제18회 한국정보처리학회 추계학술발표대회 논문집 제9권 제2호 (2002.11)

핑거프린팅 된 이미지에 대한 공모공격에 관한 연구

김원겸, 이선화, 장호욱
한국전자통신연구원, S/W·컨텐츠기술연구부
e-mail:{wgkim, seonhwa, hwjang}@etri.re.kr

A Study on Collusion Attacks of Fingerprinting Images

Won Gyum Kim, Seon Hwa Lee and Ho-Wook Jang
S/W·Contents Technology Dept., ETRI

요 약

본 논문에서는 현재 컨텐츠에 대한 저작권보호 기술로 활발히 연구되고 있는 워터마킹 기술과 핑거프린팅 기술에 대해 알아보고 핑거프린팅 기술의 구현시 문제시 되고 있는 공모공격의 유형을 분석한다. 워터마킹 기술은 소유권주장이 목표인 반면 핑거프린팅 기술은 사용자인증을 목표로 하는 것이 다르다. 즉, 판매되는 컨텐츠에 구매자(사용자)의 정보를 워터마크로 삽입한 후 배포하고 후에 불법복제가 의심되는 컨텐츠에 대해 적법구매자의 워터마크를 추출하여 불법재분배를 증명한다. 이는 불법복제가 어느 구매자로부터 이루어졌는지의 추적이 가능하게 한다. 하지만 핑거프린팅 기술은 워터마킹 기술과는 다르게 구매자 정보를 워터마크로 사용하기 때문에 마크가 삽입된 컨텐츠마다 서로 다르다는 특징을 갖는다. 의도적인 공격자는 이 특성을 이용하여 여러 개의 핑거프린팅된 컨텐츠를 공모(collusion)하여 워터마크를 제거하기 위한 다양한 공격(attack)을 가할 수 있다. 평균화(averaging)공격과 모자익(mosaiking)공격이 공모공격의 대표적인 방법으로, 본 논문에서는 이러한 공격유형을 분석하고 공모공격에 강인한 핑거프린팅 삽입/추출방법을 고찰한다.

1. 서 론

인터넷을 통한 전자상거래가 활발해 짐에 따라 소프트웨어나 멀티미디어 데이터 같은 디지털컨텐츠(digital contents)에 대한 판매와 보급이 활발해 지고 있다. 하지만 이러한 컨텐츠는 디지털데이터의 특성상 완전복제와 다량배포가 가능하여 많은 불법사용자에 의한 재분배가 매우 쉽게 이루어지고 있어 이에 대한 보호가 절실한 실정이다.

일반적으로 소프트웨어 같은 컨텐츠에 대한 저작권보호(copyright protection)는 주로 암호메카니즘[2][3][4][5][6]을 사용하여 이루어지고 있다. 판매자는 적법 구매자에게 암호화 및 압축된 패키지(package) 형태의 컨텐츠와 이를 해독할 수 있는 키(key)를 함께 배포한다. 키는 구매자에 따라 다르며 그 구매자만 소유할 수 있다. 구매자는 자신의 키로 컨테츠를 해독하여 사용한다. 후에 불법복제가 의심되는 컨텐츠에 대해서 해독이 가능한 키를 역추적함으로써 어느 구매자로부터 복제되었는가를 찾아 내어 법적조치 같은 후속 조치에 활용할 수 있다. 이는 불법복제에 의한 배포를 어느 정도 예방하는 효과를 가지기는 하지만 근본적인 복사를 방지하지는 못한다는 점과 패키지된 상태가 아닌 해독된 상태의 컨텐츠가 불법 유통된다면 추적이 불가능한 약점이 있다.

이미지나 오디오, 비디오데이터 같은 멀티미디어 컨텐츠에 대한 저작권보호 방법으로 디지털 워터마킹(digital watermarking) 기술이 최근 몇 년전부터 활발히 연구되오고 있다. 워터마킹 기술은 판매 될 컨텐츠 자체에 저작권자의 마크를 삽입하여 판매하고, 후에 이 마크를 추출함으로써 저작권자를 증명하는 저작권보호 메카니즘이다. 컨텐츠 자체에 마크가 삽입되기 때문에 불법사용자가 컨텐츠를 해독하여 배포한다고 해도 여전히 마크를 검출할 수 있다는 점이 암호메카니즘을 이용한 방법과 다르다. 삽입된

워터마크는 의도적인 공격자가 마크가 삽입된 컨텐츠를 복사하여 재판매를 시도할 경우 저작권자를 분명하게 함으로써 불법 재판매에 대한 증거로써 활용될 수 있다. 따라서 워터마크는 불법 재판매자를 법적으로 대응할 수 있도록 하는 증거이며 또한 불법 재판매에 대해 어느 정도의 예방 효과를 가진다.

