에서 효율적으로 수행가능해야 한다. 또한, 핑거프린팅 코드는 완벽한 지식이 없는 이들에 의해 삭제 및 변경이 어려워야 한다. 마지막으로 불법 배포된 컨텐츠의 작은 부분에서도 배포자를 구분 가능해야 한다. 제안 방식은 유한체상의 원시 다항식을 이용하여 효율적이고 공모 공격에 강인한 핑거프린팅 코드를 생성하는 것이다.

3.1 원시 다항식

주어진 다항식 $f(x)$에 대하여 인수분해 가능한 경우와 인수분해 불가능한 다항식이 있을 때 m차인 인수분해 불가능한 다항식으로서 $x^n+1(n=2^m-1)$은 나눌 수 있고, $x^i+1(1<i<2^m-1)$은 나눌 수 없을 때, 이때의 다항식을 원시 다항식이라 한다. GF(2)상의 벡터공간 GF(2^m)은 원시 다항식으로 정의될 수 있으며, 다음 식(1)과 같이 표현된다.

$$p(x)=x^m + p_{m-1}x^{m-1} + \cdots + p_1x + p_0$$
$$(p_i \in GF(2)\ (0 \le i \le m-1)) \qquad (1)$$

표준 기저에서 기저 벡터는 $\alpha^0, \alpha^1, \cdots, \alpha^{m-1}$이고, 이때의 α를 원시 다항식의 근으로서 필드의 원시 원소(*primitive element*)라 한다. 어떠한 GF(2^m)상의 원소 A도 식(2)로 나타낼 수 있으며, 원시 원소를 이용하여 A를 식(3)과 같은 다항식으로 표현할 수 있다. 즉, GF(2)상의 m차원 벡터로서 $\{a_0, a_1, \cdots, a_{m-1}\}$로 나타낼 수 있다.

$$A = a_0\alpha^0 + a_1\alpha^1 \cdots + a_{m-1}\alpha^{m-1} \quad (a_i \in GF(2),\ 0 \le i \le m-1) \qquad (2)$$

$$A(x) = a_0x^0 + a_1x^1 + \cdots + a_{m-1}x^{m-1} \qquad (3)$$

3.2 *trace*

F=GF(q), K=GF(q^n)이라고 두면, 아래의 정리1에 의해 K의 서브필드로서 F를 볼 수 있다. 만일 α가 K의 요소라면, 서브필드 F에 대한 *trace relative*는 식(4)와 같이 규정할 수 있다. 그리고 이러한 *trace*는 정리2와 같은 기본적인 성질들을 가진다[7].

$$Tr\,{}_K^F(\alpha) = \alpha + \alpha^q + \alpha^{q^2} + \cdots + \alpha^{q^{n-1}} \qquad (4)$$

정리1. n의 모든 약수 d에 의해, GF(p^n)은 정확하게 하나의 GF(p^d)와 동형의 서버필드를 포함한다.

정리2. 임의의 α,β∈K에서,

(1) $Tr(\alpha)\in F$

(2) $Tr(\alpha+\beta)=Tr(\alpha)+Tr(\beta)$

(3) $Tr(\lambda\alpha)=\lambda Tr(\alpha)$, $\lambda\in F$

(4) $Tr(\alpha^q)=Tr(\alpha)$

(5) Tr: $K\rightarrow F$

원시 다항식 $p(x)=x^4+x+1$에 대하여, 원시 원소와 *trace*를 구하면 표1과 같다. GF(q)에서 원의 수 q를 *order*라 부르며, 여기에서 $ord(\alpha^i)$는 '2^m 원소의 총 개수-1'의 약수로 정의된다.

[표 1] GF(2^4)의 원시 다항식 $p(x)=x^4+x+1$

i	α^i	다항식표현	벡터표현	$ord(\alpha^i)$	$Tr(\alpha^i)$
0	α^0	1	0001	1	0
1	α^1	α	0010	15	0
2	α^2	α^2	0100	15	0
3	α^3	α^3	1000	5	1
4	α^4	α+1	0011	15	0
5	α^5	$\alpha^2+\alpha$	0110	3	0
6	α^6	$\alpha^3+\alpha^2$	1100	5	1
7	α^7	$\alpha^3+\alpha$+1	1011	15	1
8	α^8	α^2+1	0101	15	0
9	α^9	$\alpha^3+\alpha$	1010	5	1
10	α^{10}	$\alpha^2+\alpha$+1	0111	3	0
11	α^{11}	$\alpha^3+\alpha^2+\alpha$	1110	15	1
12	α^{12}	$\alpha^3+\alpha^2+\alpha$+1	1111	5	1
13	α^{13}	$\alpha^3+\alpha^2$+1	1101	15	1
14	α^{14}	α^3+1	1001	15	1

