

PDF 형식 북스캔 파일 불법 유통 방지를 위한 관리 프레임워크*

이 국 현,[†] 정 현 지, 유 대 걸, 이 상 진[‡]
고려대학교 정보보호연구원 디지털포렌식 연구센터

A framework of management for preventing illegal distribution of pdf bookscan file*

Kuk-heon Lee,[†] Hyun-ji Chung, Dae-gull Ryu, Sang-Jin Lee[‡]
Digital Forensic Research Center, Korea University

요 약

다양한 스마트 기기의 보급으로 인해 종이책이 아닌 전자책(e-book)으로 책을 읽는 사용자가 증가하고 있다. 하지만 전자책으로 출판되는 책의 다양성이 부족하여 사용자가 직접 원하는 책을 스캔하여 전자책으로 만드는 북스캔(Bookscan)이 생겨났다. 현재 북스캔 서비스는 PDF(Portable Document Format), JPG파일 형태로 제공되고 있으며 파일 관리에 대한 문제점으로 인해 PDF 파일 형식이 보편적으로 사용되고 있다. 북 스캐닝은 고가의 장비와 시간 탓에 북스캔 대행업체가 시장을 형성해가고 있다. 하지만 이는 엄연히 저작권법에 위배되는 불법 행위이다. 또한 북스캔 대행업체는 저작물에 대한 보호조치가 없어 북스캔본이 웹을 통해 불법 공유되고 있다. 북스캔 서비스의 저작권 문제로 출판 산업 시장도 과거 소리바다 음원 시장 붕괴와 같은 절차를 밟을 것으로 보인다. 따라서 북스캔에 대해 법제도 정비를 위한 기술적인 뒷받침이 필요하다. 기존 불법 저작물 추적관리 시스템은 음원, 영화에 대해 적용이 되고 출판물에 대해서는 조치가 없었다. 이에 대해 본 논문에서는 실현 가능한 기술을 기반으로 PDF 형식의 북스캔 파일에 대한 관리 프레임 워크를 제시하고자 한다.

ABSTRACT

Since various smart devices are being developed, a growing number of people are reading eBooks instead of paper books. However, people started making eBooks on their own by scanning paper books because there are not enough eBooks provided from market. The term "Bookscan" was made with this reason. The number of bookscan company is increasing because the equipment is too expensive. However, the commercial activity of bookscan company is against copyright law. Also bookscan files are in danger of being illegally distributed on web, because bookscan companies are not protecting copyright. Publication market follows the same procedure with sound market which was collapsed due to copyright problem. Therefore, the technical methods should be prepared for law system against bookscan. The previous ICOP(Illegal Copyrights Obstruction Program) system has been applied to sound and movie files, but not applied to publication. This paper suggests the framework for bookscan file management based on practical mechanism.

Keywords: Bookscan service, Framework of bookscan file, management PDF, XMP, Data Hiding

접수일(2013년 5월 2일), 수정일(2013년 6월 25일), 게재
확정일(2013년 7월 10일)

* 본 연구는 산업통상자원부 및 한국산업기술평가관리원의
산업융합원천기술개발사업(정보통신)의 일환으로 수행하

였음.(10035157, 실시간 분석을 위한 디지털 포렌식 기
술 개발)

[†] 주저자, kukheon1109@korea.ac.kr

[‡] 교신저자, sangjinlee@korea.ac.kr(Corresponding author)

I. 서 론

스마트폰의 대중화와 전자책 콘텐츠 증가로 전자책 시장이 활성화 되었다[1]. 하지만 전자책의 공급이 사용자의 수요를 맞추지 못하자 사용자가 직접 책을 스캔하여 전자책으로 만드는 '복스캔' 서비스가 생겨났다[2]. 복스캔 파일은 PDF 또는 JPG 파일로 제공 되는데, 관리에 있어서 수월한 PDF 파일 형식이 보편적으로 사용 되고 있다. 복스캔 과정에 필요한 장비들은 개인이 구매하기에 비싸기 때문에, 복스캔 장비를 이용하여 복스캐닝을 대신해주는 복스캔 대행업체가 생겨났다.

복스캔 시장이 형성되어 가고 있음에도 불구하고 현재 복스캔 대행 서비스는 저작권법에 위배되는 사업이다[3]. 저작권법 제30조에 따라 저작물 소유권자가 개인의 편익을 위해 직접 스캔을 행하는 주체가 될 경우에는 저작권 침해가 아니지만, 제 3자인 복스캔 대행 서비스 업체가 스캔을 행하는 주체가 될 경우 위법 행위가 된다.

온라인상의 시장 증가는 기존 오프라인 세계를 기반으로 형성되어 온 법제도와 마찰을 빚어왔다. 예를 들어 과거 소리바다는 기존 CD, TAPE 등 물리적인 매체로 형성된 오프라인 음원 시장과 마찰을 빚었었다. MP3 파일에 대해 제도적인 기반이 없고 불법이었음에도 불구하고 소리바다는 큰 시장을 형성하게 되었고 결국 법적 소송까지 직면하게 되었다.

하지만 소리바다 사건은 이용자가 600만 명을 넘어서는 하나의 음악시장으로 인해 서비스 중지가 아닌 단순히 저작권 보호조치 판결이 나왔다[4].

소리바다 판결의 예에서 알 수 있듯이 법에 위배되는 사항을 단순히 불법으로 판단하고 해당 서비스를 중지시킬 수 없다. 과거 디지털 음원 시장과 동일하게 복스캔도 복사가 용이한 전자파일로써 토렌트(torrent), P2P(Peer-to-Peer network), 웹하드를 통해 불법 유포될 가능성이 존재한다. 시장이 활성화 되기 전에 대책을 수립하지 않는다면 과거 소리바다 음원 시장과 똑같은 절차를 밟을 것으로 보인다. 따라서 본 논문에서는 PDF 형식의 복스캔에 대한 기술적인 대책을 제안하고자 한다.

