불법콘텐츠 추적 기술 연구동향

Research Trend of Illegal Contents Trace Technology

유비쿼터스 시대를 주도할
디지털콘텐츠 기술 특집

정혜원 (H.W. Jung) 콘텐츠보호연구팀 연구원
이준석 (J.S. Lee) 콘텐츠보호연구팀 책임연구원
서영호 (Y.H. Suh) 콘텐츠보호연구팀 팀장

목 차

인터넷 환경이 급속도로 발전함에 따라 멀티미디어의 범람과 사용자들의 유료 콘텐츠 사용에 대한 인식 부족으로 디지털 콘텐츠의 지적재산권 침해가 빈번하게 발생하고 있으며, 이러한 불법콘텐츠들의 무분별한 공유는 디지털콘텐츠 산업 발전을 저해하는 심각한 문제로 대두되고 있다. 이에 따라 최근 디지털 콘텐츠의 저작권 보호를 위해서 저작권 단속활동을 강화하고 있으며 다양한 P2P/웹하드 환경에서도 불법콘텐츠를 검색 및 다운로드하고 콘텐츠를 식별하고 불법배포자를 추적할 수 있는 기술을 필요로 하게 되었다. 그러나 대부분의 불법콘텐츠들은 약 80%가 P2P, 웹하드, 동호회/카페를 통해서 전파되고 있으며 각 업체별 클라이언트 프로그램의 다양성 때문에 불법콘텐츠의 검색 및 불법복제자 추적이 매우 어려운 실정이다. 불법콘텐츠의 추적 및 차단을 위해 워터마킹/핑거프린팅, 네트워크 모니터링, 매크로 프로그램, 공개 P2P 프로토콜 조작, 페이크 파일 유포 등의 기술들이 사용되고 있다. 본 고에서는 불법콘텐츠 국내외 시장현황과 업체의 기술동향을 살펴보고 불법배포자 추적을 위해 필요한 핑거프린팅과 특징점 기반 콘텐츠 식별 기술에 대해 설명한다. 그리고 P2P와 웹하드에서 불법콘텐츠들을 자동으로 다운로드하고 불법배포자들을 추적할 수 있는 시스템 아키텍처에 대하여 설명한다.

Ⅰ. 서론

컴퓨터 성능의 급격한 향상과 AV 코딩 기술의 발전으로 다양한 콘텐츠 저작도구들을 통해 콘텐츠들을 쉽게 리코딩/편집할 수 있게 되었고, 불법으로 제작된 콘텐츠들은 초고속 인터넷망을 타고 제3자에게 빠르게 전파되고 있다. 이러한 디지털콘텐츠는 손실없이 복사될 수 있다는 특성과 사용자들의 유료 콘텐츠 사용에 대한 인식부족으로 지적재산권 침해가 빈번히 발생하고 있다. 온라인상 불법저작물 이용형태를 살펴보면, 포털사이트의 카페나 블로그에서의 불법링크, 소리바다・프루나 등의 P2P 프로그램을 이용한 개인간 공유, 대용량 서버에 불법 복제물을 저장한 후 불특정 다수와 공유하는 웹하드 등의 침해가 가장 심각한 것으로 나타나고 있다. 그러나 대부분의 P2P/웹하드는 각 업체별로 다른 인터페이스와 프로토콜을 사용하는 다양성 때문에 불법콘텐츠의 검색 및 불법복제자 추적이 매우 어려운 실정이다. 이러한 지적재산권 침해로 인한 피해규모는 전세계적으로 매우 크며 그로 인해 새로운 창작산업에 막대한 피해를 주고 있어 저작권 관련 이슈가 사회 문제로 대두되고 있다. 이에 최근 저작권 관련 기관 및 협회에서는 법・제도를 정비하고 저작권 단속활동을 강화하고 있다. 저작권 침해를 막기 위한 기술적인 보호조치로 디지털 저작권 관리(DRM), 워터마킹(WM) 기술들이 각광 받고 있으며, 불법콘텐츠의 추적 및 차단을 위해서 핑거프린팅(FP), 네트워크 모니터링, 매크로 프로그램, 공개 P2P 프로토콜 조작, 페이크(fake) 파일 유포 등의 기술들이 사용되고 있다. 여기에서는 이러한 불법콘텐츠들을 자동으로 검색하고 수집할 수 있는 기술들에 대해 최신 현황을 소개하고 이러한 불법콘텐츠 유포의 온상이 되고 있는 P2P/웹하드에서의 불법복제자를 추적하기 위한 방법들에 대해서 설명하고자 한다. 또한 핑거프린팅과 특징점 기반 콘텐츠 식별 기술을 이용하는 불법콘텐츠 추적 시스템에 대한 아키텍처를 제안하고자 한다.