저작권보호를 위한 또 다른 기술로, 소유권자의 정보를 삽입하는 워터마킹 기술과는 다르게 컨텐츠의 판매시 구매자의 정보를 마크로 삽입함으로써 저작권을 보호하려는 핑거프린팅(fingerprinting) 기술이 있다[7]. 이는 워터마킹 기술의 확장으로 인터넷 상에서 불법으로 유통되고 있는 디지털 컨텐츠를 발견하였을 때, 디지털 워터마킹 기법을 사용한 컨텐츠라면, 그 컨텐츠의 원래 저작권자가 누구인지는 알 수 있지만, 누가 합법적으로 구매한 후에 불법으로 배포하였는지는 알아낼 수 없다. 핑거프린팅은 이를 보완하는 기술로 구매자의 정보를 워터마크로 삽입하고 추출함으로써 불법복제로 의심되는 컨텐츠가 어느 구매자로부터 불법유통되었는지를 역추적(traitor tracing)할 수 있도록 해 준다. 즉, 워터마킹 기술은 저작권자에 대한 소유권인증(owners-hip authentication)이 목표라면 핑거프린팅 기술은 구매자에 대한 사용자인증(user authentication)이 목표라 할 수 있다.

본 논문에서는 대표적인 저작권보호 기술인 워터마킹 기술과 핑거프린팅 기술의 차이점을 논한다. 또한 이미지 데이터에 핑거프린팅 기술을 구현함에 있어서 해결해야 할 조건중에 하나인 공모공격(collusion attack)에 대한 유형을 분석하고 이에 대한 보완책을 제시한다.

2. 워터마킹과 핑거프린팅

이미지 데이터에 대한 워터마킹 기술은 워터마크의 삽입영역에 따라 크게 공간영역(spatial domain)삽입방식과 주파수영역(frequency domain)삽입방식 두 가지로 분류된다. 또한 추출시 원본의 필요 유무에 따라 넌블라인드(non-blind)방식과 블라인드(blind)방식으로도 분류된다.

공간영역 삽입방식은 이미지데이터의 RGB, LC_rC_b 혹은 YUV영역에 직접 워터마크를 삽입하는 방식이며 주파수영역삽입방식은 이미지를 FFT, DCT, DWT등의 변환을 사용하여 주파수영역으로 변환한 뒤 적절한 주파수밴드를 선택하여 워터마크를 삽입한 후 다시 역변환하는 방식이다. 공간영역 삽입 방식은 주로 RGB의 블루(blue)영역이나 휘도(luminance) 영역에 워터마크를 삽입한다. 그 이유는 사람의 시각모델이 이 채널의 변화에 덜 민감하기 때문이다. 이 방식은 삽입방식이 간단하긴 하지만 여러 가지 공격(attack)이나 압축같은 신호처리에 취약한 단점이 있다.

주파수영역 삽입방식은 워터마크를 주파수변환을 거친 계수에 삽입하는 방식으로 연구 초창기에는 DCT나 FFT를 이용한 방법이 많이 연구되었다가 현재는 DWT를 이용한 방법이 주로 연구되고 있다. 이 방법은 워터마크가 주파수영역의 어떤 부분에 삽입되는 지를 결정하는 것이 중요하다. 저대역(low frequency band)부분에 워터마크가 삽입 될 경우에는 압축이나 신호처리에는 강인하지만 비지각성(imperceptibility)의 품질을 저하시키는 반면에 고대역(high frequency band)부분에 마크가 삽입될 경우에는 비지각성의 품질은 높은 반면 강인성이 저하되는 성질이 있다. 따라서 두가지 특성을 고려하여 적당한 삽입대역을 정하는 것이 중요하다. 또한 비지각성의 품질을 높이기 위해서 시각심리모델(psychovisual model)을 이용하여 워터마크의 삽입강도를 조절한다. 이는 이미지상에서 사람의 시각모델이 상대적으로 변화에 민감한 부분과 덜 민감한 부분(예를 들면, 에지(edge))을 계산하여 민감한 부분에는 작은 강도로 삽입하고 덜 민감한 부분에는 큰강도로 삽입함으로써 견고성도 강인하게 하고 비지각성의 품질도 높이는 삽입 방법이다.