II. 복스캔 서비스의 정의 및 문제점

2.1 복스캔 서비스의 정의

복스캔이란 인쇄물 형태의 종이책을 스캔장비를 이

용하여 jpg 파일 또는 PDF 파일과 같은 전자책으로 변환하는 작업을 의미한다. 스마트폰과 태블릿PC와 같은 스마트 기기의 사용자가 늘어남에 따라 전자책의 필요성이 증가하였다[5]. 하지만 출판업계 입장에서는 각기 다른 전자책 단말기를 지원해야 하는 번거로움과 출판 구조문제로 인하여 종이책에 비해 전자책의 출간 비율이 매우 낮다[6]. 이로 인해 출판업체가 아닌 저작물 소유권자가 직접 스캐너 장비를 이용하여 종이책을 전자문서로 변환하기 시작하였다. 개인이 구매하기에 비싼 복스캔 장비들로 인해 '복스캔 대행 서비스'라는 새로운 형태의 서비스가 생겨났다. 복스캔 대행업체는 금품을 대가로 소비자가 위탁한 종이책을 스캔하여 JPG 파일 또는 PDF 파일과 같은 전자책 형태로 제공한다. 보통 복스캔 대행업체는 대량의 JPG 파일을 관리하는 것이 어려워 통합된 PDF 파일로 제공한다.

복스캔본은 파일 특성상 플랫폼에 제한받지 않으며, 다양한 스마트 기기에서 열람이 가능하며 또한 사용자가 원하는 책을 전자책 형태로 소유할 수 있어 사용률이 점차 증가할 것으로 예측된다.

2.2 복스캔 시장

태블릿 PC 및 스마트폰의 보급으로 인해 전자책의 수요가 증가하고 있다. 하지만 아직도 전자책 콘텐츠 공급은 수요에 비해 부족한 현실이다. 한국콘텐츠진흥원의 보고서에 의하면, 현재 세계 출판시장은 2011년 기준 1,120억 달러로서 종이책 시장 1,065억 달러(95.1%)와 전자책 시장 55억 달러(4.9%)로 시장이 형성되어 있다. 향후 5년간 세계 연평균 0.6%씩 성장하여 2016년 1,157억 달러로 현 시장을 유지할 것으로 전망하고 있다. 종이책 시장은 2016년 까지 연평균 2.3%씩 감소해 949억 달러 규모가 되는 반면, 전자책 시장은 2016년까지 연평균 30.3%씩 성장해 208억 달러(16%)에 이를 것으로 전망하고 있다. 전자책의 공급으로 인한 복스캔 시장 형성은 국내뿐만 아니라 미국, 영국, 일본 등의 전 세계적인 문제로 부상하고 있다[7].

2010년 4월 일본의 '복스캔(Bookscan.co.jp)' 업체가 처음 복스캔 대행 서비스를 시작하였다. 해당 업체는 저렴한 스캐닝 가격을 통해 인지도를 높였으며, 국내의 복스캔 업체와 다르게 ISBN(International Standard Book Number)을 기준으로 저작권자, 출판업체, 소비자에게 정보를 제공하고 2차 출판에 대

한 허가를 가능하게 북스캔 저작권 통제 센터(Book-scan Right Control Center)를 통해 자체적인 저작권 관리 시스템을 형성하였다.

국내에서는 2011년 5월 아이패드 판매를 기점으로 북스캔 대행 서비스 업체가 생겨나기 시작하였고, 대형 포털 사이트 기준 10여개의 업체가 등록되어 있다. 개인이 북스캔 대행 서비스를 제공하는 1인사업자 형태의 서비스도 존재하기 때문에 업체의 수를 정확하게 추정하기는 어렵다. 국내 북스캔 대행 서비스 업체는 저작권 보호를 위해 소극적인 불법 복제 방지 조치를 취하고 있다. '악어스캔'은 북스캔본에 DRM(Digital Rights Management) 기술을 적용하여 파일의 고유값을 관리하고 있다. 혼민스캔은 안드로이드 기반의 어플리케이션을 통하여 북스캔본을 관리하고 있다. 또한 저작권과 관련한 모든 책임을 북스캔본 구매자에게 위임하는 업체도 있다.

북스캔 대행 서비스라는 시장이 형성되면서 저작권과 관련한 불법복제 및 저작권 침해에 대한 문제는 피할 수 없을 것이다. 이로 인해 기존 전자책 시장은 물론 저작권 침해로 인한 창작자의 창작 의욕이 저해되어 출판업계의 전반적인 침체가 우려된다. 이는 과거 온라인 시장에서의 음악 및 영화 산업 붕괴와 유사한 전철을 밟을 가능성이 있다. 북스캔에 대해 불법적인 서비스를 제공하는 사업자에 대해 단속하는 것이 맞지만, 합법적인 서비스를 지향하는 사업자에 대해 상응하는 조치가 필요하다.

2.3 북스캔 서비스의 문제점

2.3.1 법제도적 측면

급속도로 변화하는 사회 현실을 즉각적으로 법제도에 반영하기 어렵다. 2011년 5월 문화체육관광부가 발표한 자료에 따르면 북스캔 대행 서비스는 기존 법에 위배된다. 최근에는 새로운 서비스에 맞추어 법제도 개정방안이 필요하다는 목소리가 나오고 있다 [8][9]. 북 스캔 대행 서비스가 불법이라는 내용의 조항은 다음과 같다.

저작권법[10] 제4조(저작물의 예시 등)에 의하면 책의 그림 및 글은 어문저작물, 사진저작물에 해당한다고 볼 수 있다. 북스캔본의 불법복제행위 및 불특정 다수에게 배포는 저작물 저작자의 저작재산권을 침해하는 행위로서 그 손해배상책임이 인정된다.

저작권법 제30조(사적이용을 위한 복제) '공표된

저작물을 영리를 목적으로 하지 아니하고 개인적으로 이용하거나 가정 및 이에 준하는 한정된 범위 안에서 이용하는 경우에는 그 이용자는 이를 복제할 수 있다. 다만, 공중의 사용에 제공하기 위하여 설치된 복사기기에 의한 복제는 그러하지 아니하다.' 조항을 살펴보면 '개인적으로 이용하거나 가정 및 이에 준하는 한정된 범위' 라는 것은 책을 서점 또는 인터넷으로 구매한 저작물 소유권자가 개인의 편익을 위해 직접 스캔하는 주체가 되는 경우에만 저작권 침해가 아니라는 말이다.