본 고의 제Ⅱ장에서는 국내외 시장현황 및 업체의 기술동향을 살펴보고 제Ⅲ장에서는 불법콘텐츠 추적 및 차단을 위한 기술들에 대한 소개를 한다. 그리고 제Ⅳ장에서는 P2P와 웹하드에서 불법콘텐츠들을 자동으로 다운로드하고 불법배포자들을 추적할 수 있는 시스템 아키텍처에 대하여 설명한다.

Ⅱ. 시장 및 업체 동향 분석

1. 불법콘텐츠 현황

불법복제가 난무하는 이유는 저작권법에 대한 인식 부족, 파일공유가 가져다 주는 편리함과 이익, 수많은 사람들의 관행 속에서 오는 군중심리, 현실적으로 형사상 적발되더라도 처벌강도가 낮을 것이라는 안일함 등이 요인이라 할 수 있다. 온라인에서 유통되고 있는 이미지, 음악, 게임, 영화 등의 디지털 콘텐츠들은 무료라는 인식아래 무분별하게 공유되어져 왔고 이러한 무단복제 현상은 위험수위에 다달았다. 불법복제로 인한 피해규모를 살펴보면, 전세계 음반불법복제에 따른 피해액은 세계시장의 35%인 45억 달러(출처: IFPI)이며, 영상 불법복제로 인한 전세계 추정 피해액은 약 30억 달러(출처: MPA)에 달한다고 한다. 국내 온라인 불법 영상물 피해액은 연간 1000억 원(출처: 한국영상협회)이며, 온라인 불법복제 시장은 2004년 4천400억 규모로 추정(출처: 문화관광부)하고 있다.

초고속망 인프라와 인터넷 문화가 발달한 우리나라의 경우 누구나 쉽게 불법콘텐츠에 접근할 수 있게 되었고, 이러한 이유로 미국 무역대표부(USTR)

〈표 1〉 이용형태에 따른 불법콘텐츠 유통경로

구분	유통경로(%)	내용
P2P	38.5	소리바다, 당나귀, 구루구루, 온파일 등
웹하드	20.3	디스크팝, 네오폴더, 아이팝, 피디박스 등
동호회/카페	20.3	기존 커뮤니티를 활용한 불법콘텐츠 유통 및 정보교환

<자료>: 한국콘텐츠산업협회(KIBA), 2005.

〈표 2〉 국내 P2P/웹하드 분류

동작방식	설명	서비스 명
중앙 중재자형 P2P	사용자의 PC에 설치된 프로그램에서 자료 공유를 하지만, 자료 검색은 공유된 리스트들을 가지고 있는 서버에서 진행하는 방식	오렌지파일, 온파일, 쿨피, 엔진닷컴, 파일 박스, 파일나라, 피투피아 등
순순 분산형 P2P	자료의 공유여부와 자료의 검색여부를 모두 개개인의 PC에 설치된 특정 프로그램으로 실행하는 방식으로, 파일공유를 위한 특정의 서버는 없는 형태임	고부기, 브이쉐어, 소리바다, 이동키, 이동키2000, 이물, 프루나, 썬파일 등
웹하드	회사에서 대형 저장공간을 제공하고 사용자들이 일부분을 임대하여 사용하는 형태. 개인적인 용도로 사용되는 저장공간은 불법자료들로 채워지고 있음	네오폴더, 아이팝, 팝폴더, 폴더플러스, 큐빅, 토토디스크, 디스크피아, 로드빅 등
스트리밍 웹하드	회사 서버에서 전용 프로그램으로 실시간 영상/음성을 전송해주는 방식	모기의 주크박스, 뮤직통, 추억속으로 떠나는 음악여행 등
클럽형 웹하드	웹하드 방식과 비슷하며 클럽 관리자에게 모든 법적인 책임을 부여하고 클럽회원에게만 자료를 공유	아이디스크, 엑스톡, 디스크팟, 애니파일, 폴더로, 하나포스, 클럽폴더 등
웹방식 웹하드	웹하드방식과 비슷하지만 사용자들이 회사의 서버에 자료를 올리면 해당회사에서의 웹 게시판을 통해서 사용자들이 다운로드하는 방식	위디스크, 인포마스터, 짱공유넷 등
링크 웹하드	자료를 직접 제공하지는 않지만 공유된 자료의 위치를 게시판에 공지하여 사용자들이 쉽게 자료를 찾게 해주는 방식	짱공유닷컴, 오지지코리아, 공유넷, 꼬게네, 다운박스넷, 따오기, 러빙카페 등

에서는 한국을 지적재산권 침해 우선감시 대상국(PWL)으로 지정했다. 이에 저작권심의위원회, 한국영상협회 및 저작권보호센터 등에서는 저작권 침해 단속활동, 공정한 저작물 이용 촉진 및 교육·홍보 사업을 추진하며 권리자와 이용자 간에 합리적인 저작권 이용 질서 확립을 위해 노력하고 있다.