핑거프린팅 기술은 이러한 워터마킹 기술의 확장이라 할 수 있다. 워터마킹 기술의 조건인 견고성과 비지각성이 핑거프린팅 기술에도 포함된다. 가장 큰 차이는 각 구매자의 정보가 워터마크로 삽입되기 때문에 워터마킹 된 컨텐츠가 서로 다르다는 점이다. 공격자는 컨텐츠의 이런 특징을 이용하여 워터마크를 제거하기 위해 여러 가지 다양한 공격을 가할 수 있다. 핑거프린팅 기술에서 구매자(사용자) 인증이라는 목적을 달성하기 위해 워터마킹 기술과는 달리 추가적으로 만족해야 할 조건들을 살펴보면 다음과 같다.

- 견고성 - 공모허용(collusion tolerance)

워터마킹 기술이 갖추어야 하는 견고성 이외에 공격자가 다수의 핑거프린팅 된 컨텐츠에 접근이 가능하더라도 컨텐츠를 서로 비교하여 삽입된 워터마크를 찾거나 생성, 혹은 제거할 수 없어야 한다는 조건이다.

- 비대칭성(asymmetry)

핑거프린팅된 컨텐츠는 구매자만이 접근할 수 있고 판매자는 접근이 불가능해야 한다. 반대의 개념인 대칭성 핑거프린팅 방법에서는 핑거프린팅된 컨텐츠를 판매자 또한 생성할 수 있기 때문에 구매자에 대한 불법재배포를 증명하는 완벽한 증거가 될 수 없다.

- 익명성(anonymity)

인터넷을 이용한 전자상거래의 특성상 구매자는 익명으로 컨텐츠를 구입할 수 있어야 한다. 구매자에게는 익명성을 보장하면서도 후에 구매자를 식별할 수 있도록 하는 마크가 삽입되어야 한다. 비대칭성과 익명성 같은 성질 때문에 핑거프린팅 기술은 워터마킹 기술과 같이 단순히 마크를 삽입/추출하는 방법이 아닌 암호프로토콜을 이용한 메카니즘 형태로 개발되어야 한다.

본 논문에서는 핑거프린팅 기술을 워터마킹 기술의 확장이라 제한하고 위에서 제시한 공모공격의 유형에 대해서만 정의하고 분석한다. 강인한 핑거프린팅 기술을 설계하기 위해서는 압축이나 RST(rotation, scaling and transition)공격같은 기존의 워터마킹에 대한 공격법[1][2]에 견고해야 됨은 물론 공모공격에 대해서도 강인해야 한다. 다음장에서는 공모공격의 정의와 유형에 대해서 고찰한다.

3. 공모공격(collusion attack)

공모공격은 공격자가 핑거프린팅 된 다수 이미지데이터에 접근 가능하고 그 이미지가 삽입된 핑거프린팅 데이터에 의해 서로 조금씩 다르다는 특징을 이용한 공격법으로 평균화(averaging)공격, 모자익(mosaiking)공격이 있다.

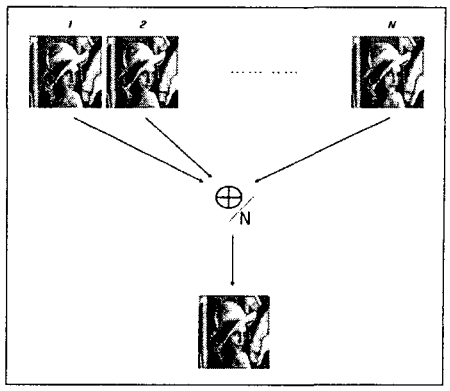
3.1 평균화 공격

평균화 공격은 핑거프린팅된 다수의 이미지의 픽셀(pixel)값을 서로 평균하여 새로운 이미지를 생성하는 핑거프린팅 제거 공격법이다. 워터마킹된 이미지의 경우, 일반적으로 저작권을 소유하는 소유주가 한명이기 때문에 삽입되는 워터마크와 비밀키가 동일하다. 하지만 워터마크를 삽입하는 소유주가 삽입되는 워터마크를 이미지마다 서로 다르게 하거나 혹은 비밀키를 서로 다르게 사용하게 되면 삽입되는 워터마크가 달라 워터마킹된 이미지가 서로 다르게 된다. 이런 경우에는 핑거프린팅 이미지와 같이 평균화공격에 의해 삽입된 워터마크가 손상될 수 있다.