저작권법 제35조의2(저작물 이용과정에서의 일시적 복제) '컴퓨터에서 저작물을 이용하는 경우에는 원활하고 효율적인 정보처리를 위하여 필요하다고 인정되는 범위 안에서 그 저작물을 그 컴퓨터에 일시적으로 복제할 수 있다. 다만, 그 저작물의 이용이 저작권을 침해하는 경우에는 그러하지 아니하다.' 조항에서 개인의 편익을 위해 스캔한 경우 스캔을 한 주체가 되는 대상 PC에서 파일을 일시적으로 저장 및 사용하여야 한다. 하지만 북스캔은 스캔을 진행한 PC에서 스마트 기기로의 전송 및 복제가 이루어지기 때문에 위법이다.

과거 소리바다, 대학가 출판물 불법복제처럼, 시장 출현 초반에 제도를 확립하지 않는다면 법적 기준을 교묘히 빠져나가는 불법 업체가 성행할 것이다. 또한 불법 행위로 인한 출판, 저작권 관련 시장에 타격을 줄 것으로 보인다.

2.3.2 불법복제 문제

북스캔은 인터넷의 발달과 함께 익명성 및 전파성이라는 사이버공간의 특성으로 인하여 저작권침해자의 확정이 곤란해졌다는 점이다. 즉, 저작권자가 사이버공간에서의 침해자를 찾아내어 일일이 손해배상청구권을 행사하거나 형사 고소한다는 것은 거의 불가능에 가까운 것이 현실이다[11]. 또한 불법 복제물을 단속해 폐기하는 것이 어렵기 때문에 아직까지도 큰 문제점으로 남아 있다. 이를 지각하고 있는 출판사들은 자사의 베스트셀러 등의 전자책 출간을 피하고 있다. 불법복제에 대한 문제 때문에 전자책의 공급이 줄어들게 되고 수요를 해결하기 위해 북스캔 대행 서비스 시장이 급속도로 성장하고 있다.

현재 국내에서 디지털 콘텐츠의 저작권 보호를 위해 사용하는 기술 중 하나가 DRM 기술이다. DRM은 출판자 또는 저작권자가 그들이 배포한 디지털 자

료나 하드웨어의 사용을 제어하고 이를 불법적인 용도로 사용하는 것을 제한하는 데 사용되는 모든 기술을 지칭하는 용어다[12]. 하지만 국내 DRM 기술은 사용자가 구매한 콘텐츠를 사용할 수 있는 디바이스에 대해 제한적이고 별도의 추가 프로그램이 필요하기 때문에 외면 받는 것이 현실이다.

DRM 기술이 저작권 보호를 위한 만능 조치는 아니다. DRM을 설정할 경우 디지털 기기의 호환성은 물론 사용자들의 불편이 증가하여 오히려 콘텐츠 이용률이 떨어진다는 의견도 있다[13]. 또한 DRM 기술은 사전 조치일 뿐이지 문제가 발생한 후, 즉 사후 처리에 있어서는 유포와 관련된 정보를 알 수 없다. 이러한 DRM 기술을 해제하여 사용자의 편의성을 증가시켜 콘텐츠 시장의 활성화를 이끌어 내야 한다. 애플사를 살펴 볼 경우 애플사가 주도한 DRM-Free 정책을 통하여 디지털 기기의 호환성을 높이게 되었고, 이에 따라 미국 내 디지털 음반 시장의 성장은 물론 사용자들의 인식까지 변화하게 되었다[14].

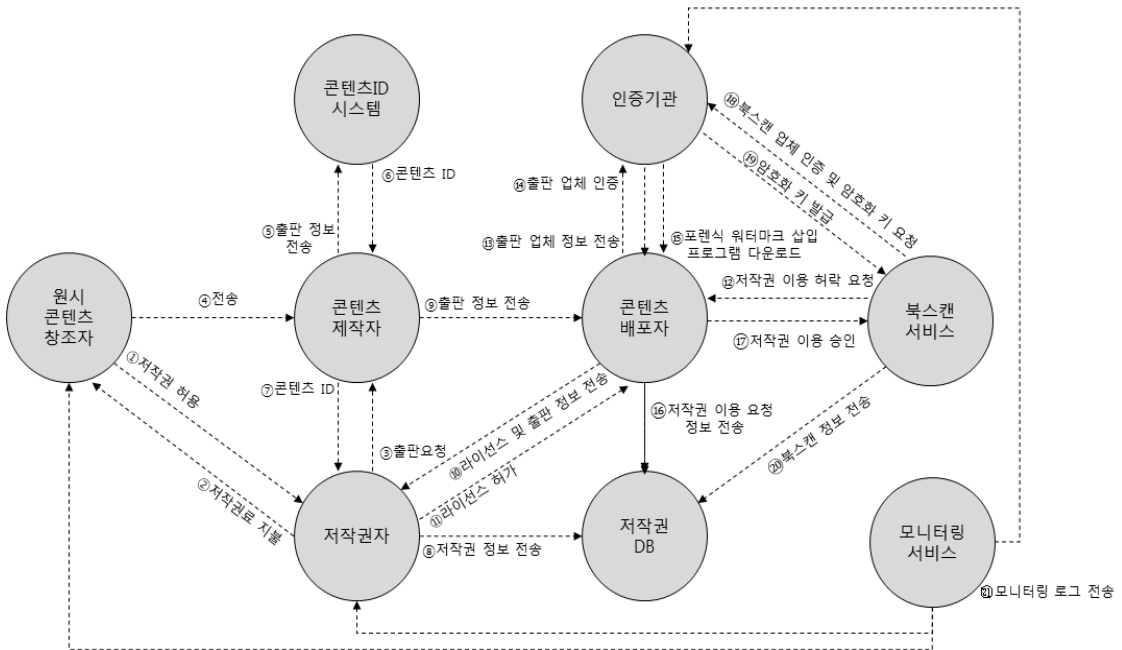
본 논문에서는 프로그램에 비종속적인 불법유통 방지 조치를 취하기 위해 3절 복스캔 관리 체계에서 기술적인 방안을 제안한다.