3.3 원시 다항식을 이용한 2-detecting 핑거프린팅 코드 구성

원시 다항식을 이용한 2-detecting 핑거프린팅 코드 구성은 다음과 같은 절차에 의해 이루어진다.

(1) 주어진 원시 다항식 $p(x)$에 대하여 근 α를 구한다.

(2) α로부터 *trace*를 계산한다.

- $Tr(1)=1+1+1+1=0$
- $Tr(\alpha)=\alpha+\alpha^2+\alpha^{2^2}+\alpha^{2^3}=\alpha+\alpha^2+\alpha^4+\alpha^8$ $= 0$
- $Tr(\alpha^2)=\alpha^2+\alpha^{2^2}+\alpha^{2^4}+\alpha^{2^5}=\alpha^2+\alpha^4+\alpha^8+\alpha^1$ $= 0$

……

- $Tr(\alpha^{14})=\alpha^{14}+\alpha^{14^2}+\alpha^{14^4}+\alpha^{14^8}$ $= \alpha^{14}+\alpha^{13}+\alpha^{11}+\alpha^7 = 1$

(3) *trace*를 이용한 $2^m-1\times 2^m-1$ 순환수열 M을 생성한다.

$$M = \begin{pmatrix} 000100110101111 \\ 001001101011110 \\ 010011010111100 \\ 100110101111000 \\ 001101011110001 \\ 011010111100010 \\ 110101111000100 \\ 101011110001001 \\ 010111100010011 \\ 101111000100110 \\ 011110001001101 \\ 111100010011010 \\ 111000100110101 \\ 110001001101011 \\ 100010011010111 \end{pmatrix}$$

(4) 생성된 M으로부터 각 행의 수열을 핑거프린팅 정보로서 유통될 컨텐츠에 삽입한다.

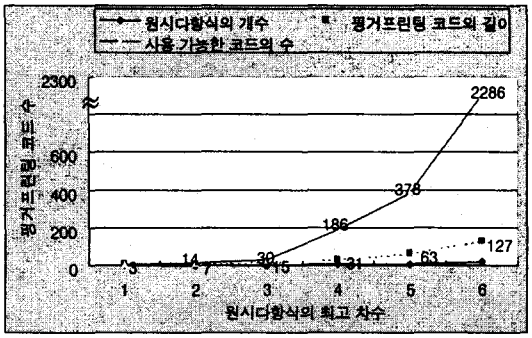
4. 제안 방법에 대한 평가 및 고찰

4.1 핑거프린팅 코드 생성에 관한 고찰

기존의 Dittmann에 의해 제안된 d-detecting 코드에서는 7비트의 핑거프린팅 코드를 이용하여 7명 미만의 사용자에게 분배 가능하다. 제안 방법은 7명의 사용자를 위해 3차 원시 다항식을 이용하여 7비트로 이루어진 핑거프린팅 코드를 사용할 수 있다. 기존의 방법에서는 임의의 사용자 확장시, 코드 생성 및 추가 작업이 어려운 문제점을 가지고 있으나, 제안 방법은 핑거프린팅 코드의 새로운 생성을 위해서 서로 다른 원시 다항식을 적용하므로 쉽게 코드의 생성 및 추가가 가능하다.

이용자의 증가에 따라 원시 다항식의 차수를 높여서 코드를 생성하므로 기존의 방법에 비해 훨씬 효율적인 코드 생성이 가능하다. 그림1은 원시 다항식의 차수에 대한 구매자의 수와 핑거프린팅 코드의 길이에 관하여 나타내고 있다. 고차의 원시 다항식

을 이용하여 기존의 핑거프린팅 코드에서 가지는 다수의 구매자에 대한 문제점을 해결하고 상대적으로 코드의 길이를 줄일 수 있다.