최근 MP3 시장이 유료화의 길을 걷고 있는 가운데 51.1%에 달하는 네티즌이 아직도 무료 P2P 사이트에서 MP3의 음원을 다운로드 하고 있고, <표 1>에서 보는 바와 같이 79.1%의 불법콘텐츠들이 P2P, 웹하드, 동호회/카페를 통하여 유통되고 있다. 사용자들이 P2P와 웹하드를 선호하는 이유는 고사양의 서버와 네트워크를 제공하여 빠른 속도를 보장하고 파일 공유의 편리한 인터페이스와 커뮤니티를 제공하기 때문이다. 국내에서 활발하게 서비스되고 있는 P2P/웹하드는 40여 종이며 업체별로 클라이언트 프로그램과 커뮤니티를 제공하고 있고 각기 다른 프로토콜을 사용하고 있다. <표 2>에서는 P2P/웹하드 서비스들은 동작방식에 따라 분류한 것이다.

2. 업체 기술 동향

국내에서는 최근 들어 이러한 저작권 침해 피해를 줄이기 위해 저작권 단속활동을 강화하고 있다. 정보통신윤리위원회에서는 웹 검색 로봇을 이용하여 음란 사이트나 도박, 카지노, 자살 등의 불법사이트를 검색하고 단속하는 활동을 하고 있다. 음란사이트의 경우 텍스트와 이미지를 분석하여 유해사이트인지를 판별하는 기술을 이용하고 있다[1]. ㈜노프리에서는 P2P를 통한 불법적인 자료 공유로 인해 발생되는 저작권 침해를 막는 솔루션인 CPS를 개발하여 서비스하고 있다[2]. 이 시스템은 다량의 페이크(fake) 파일을 만들어 온라인에 유통시킴으로써 원본을 방어하는 기술이다. ㈜지란지교소프트에서는 사용자 PC에서 음란 동영상을 보려고 할 때 서버에서 보내주는 해시 값과 비교하여 일치할 경우에 플레이어를 종료시키게 하는 기술을 개발하였다[3]. ㈜소프트세이브, ㈜펀팡, ㈜이카피라이트코리아에서는 현재 개봉중이거나 개봉예정인 영화들의 단속 대행권을 확보하고 웹하드 및 P2P를 대상으로 적극적인 단속활동을 수행하고 있다.

국외 업체들의 경우, 네덜란드의 Philips사에서는 FP을 이용하여 디지털 비디오 파일을 식별해 전송을 중단하는 기술을 개발하고 있다[4]. 이미 몇몇 대학의 네트워크에서 음악 파일의 복사를 추적·방지하는 데 사용하는 것과 유사한 방식으로 음악 파일에 FP 정보를 삽입하고 네트워크 내부에 설치된

소프트웨어가 교환되는 파일을 감시하고 데이터베이스와 대조하여 불법이면 파일전송을 중단시키는 방법이다. Digimarc사에서는 이미지 안에 저작권 정보, 디지털 권리, 사용권, 웹에서 사용할 수 있는 권한 등의 디지털 워터마크 정보를 삽입하고 추출할 수 있는 프로그램인 PictuerMarc를 개발하였다[5]. MarcSpider는 검색된 이미지들 중 워터마크가 삽입되어 있는 이미지만을 추출하여 보여줌으로써 저작권자가 자신의 이미지가 어디에서 어떻게 사용되고 있는지를 알 수 있도록 해주는 시스템이다. 하지만 저작권 정보에 대한 워터마크이므로 불법배포자를 추적할 수는 없다. BayTSP사에서는 콘텐츠에 Signature Acronym을 삽입/편집할 수 있는 My-Gudio를 개발하였다[6]. 이것은 DB에 저장된 FP 정보를 이용하여 클라이언트에서 권한이 부여된 콘텐츠만을 검색할 수 있도록 하는 기능을 제공하고 있다. 또한 공개 P2P 공유 서버인 BitTorrent와 eDongey 상의 불법복제물을 검색하여 자동으로 최초 배포자를 추적하여 찾아내는 FirstSource를 개발하였다. AudibleMagic사에서는 오디오 콘텐츠의 Psycho-Acoustical 속성을 이용 불법복제물을 인식하는 FP 기반의 CopySense을 개발하였다. Over-peer사는 ㈜노프리와 같은 방법으로 P2P상에 방대한 양의 페이크 파일들을 업로드하여 원본 콘텐츠를 보호하는 솔루션을 제공하고 있다. 이 외에도 워터마킹 기술을 보유하고 있는 업체로는 Alpha-Tec사, Giovanni사, Signum Technologies사, Media-Sec사 등이 있다.