III. 복스캔 파일 관리 및 협력 체계

3.1 복스캔 관리 체계

복스캔 서비스에 대해 관련 정책과 법제도가 정비되기 위해서는 이를 뒷받침하기 위한 기술적인 보완 조치가 필요하다. 복스캔 파일에 대한 관리 및 불법 유포 방지를 위해 본 논문에서 제안하고자 하는 복스캔 파일 유통 관리 체계는 [그림 1]과 같다. 복스캔 파일 관리 체계는 IMPRIMATUR[15]을 참조하여 구성되었다. IMPRIMATUR은 유통 흐름에 따른 주체와 기능이 명확히 정의되어 있는 비즈니스 모델이다. 현재 디지털 콘텐츠 유통 모델로 사용되고 있다. 복스캔 파일 관리 체계는 저작물보호조치 및 복제 제한을 위하여 컴퓨터 기술, 워터마크 기술, 암호 기술, 정보 은닉 기술 등 실현 가능한 기술을 기반으로 한다.

복스캔 파일은 DRM과 다르게 플랫폼에 비종속적이며, 파일에 대한 개별 관리가 되어야 한다. 따라서 플랫폼 기반 기술이 아닌 PDF 문서에 대해 기술을 구현해야 한다. 기술적으로 간략하게 살펴본다면, 복스캔 파일이 어떠한 내용을 담고 있는지에 대한 저작물 식별 정보를 삽입할 수 있도록 워터마크 삽입 기술이 요구된다[16]. 워터마크를 삽입하는 이유는 비가시적



(그림 1) 복스캔 관리 체계

이고 기존 데이터에 위변조를 가하더라도 워터마크 영역을 확인하여 불법 유포자를 찾기 위한 방안이다.

복스캔 대행업체는 일반 사용자들이 사용자 식별 정보를 알아볼 수 없도록 난독화 과정이 필요하다. 난독화하기 위해 암호 기술 및 인코딩 기술이 요구되고, 정보를 삽입하기 위해 워터마크 삽입 기술, 정보 은닉 기술이 요구된다. 이때 사용하는 소프트웨어는 워터마크를 삽입할 뿐만 아니라 해당 식별 정보를 제 3의 신뢰기관의 저작물 관리 서버로 전송해야 하기 때문에 네트워크 기술이 요구된다.

제 3의 신뢰기관은 해당 소프트웨어가 신뢰할 만하다는 것을 검증해주기 위해 소프트웨어 검증 기술이 요구된다. 또한 제 3의 신뢰기관은 웹 크롤링 기술을 이용하여 복스캔본을 수집하여, 복스캔본 내에 삽입된 워터마크와 은닉 정보를 찾아내기 위해 워터마크 검출 기술, 암호 기술, 그리고 은닉 정보 탐지 기술이 요구된다.

출판 업체, 복스캔 대행업체, 그리고 인증기관은 위와 같은 기술을 사용할 수 있는 환경을 만들어야 한다.

3.2 협력 체계

본 논문에서 제안한 복스캔 파일 불법 유통 방지 체계가 의의를 갖고 복스캔 서비스와 관련한 문제를 해결하기 위해서는 크게 출판 업계, 복스캔 대행 업계, 제 3의 신뢰기관이 각자의 역할을 잘 수행할 수 있도록 협력해야 한다. 각 업체 및 기관의 역할은 다음 절과 같다.

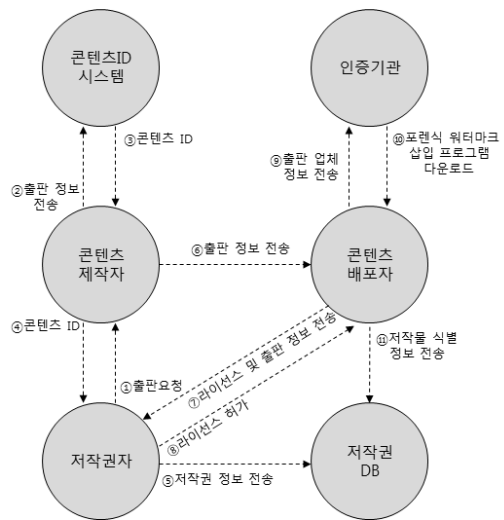
3.2.1 출판 업체

출판업체는 인쇄되는 종이책에 저작물 정보를 식별할 수 있는 포렌식 워터마크를 삽입하고 해당 정보를 저작물 서버에 전송하는 역할을 한다. 자세한 내용은 다음 절과 같다.

3.2.1.1 포렌식 워터마크 삽입

콘텐츠 제작 및 배포자는 시장 구조로 인해 각각 다른 업체이거나, 하나의 업체일 수 있다. 이 때 콘텐츠 제작 및 배포자는 [그림 1]에서 출판 업체에 해당한다. 출판 업체의 역할은 저작권자로부터 콘텐츠 출판 요청이 있을 경우 콘텐츠 ID 시스템으로 출판물 정보를 전송하고 콘텐츠 ID를 발급 받는다. 그 다음 인증

기관으로부터 출판 업체 인증 후 콘텐츠 정보 삽입이 가능한 포렌식 워터마크 삽입 프로그램을 다운로드 받아 콘텐츠에 삽입한다. 포렌식 워터마크란 저작물의 식별정보 및 저작권과 관련된 정보를 포함하고 있는 워터마크이다. 포렌식 워터마크는 인쇄되는 출판물에 같이 삽입된다. 차후 포렌식 워터마크의 정보를 확인하여 복스캔 업체에서 사용하거나, 2차 출판물 제작 과정에 사용된다. 배포될 콘텐츠에 대한 저작 정보는 저작권자 및 저작권 DB에 전송하여 등록한다. 출판 업체의 역할은 [그림 2]와 같다.



(그림 2) 출판 업체 역할

3.2.1.2 포렌식 워터마크 전송

출판업체가 인증기관으로부터 다운받은 프로그램은 포렌식 워터마크의 내용을 인증기관이 관리하는 저작권 DB 및 모니터링 서버로 전송하는 역할을 한다. 포렌식 워터마크를 서버에 별도로 관리하는 이유는 위변조에 대해 방지하고 불법 유포자 발견 시 불법 유포 파일과 원본의 내용을 비교하여 유포자에 대한 정보를 정확하게 확인하기 위해서다.