[그림 1]은 핑거프린팅 이미지에 대해 평균화공격을 가하는 과정을 보인것이다. N개의 핑거프린팅 이미지가 있다고 가정할 때 공격자는 N개의 이미지에 대해 평균을 가하여 새로운 이미지를 생성한다.

$$x_i' = x_i + w_i = x_i + \alpha p_i c_i \quad \text{--- (1)}$$

수식(1)은 삽입할 워터마크를 생성하는 일반적인 수식이다. x_i는 삽입할 도메인의 계수이고 a는 삽입강도, p_i는 비밀키에 의해 생성된 무작위비트열, c_i는 구매자 정보, 따라서 w_i는 삽입할 핑거프린팅 신호를 나타낸다. 여기에 N개의 평균화 공격을 가하면 수식(2)와 같다.



[그림 1] 핑거프린팅 이미지에 대한 평균화공격

수식(2)에서 알수 있듯이 일반적으로 삽입할 핑거프린팅 데이터(w_i)를 {+1, -1}의 무작위비트열을 사용하고 N이 충분히 크다고 가정할 때 p_i도 무작위 순열이므로 ap_iw_i 값이 0에 근사하게 되어 삽입한 데이터는 거의 손실된다

$$\frac{\sum_{i=1}^{N} x_i'}{N} = \frac{\sum_{i=1}^{N} (x_i + \alpha p_i c_i)}{N} = \frac{\sum_{i=1}^{N} x_i}{N} + \frac{\sum_{i=1}^{N} \alpha p_i c_i}{N} = x_i + \Delta \ (\approx 0) \quad \text{--- (2)}$$

평균화 공격에 강인하게 핑거프린팅을 하기 위해서는 이미지 내에서 삽입되는 위치를 서로 다르게 삽입하거나 원본정보를 최대한 유지하면서 삽입하는 방법이 필요하다. 하지만 삽입위치를 다르게 하여 삽입하는 방법은 구매자의 수가 아주 많을 경우에는 적용이 어렵고 삽입되는 정보의 양도 제한적이다. 따라서 기밀데이터의 배포 같이 배포자가 제한적인 경우에만 적용가능하다.

원본정보를 최대한 유지하면서 핑거프린팅하는 방식은 마크의 삽입시 추출 조건을 만족하면 삽입을 하지 않는 방식으로 원본의 값을 최대한 유지시켜 평균화공격의 영향을 덜 받도록 한다. 수식(3)은 삽입조건을 나타낸 것이다.

$$w_i = \begin{cases} 0 & \text{if } f(x_i, x_i') = c_i \\ \alpha p_i c_i & otherwise \end{cases} \quad \text{--- (3)}$$

즉, 워터마크를 삽입할 계수가 삽입할 마크와 같은 추출조건을 이미 가지고 있으면 워터마크를 삽입하지 않음으로써 원본정보를 최대한 유지하고 그렇지 않을 경우에만 마크를 삽입한다.

또한 공모에 참여한 공격자를 알아내는 방법으로 Dittmann[7]의 방법이 있다. Dittmann은 공모자의 수가 제한되어 있을때 모든 공모자들을 색출할 수 있는 유한사

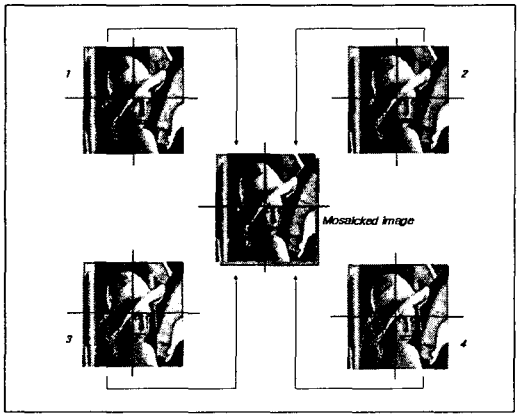
영기하학을 기반으로 하는 d-dectecting 핑거프린팅 알고리즘을 제안하였다. 여기서 d는 공모자의 수이다. 예를 들어 3명의 구매자(q)가 있고 2명(d)의 제한적 공모만을 허용할 때 구매자 3명에 대한 13비트(n)의 핑거프린팅 코드는 다음과 같다.

v_1={0 0 0 1 0 0 **1** 0 1 0 **1** 0 0}
v_2={0 0 1 0 0 0 **1** **1** 0 0 0 1 0}
v_3={0 1 0 0 1 0 0 **1** 0 0 **1** 0 0}

비트 1의 위치에 워터마크가 삽입되었고 3명의 구매자 중 어느 2명이 공모했다고 가정한다면 3명의 코드중에서 공통되는 1의 값을 갖는 1비트가 반드시 존재하므로 추출시에 그 유사도를 가지고 공모자를 찾아내는 방법이다.