[그림 1] 원시 다항식 및 사용 가능한 핑거프린팅 코드 수

4.2 공모 공격에 관한 고찰

공모자가 2명 이라고 할 때, 각각의 구매자에 대한 핑거프린팅 코드는 위에서 구해진 행렬 M의 행벡터로 생성된다. 임의의 행벡터로 서로 다른 구매자에 대한 핑거프린팅 코드를 삽입한 컨텐츠는 서로 다른 코드끼리의 1/4 개 만큼의 공통부분을 가지게 된다. 이 공통부분의 위치 벡터는 구매자마다 유일하게 구성되며, 이러한 위치 벡터를 이용하여 공모 공격을 시도한 부정자 2명의 검출이 가능하다.

[표 2] 제안 방식에 대한 강인성 평가

공격 형태	평균 공격	모자익 공격		평균공격+ mp3 압축	모자익공격 +mp3 압축
		1000 sampls	1초		
공모자 검출률	100%	100%	100%	90%	93%

제안 방법을 44.1KHz, 스테레오 음원을 이용하여 공모 공격에 대한 강인성을 평가하였다. 핑거프린팅 코드 삽입을 위하여 심리음향모델을 이용한 오디오 워터마킹 알고리즘을 이용하여 핑거프린팅 코드를 워터마크 정보로 삽입하였다[8]. 삽입된 컨텐츠에 대하여 평균공격과 모자익 공격, 그리고 공격한 파일의 MP3 압축된 음원에 대하여 핑거프린팅 코드의 검출 결과를 표2에서 나타내고 있다.

표2의 결과에서 알 수 있듯이 제안 방식은 기존의 핑거프린팅 공격에 대한 강인성은 매우 우수함을 알 수 있다. 또한 인터넷 상에서 취해질 수 있는 각종 신호처리적인 공격에서도 핑거프린팅 코드의 검출률은 매우 높은 것을 확인할 수 있다.

5. 결론

본 논문에서는 유한체 상의 원시 다항식을 이용하여 공모 공격에 강인한 핑거프린팅 코드 생성 방법을 제안하였다. 실험 결과에서 알 수 있듯이, 제안 방법은 평균공격 및 모자익 공격과 같은 기존의 공모 공격에 대하여 매우 강인한 것을 알 수 있고, 또한 기타 다양한 공격에서도 검출률이 뛰어남을 알 수 있다. 또한, 기존의 핑거프린팅 코드 구현에서 문제가 되는 사용자 확장 부분을 원시 다항식의 순환행렬에 의해 쉽게 확장 가능함을 보였다.

그러나, 본 논문에서 제안한 핑거프린팅 코드는 다수의 복사본에 대하여 2명의 공모 공격만 고려하고 있다. 따라서 향후 다수의 공모 공격에 대하여 강인성을 가지는 확장된 코드 구성법이 요구된다.

참고문헌

[1] M. Swanson, M. Kobayashi, A. Tewfik, "Multimedia Data Embedding and Watermarking Technologies," Proc. of IEEE, Vol. 86, No. 6, pp.1064-1087, June. 1998.

[2] B.Pfitzmann and M.Schunter, "Asymmetric Fingerprinting", EUROCRYPT'96, LNCS 1070, pp.84-95, 1996.

[3] D.Kirovski, H.S.Malvar, and Y.Yacobi, "Multimedia content Screening Using a Dual Watermrking and Fingerprinting System", ACM Mutimedia, 2002.

[4] D. Boneh and J. Shaw, "Collusion-Secure Fingerprinting for Digital Data", CRYPTO'95, LNCS 963, pp.542-465, 1995.

[5] J.Dittmann, A.Behr, M.Stabenau, P.Schmitt, J. Schwenk and J.Ueberberg, "Combining digital Watermarks and collusion secure Fingerprints for digital Image", SPIE J. Electron. Image, Vol. 9, pp.456-467, 2000.

[6] W.Trappe, M.Wu, J.Wang, and K.J.Ray Liu, "Anti-collusion Fingerprinting for Multimedia", IEEE Transction on Signal Processing, Vol. 51, N0. 4, pp.1069-1087, 2003.

[7] Robert J. MacEliecs, Finite Fields for Computer Scientists and Engineers, Kluwer, 1986.

[8] 이진홍, 박지환, "디지털 오디오 데이터의 저작권 보호를 위한 오디오 워터마크 기술", 한국정보보호학회 영남지부 학술발표회논문집, pp.35-42, 2002.