3.2.2 복스캔 대행 업체

불법 유포 방지를 위해 종이책을 스캔하여 복스캔 파일로 제작하는 복스캔 대행업체의 역할이 중요하다. 업체에서는 복스캔 파일에 대해 실질적인 저작물 보호 조치를 한다. 보호 조치에 대한 내용은 다음 절과 같다.

3.2.2.1 디지털 포렌식 워터마크 삽입

출판업체에서 삽입한 포렌식 워터마크를 통해 저작물 정보를 확인한다. 이 과정에서 포렌식 워터마크에 대한 유효성을 검사하며 만일 포렌식 워터마크에 포함된 저작물 정보와 실제 저작물 정보가 상이할 경우 콘텐츠 ID 시스템, 인증기관에 통보하여 저작권 DB에 등록된 내용이 수정 되도록 한다. 포렌식 워터마크가 없을 경우 인터넷상에서 별도로 조회하여 확인한다. 저작물 내용과 복스캔 서비스를 요청한 사용자 식별 정보를 텍스트와 이미지 형태의 워터마크로 삽입하는데 이를 디지털 포렌식 워터마크라 정의한다. 디지털 포렌식 워터마크는 포렌식 워터마크와는 다르게 저작물 보호조치 유무를 시각적으로 확인이 불가능하다. 사용자 식별 정보는 평문이 아닌 암호화와 인코딩 과정을 통해 암호문으로 삽입한다. 암호화 과정에서 사용하는 키는 제 3의 신뢰기관으로부터 발급 받아 사용한다. 디지털 포렌식 워터마크는 위변조를 방지하고자 여러 페이지에 중복 저장한다. 복스캔 대행업체의 역할은 [그림 3]과 같다.

텍스트 형태의 워터마크는 암호문을 의미하며, 복스캔 파일의 메타데이터 영역인 XMP(Extensible Metadata Platform)에 삽입된다. XMP 영역과 관련한 자세한 내용은 4절 PDF 파일을 이용한 식별 정보 은닉 기법에서 보다 자세하게 설명한다. 이미지 형태의 워터마크는 OCR(Optical Character Reader) 기술을 이용해 사용자 식별정보를 이미지

로 변환한 것을 의미한다. 텍스트와 이미지를 함께 사용하는 이유는 각각의 워터마크를 검출하여 위변조를 확인하고자 할 때 사용한다. 워터마크는 제거할 가능성이 존재하기 때문에 1차적인 저작물 보호 조치로서의 역할을 한다.

3.2.2.2 스테가노그래피(정보 은닉 기법) 적용

디지털 포렌식 워터마크가 삽입되는 PDF 문서의 메타데이터 영역을 확인하면 ‘-(hyphen)’으로 연결된 데이터들이 존재하는 걸 확인할 수 있다. 이를 이용하여 사용자 식별 정보들을 ‘-’으로 연결하여 기존 데이터들과 유사하게 함으로써 데이터 은닉이 가능하다.

3.2.2.3 수정암호 적용

복스캔 파일에 디지털 포렌식 워터마크를 삽입한 뒤 저작물 보호 조치를 위해 수정암호[17]를 설정한다.

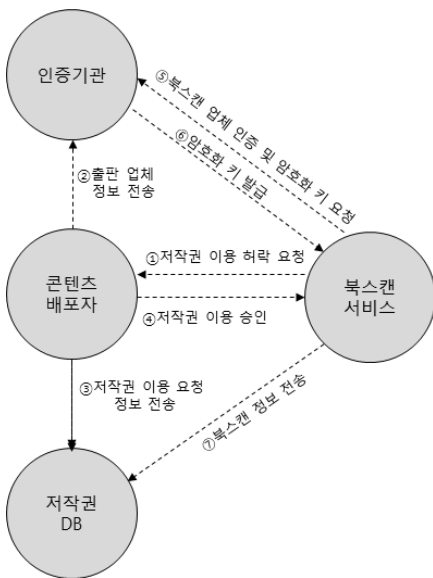
PDF 파일은 두 가지 종류의 암호 방식, 열기암호와 수정암호를 제공한다. 열기암호란 사용자가 복스캔 파일을 열람하기 위해서 비밀번호를 입력하도록 설정하는 암호이다. 열기암호를 설정함으로써 사용자는 저작물 보호 조치에 대해 인지함으로써 저작물 보호 조치를 없애려는 노력을 할 수 있다. 저작물 보호 조치가 없어질 가능성을 낮추기 위해서는 PDF가 제공하는 수정암호를 사용하는 것이 바람직하다. 수정암호는 PDF 내의 콘텐츠를 복사/인쇄 하기 전에는 암호가 설정되어있다는 것을 인지하기 어렵기 때문이다. 또한 수정암호를 통해 사용자가 복스캔 파일을 무단으로 수정하거나 복사하지 못하도록 할 수 있다. 이때 사용되는 수정암호는 제3의 신뢰기관이 발급한 키를 사용하여 복스캔 업체와 제3의 신뢰기관이 보관하도록 한다.

3.2.2.4 디지털 포렌식 워터마크 및 은닉 정보 전송

제 3의 신뢰기관으로부터 다운로드 받은 소프트웨어는 디지털 포렌식 워터마크 삽입 및 스테가노그래피 적용, 수정 및 열기 암호를 설정할 수 있다. 또한 암호화 및 인코딩된 사용자 식별 정보를 제 3의 신뢰기관이 관리하는 저작물 관리 서버로 전송하여 관리한다.

3.2.3 제 3의 신뢰기관

제 3의 신뢰기관은 한국저작권위원회, 대한출판문화협회, 한국전자출판협회, 한국복사전송권협회와 같은 기관이 역할을 할 수 있다. 제 3의 신뢰기관은 독립적인 기관으로써 외부의 압력에서 자유로우며, 저작



[그림 3] 복스캔 대행 업체

권과 관련한 정책을 수립할 수 있어야 한다.