3.2 모자익 공격

모자익 공격은 다수의 핑거프린팅 이미지를 작은 조각으로 나누어 각각의 조각을 모자이크 하듯이 끼워 맞추어 새로운 이미지를 생성하는 공격법으로 평균화 공격과 마찬가지로 핑거프린팅 제거 공격법이다. 핑거프린팅 이미지에 대한 모자익 공격은 워터마킹 이미지에 대한 모자익 공격과는 다르다. 워터마킹 이미지에 대한 모자익 공격은 웹을 통해 로봇같은 검색엔진을 이용하여 워터마크를 검색할 경우 워터마크된 이미지를 작은 조각으로 잘라 보관함으로써 이런 방식의 워터마크 검색을 피하는 방식이다. 따라서 워터마크된 이미지 자체를 작은 조각으로 나누어 보관하는 것이 핑거프린팅 이미지에 대한 모자익 공격과 다른점이다.



[그림 2] 핑거프린팅 이미지에 대한 모자익 공격

[그림 2]는 4명의 공모자가 모자익 공격으로 새로운 이미지를 생성하는 과정을 보인것이다. 4개의 이미지에서 각각 1/4 이미지 조각으로 새로운 이미지가 생성된다. 이 공격은 기존의 워터마킹 기법에 대한 공격법중 잘림(cropping)공격과 같은 부류로 원이미지의 1/4에서 워터마크가 검출되지 않는다면 4개의 핑거프린팅 데이터는 모두 손실된다. 모자익 공격에 강인하게 하기 위해서는 가능한한 이미지의 작은 영역에 핑거프린팅 데이터를 삽입해야 한다.

4. 결 론

본 논문에서는 이미지에 대한 핑거프린팅 기법을 워터마킹 기법과 비교, 설명하였다. 핑거프린팅 기법은 워터마킹의 확장기술로 워터마킹과는 달리 구매자 정보를 워터마크로 삽입하기 때문에 같은 이미지라도 핑거프린팅 이미지는 서로 조금씩 다른 특징을 갖는다. 이는 핑거프린팅 데이터를 없애는 여러 가지 공격을 가능하게 하는데 특히 구매자 여러명이 공모하여 핑거프린팅 데이터를 지우는 공모공격이 대표적으로 평균화공격과 모자익공격이 있다. 평균화 공격은 핑거프린팅된 다수의 이미지의 픽셀값을 서로 평균하여 새로운 이미지를 생성하는 핑거프린팅 제거 공격법으로 이에 강인하게 워터마크를 삽입할려면 최대한 원본 정보를 유지하면서 마크를 삽입해야 한다. 모자익 공격은 다수의 핑거프린팅 이미지를 작은 조각으로 나누어 각각의 조각을 모자이크 하듯이 끼워 맞추어 새로운 이미지를 생성하는 공격법으로 이에 강인하게 하기 위해서는 가능한한 이미지의 작은 영역에 핑거프린팅 데이터를 삽입하는 기술이 필요하다.

참 고 문 헌

[1] Fabien A. P. Petitcolas, Ross J. Anderson, Markus G. Kuhn, "Attacks on copyright marking system," in *Proceedings of Information Hiding, Second International Workshop, IH'98*, Portland, Oregon, U.S.A., April 15-17, 1998

[2] D. Boheh and J. Shaw, "Collusion-Secure Fingerprinting for Digital Data," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1897-1905, Sept. 1998.

[3] B. Pfitzmann, and M. Schunter, "Asymmetric Fingerprinting," *Journal of the ACM*, Vol.33, pp.792-807, 1986

[4] B. Pfitzmann, and M. Waidner, "Anonymous Fingerprinting," in Advances in Cryptology, *Proc. of EUROCRYPT '97*, Vol.1233, of Lecture Notes in Computer Science, Springer-Verlag, pp.88-102, 1997

[5] B. Pfitzmann, and A. Sadeghi, "Coin-Based Anonymous Fingerprinting," in Advances in Cryptology, *Proc. of EUROCRYPT '99*, Vol.1592, of Lecture Notes in Computer Science, Springer-Verlag, pp.150-164, 1999

[6] J. Domingo-Ferrer, "Anonymous Fingerprinting of Electronic Information with Automatic Identification of Redistributors," *Electornics Letters* 34/13, pp.1303-1304, 1998

[7] J. Dittmann, "Combining digital watermarks and collusion secure fingerprinting for customer copy monitoring," *Proc. IEE Seminar Sec. Image & Image Auth.*, pp.128-132, Mar., 2000