3.2.3.1 프로그램 신뢰성 검증

제 3의 신뢰기관은 포렌식 워터마크 삽입 프로그램과 디지털 포렌식 워터마크 삽입 프로그램의 기능이 우회되지 않고 제대로 동작하는지에 대한 신뢰성을 검증해야 한다. 삽입 프로그램은 실질적인 저작물 보호 조치를 할 수 있는 도구이기 때문에 프로그램에 대한 신뢰성 검증은 필수적이다. 프로그램의 규격화된 분석을 통하여 검증이 이루어져야 한다[18].

3.2.3.2 암호화 키 발급 및 관리

제 3의 신뢰기관은 복스캔 대행업체가 사용자 식별 정보를 암호화할 때 필요한 키를 발급하고 관리해야 한다. 키에 대한 취약성이 발견되지 않도록 내부적인 관리가 필요하다.

3.2.3.3 포렌식 마크 및 디지털 포렌식 마크 정보 저장

출판 업체가 전송한 포렌식 워터마크와, 복스캔 대행업체가 전송한 디지털 포렌식 워터마크 및 은닉 정보를 차후 불법 유포자 식별을 위해 저작물 관리 서버에 저장 및 관리하여야 한다.

3.2.3.4 불법 유통되는 복스캔본 수집

다수의 OSP(Online Service Provider)에 의해 운영되는 웹 사이트를 대상으로 웹 크롤링(Web Crawling) 기술을 이용하여 복스캔 파일들을 수집한다. 이때 별도의 수집 프로그램이 아닌 저작권 보호센터의 불법저작물 추적 시스템(ICOP)으로도 사용이 가능하다.

3.2.3.5 저작권 침해 여부 판단

웹에서 복스캔 파일이 수집된 경우 다음과 같은 과정을 통해 저작권 침해 여부를 판별한다.

첫째, 복스캔 파일에 삽입된 텍스트, 이미지 형태의 디지털 포렌식 워터마크를 검출하여 저작물에 대한 정보를 확인한다.

둘째, 암호화 및 인코딩되어 있는 사용자 식별 정보와 복스캔 파일의 메타데이터 영역에 은닉되어 있는 사용자 식별 정보를 검출하여 디코딩 및 복호화 하여 데이터가 위변조 되었는지 확인한다.

셋째, 서버에 등록되어 있는 저작물 정보와 비교하여 침해 여부를 판단한다.

위 과정을 통해 불법 유포된 파일인 경우 해당 사이트에 대한 증거화면과 사용자의 정보를 저작권자에게 통보한다.

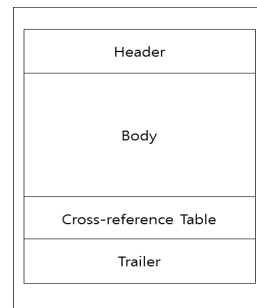
IV. PDF 파일을 이용한 식별정보 은닉 기법

3절에서 제안한 프레임워크 중 가장 중요하게 사용될 디지털 포렌식 워터마크 삽입 및 검출기 구현방안에 대해 설명한다. 4절에서 PDF의 파일 구조와 데이터 은닉 방법에 대해 제시하고, 5절에서 실제 구현한 프로그램에 대해 설명한다.

4.1 PDF 기본 구조

PDF는 미국 어도비시스템즈(Adobe Systems)에서 만든 문서파일 형식이다. 윈도우즈, 맥, 안드로이드 등 거의 모든 운영체제에서 읽거나 인쇄가 가능하며, 원본의 글꼴, 이미지, 그래픽, 문서의 형태 등이 그대로 유지된다는 특성이 있다[19]. 특정 소프트웨어 및 하드웨어에 의존적이지 않기 때문에 사용률이 높다. PDF 파일 문서 포맷은 (그림 4)와 같다.

메타데이터란 데이터를 위한 데이터 이다. 즉 데이터에 대해 추가적인 정보를 기록하기 위한 데이터를 말한다. 예를 들어 사진파일에 대한 메타데이터는 시간, 노출, 플래시 사용여부, 썸네일 등의 정보를 저장하고 있다. PDF 문서파일은 메타데이터를 저장하기 위해 XMP 규격을 따르고 있다. XMP에 대한 기술은 ISO 국제 표준으로 정의되어 있다[20].



(그림 4) PDF 문서 파일 형식

4.2 XMP 영역을 이용한 데이터 은닉

XMP란 다양한 응용프로그램에서 메타 데이터를 보다 효과적으로 관리하기 위한 표준이다. XMP 표준

은 데이터 모델, 저장 모델, 스키마 모델, 메타데이터 모델로 사용된다. PDF 문서뿐만 아니라 TIFF, JPEG, PNG, GIF 파일에도 기록이 가능하다.

PDF 파일의 XMP 영역을 확인하는 방법은 다음과 같다. PDF 파일은 트리구조를 따라 파일을 저장한다. 파일에서 XMP 영역에 접근하기 위해서는 XMP 영역에 대한 주소가 필요하다. 모든 노드에 대한 주소는 /Root 노드에 저장되어 있으며, /Root 노드의 /Metadata 주소를 찾아 접근이 가능하다. [그림 5]는 XMP 영역을 찾기 위한 offset 참조 방법이다. [그림 5]에서 /Root 노드 주소는 '2 0'이다. 해당 노드의 내용 중 /Metadata 노드 주소가 '351 0'임을 확인할 수 있다. '351 0' 노드가 바로 XMP 영역이다.

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
000B4A0	20	34	32	30	0A	2F	52	6F	6F	74	20	32	20	30	20	52
000B4B0	0A	2F	49	6E	66	6F	20	34	20	30	20	52	0A	2F	49	44

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
000B2C0	5D	20	3E	3E	0A	65	6E	64	6F	62	6A	0A	32	20	30	20
000B2D0	6F	62	6A	0A	3C	3C	20	2F	54	79	70	65	20	2F	43	61
000B2E0	74	61	6C	6F	67	0A	2F	50	61	67	65	73	20	31	20	30
000B2F0	20	52	0A	2F	4D	61	72	68	49	6E	66	6F	20	3C	3C	2F
000B300	4D	61	72	68	65	64	20	66	61	6C	73	65	3E	20	2F	
000B310	4C	61	6E	67	28	6B	6F	2D	4B	52	29	20	2F	50	61	67
000B320	65	4C	61	79	6F	75	74	20	2F	53	69	6E	67	6C	65	50
000B330	61	67	65	0A	2F	4D	65	74	61	64	61	74	61	20	33	35
000B340	31	20	30	20	52	0A	2F	4F	75	74	70	75	74	49	6E	74

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
000B4C0	73	74	72	65	61	6D	0A	65	6E	64	6F	62	6A	0A	33	35
000B4D0	31	20	30	20	6F	62	6A	0A	3C	3C	20	2F	54	79	70	65
000B4E0	20	2F	4D	65	74	61	64	61	74	61	0A	2F	53	75	62	74
000B4F0	79	70	65	20	2F	58	4D	4C	0A	2F	4C	65	6E	67	74	68
000B500	20	34	31	39	20	30	20	52	0A	3E	3E	0A	73	74	72	65
000B510	61	6D	0A	3C	3F	78	70	61	63	6B	65	74	20	62	65	67
000B520	69	6E	3D	22	EF	BB	BF	22	20	69	64	30	22	57	35	4D
000B530	30	4D	70	43	65	68	69	48	7A	72	65	53	7A	4E	54	63
000B540	7A	6B	63	64	22	3F	3E	3C	78	3A	78	6D	70	6D	65	

(그림 5) XMP 영역

사용자의 식별 정보를 메타데이터에 은닉하는 이유는 원본 데이터에 영향을 미치지 않고 XMP 영역에 대해 데이터를 확인하기 어렵기 때문이다. 또한 데이터를 암호화 및 디코딩하는 방법으로 난독화 하기 때문에 해당 데이터가 사용자 식별 정보인지 구별할 수 없다. 본 논문에서는 사용자 식별 정보 및 기타 정보에 대해 XMP 영역에 기술한다.

V. 사용자 정보 삽입 및 검출기 구현

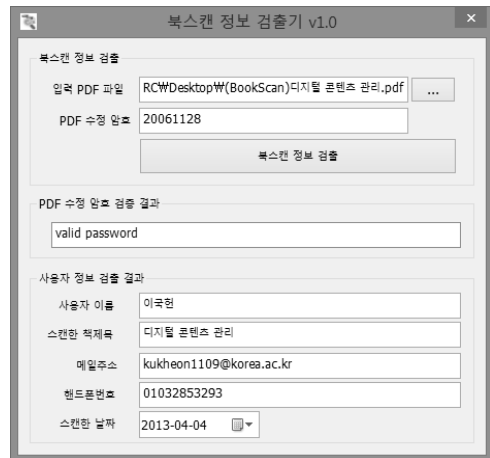
사용자 정보 삽입기 및 검출기는 MFC(Microsoft Foundation Class)와 PDF API를 기반으로 구현하였고, 디지털 포렌식 워터마크를 삽입 및 검출을 위해 AES 암호화 알고리즘과 Base85 인코딩 방식을 사용 하였다. 이때 암호화 과정에서 사용하는 키는 인증기관에서 발급하는 키를 사용하도록 한다. 사용자 정보 삽입기는 사용자 식별 정보를 AES(Ad-

vanced Encryption Standard) 암호화 및 Base85를 이용한 인코딩 과정 후 앞서 4장에서 설명한 PDF 파일의 XMP 영역에 은닉한다. 본 논문의 구현 단계에서 사용자 식별 정보는 사용자 이름, 책 제목, 사용자 메일주소, 연락처, 스캔일자를 포함하며 추후 실제 사용 환경에서는 식별 정보를 추가하거나 제외할 수 있다. 데이터를 은닉한 후, PDF 파일의 수정 암호를 설정하여 사용자가 데이터를 출력 및 편집하는 행위를 방지한다. 사용자 정보 삽입기는 [그림 6]과 같다.



(그림 6) 복스캔 정보 삽입기 실행 화면

사용자 정보 검출기는 PDF 파일에 삽입한 은닉데이터가 있는지, PDF 파일에 설정된 수정암호가 맞는지 확인한다. 수정암호가 일치할 경우 은닉된 정보를 표시한다. [그림 8]과 같이 은닉된 데이터가 없을 경우 'Not Protected' 수정암호가 불일치 경우 'Invalid password' 수정암호가 일치할 경우 'valid



(그림 7) 복스캔 정보 검출기 실행 화면



(그림 8) 수정암호 검증 결과

password'를 표시하며 [그림 7]와 같이 은닉된 데이터를 복호화 및 디코딩하여 표시한다.

VI. 북스캔 관리 프레임워크 분석

본 논문에서 제안한 북스캔 파일 관리 프레임워크는 현재까지 논문으로 발표된 주제가 없었다. 유사한 주제로 디지털 콘텐츠와 관련된 프레임워크에 대해 최홍식[21]의 “디지털 저작물의 관리를 위한 저작권정보 구축시스템 설계 및 이용프로세스에 관한 연구”에서 DRM을 기반으로 한 디지털 저작물 관리 시스템을 제안하였다.

본 논문에서는 플랫폼에 종속적인 DRM 기반이 아닌, PDF 파일 구조를 이용한 플랫폼에 대해 독립적인 방법을 제시하였으며, 제안한 관리 프레임워크의 기반이 되는 IMPRIMATUR과 비교하였다.

기존 IMPRIMATUR 구조에서는 사용자 주체가 1차 출판물을 소유 및 사용하는 것을 가정하여 구성되었기 때문에 북스캔 파일을 관리하기에는 한계점이 존재하여 이를 보완하였다.

보완 방법은 사용자 주체를 북스캔 서비스로 변경하여 각 주체들에 대한 활동을 재구성하였다. 디지털 콘텐츠 시장 구조에서 실제 저작물 소유자가 2차 출판물에 해당하는 북스캔을 행하는 주체가 아닌 북스캔 서비스 업체에 위탁하여 북스캔을 행하기 때문에 각 주체에서 요하는 정보와 활동이 다르기 때문이다.

변경된 주체에 대해 북스캔 서비스 주체가 요하는 저작권 이용, 북스캔 업체 정보, 업체 인증, 포렌식 워터마크 프로그램, 암호화 키와 관련하여 각 주체들의 활동이 추가/변경 되었다.

VII. 결 론

스마트 기기의 사용자 증가 및 북스캔 대행 서비스 시장 출현으로 인해 많은 사용자들이 북스캔 대행 서비스를 이용하고 있다. 하지만 현재 북스캔 대행 서

비에 대한 정책과 법제도에 대한 정립이 없다. 새로운 시장에 대해 대처를 하지 않는다면 과거 소리바다의 음원 시장 붕괴와 같이 불법공유로 인해 출판 산업에 큰 타격을 줄 것으로 예측된다.

디지털 콘텐츠에 대한 법제도 정비를 위해서는 기술적인 보완 조치가 필요하다. 본 논문에서는 실현 가능한 기술을 기반으로 하여 북스캔 파일 관리 체계를 제안하고 사용자 식별 정보를 삽입 및 검출할 수 있는 프로그램을 구현하였다. 제안한 체계에서 필요한 기술은 PDF 파일에 데이터를 은닉하는 기술, 소프트웨어 검증 기술, 암호 기술 등이다.

본 논문에서 제안한 프레임워크가 의미를 갖기 위해서는 제도적인 강제성을 바탕으로 출판 업체와 북스캔 대행업체, 그리고 제 3의 신뢰기관이 협력하여 실행해야 한다. 제안한 통합관리 체계가 수행된다면, 북스캔 서비스와 관련한 불법 유통이 일어날 가능성을 줄일 수 있을 것이다.

참고문헌

- [1] 남동선, “전자책 및 전자출판 시장 동향,” 정보과학회지, 28(10), pp. 16-25, 2010년 10월
- [2] 문화체육관광부, “전자출판산업 육성방안,” 2010년 4월
- [3] 경향신문, “문화부, 전자책 파일 서비스 저작권법 위반,” 2011년 5월
- [4] 김정호, 이완재, “소리바다 사건에 대한 법정제학적 분석,” 비교사법, 9(1), pp. 451-481, 2002년 4월
- [5] 김원재, “e-Book 시장의 현황 및 과제,” 한국통신사업자연합회, 52, pp. 72-77, 2010년 6월
- [6] 두일철, 김용범, “모바일 플랫폼에서의 e-Book 콘텐츠 활성화 연구,” 인문콘텐츠, 21, pp. 221-245, 2011년 6월
- [7] 노준석, 이용준, “세계 전자책(e-Book) 시장의 현황과 이슈 분석,” 한국콘텐츠진흥원, 60, pp. 1-29, 2012년 10월
- [8] 김윤명, “이러닝(e-Learning)의 활성화를 위한 법적 검토 : 저작권법을 중심으로,” 한국산업재산권법학회, 31, pp. 303-350, 2010년 4월
- [9] 매일경제, “북스캔 저작권 침해 논란 확산,” <http://news.mk.co.kr/newsRead.php?year=2012&no=722246> 2012년 11월,
- [10] “저작권법 (일부개정 2011.12.2. 법률 제 11110

- 호 시행일 2013.3.15.)”, <http://www.law.go.kr/lsSc.do?menuId=0&p1=&subMenu=1&nwYn=1&query=%EC%A0%80%EC%9E%91%EA%B6%8C%EB%B2%95&x=-687&y=-208#liBgcolor0>
- [11] 정완, “소프트웨어 불법복제 실태와 법제도적 개선 방안,” 한국형사정책연구원, pp. 9-178, 2003년 12월
- [12] 오원근, “DRM 표준화 및 평가 기술,” 한국전자통신연구원, 20(4), pp. 139-154, 2005년 8월
- [13] 정현일, “폐쇄적 DRM 전략이 음악시장에 미친 영향과 정책적 시사점,” 한국문화관광연구원, 2008년
- [14] 김평수, “애플이 주도한 DRM-Free 전략의 경제성 분석,” 글로벌문화콘텐츠학회, 2, pp. 193-216, 2009년 6월
- [15] IMPRIMATUR, <http://www.imprimatur.alcs.co.uk/index.htm>
- [16] 김진목, 유황빈, “워터마크와 공개키 암호 알고리즘을 이용한 멀티미디어 콘텐츠 보호에 관한 연구,” 한국통신학회지, 17(3), pp. 54-63, 2000년 3월
- [17] Owner Password, <http://pcsupport.ab\out.com/od/erroro/g/pdf-owner-password.html>
- [18] 정혜정, 박인수, 양승배, “소프트웨어의 신뢰성 평가 방법,” 정보과학회지, 23(3), pp. 56-61, 2005년 3월
- [19] Adobe systems incorporated, PDF Reference, 6th Ed., Nov. 2006
- [20] Adobe systems incorporated, “XMP specification- Part 1: Data model, serialization and core properties,” ISO 16684-1:2012, Feb 2012.
- [21] 최홍식, 이정수, “디지털저작물의 저작권관리를 위한 저작권정보 구축시스템 설계 및 이용프로세스에 관한 연구,” 한국정보관리학회지, 22(3), pp. 163-181, 2005년 9월

 <저자소개>



이 국 헌 (Kuk-heon Lee) 학생회원
 2012년 2월: 배재대학교 컴퓨터공학 공학사
 2012년 3월~현재: 고려대학교 정보보호대학원 석사과정
 <관심분야> 디지털 포렌식



정 현 지 (Hyun-ji Chung) 학생회원
 2010년 2월: 고려대학교 컴퓨터정보공학, 산업시스템공학 공학사
 2012년 2월: 고려대학교 정보경영공학전문대학원 석사
 2012년 3월~현재: 고려대학교 정보보호대학원 박사과정
 <관심분야> 디지털 포렌식



유 대 결 (Dae-gull Ryu) 학생회원
 2007년 2월: 고려대학교 수학, 컴퓨터학 이학사
 2012년 3월~현재: 고려대학교 정보보호대학원 석사과정
 <관심분야> 저작권 보호 기술



이 상 진 (Sang-jin Lee) 종신회원
 1989년 2월~1999년 2월: 한국전자통신연구원 선임 연구원
 1999년 2월~2001년 8월: 고려대학교 자연과학대학 조교수
 2001년 9월~현재: 고려대학교 정보경영공학전문대학원 교수
 <관심분야> 대칭키 암호, 정보은닉이론, 컴퓨터 포렌식