Ⅲ. 불법콘텐츠 추적 및 차단 기술

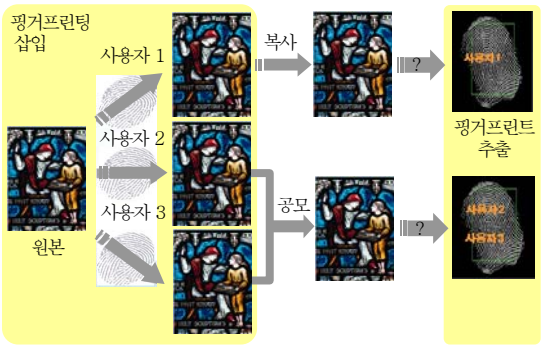
적법한 콘텐츠를 보호하는 방법에는 불법배포자를 추적하는 기술과 불법콘텐츠를 다운로드 하지 못하게 차단하는 기술로 나눌 수 있다. 불법배포자를 추적하기 위해서는 P2P나 웹하드에서 불법으로 공유한 콘텐츠들을 다운로드하는 기술과 불법콘텐츠인지를 식별할 수 있는 기술이 필요하다. 본 절에서는 이러한 기술들에 대해서 설명한다.

1. 워터마킹 및 핑거프린팅 기술

디지털 워터마킹(digital watermarking)이란 사진 이미지나 음악 파일 같은 멀티미디어 콘텐츠에 인간의 시각이나 청각으로는 식별이 어렵도록 저작권정보를 삽입하여 배포하고 저작권 분쟁이 발생하였을 경우 이를 추출하여 저작권을 보호하는 방법이다[7]. 이때 메시지는 단지 파일 뒤에 첨가되는 것이 아닌 파일 내용 안에 뒤섞이는 것으로써, 원래 파일보다 크기가 늘어나지 않고, 원래 파일 포맷 변경이 되지 않는다.

핑거프린팅(FP) 기법은 모든 콘텐츠에 동일한 저작권 정보를 삽입하는 워터마킹 기법과는 달리 콘텐츠마다 각기 다른 구매자정보를 삽입함으로써 불법복제 및 유통행위가 발견되었을 때 불법 배포자를 추적하고자 하는 기술이다[8]. 저작권 정보만을 이용하는 워터마킹보다 적극적인 의미의 보호 기법이라 할 수 있으며 부정자 추적(traitor tracing) 기술로 활용되고 있다[9],[10]. (그림 1)에서는 원본 콘텐츠에 핑거프린팅 정보가 삽입되는 과정과 추출되는 과정을 보여주고 있다.

같은 콘텐츠라도 FP 후의 콘텐츠가 조금씩 다르다는 특성 때문에 구매자들이 서로 그 차이를 공모하여 삽입된 FP를 지우려고 하는 공모공격(collusion attack)이 존재하게 된다[8]. FP 기술은 이러한 삭제 및 간섭, 기하학, 암호화, 프로토콜 등 공모공격을 고려하여 강인하도록 개발되어야 한다.



(그림 1) 핑거프린팅 삽입 및 추출과정

2. 매크로 프로그램

매크로(macro)는 반복적인 과정을 정의하여 스크립트나 어셈블리로 명령어를 지정하고 같은 동작을 반복하게 하는 것을 말한다. 매크로 작성 방식은 스크립트를 이용하여 수행할 작업을 순서대로 지정하는 방법과 수행할 작업을 사용자의 행위를 리코딩하여 작성하는 방법이 있다.

매크로는 클라이언트 프로그램의 형태에 구애받지 않고 적용할 수 있다는 것과 사용자의 단순 반복적인 작업을 효율적으로 대신할 수 있다는 장점이 있다. 그러나 복잡한 스크립트를 이해해야 하는 번거로움이 발생하며 다소 복잡한 매크로를 작성할 경우, 스크립트 프로그래밍도 복잡해지는 경향이 있다. 또한 매크로만으로는 불법콘텐츠 추적을 위해서 필요한 정보를 가져오지 못하는 경우도 발생한다. 불법콘텐츠 추적을 위해서 매크로 작성 시 필요로 하는 스크립트의 종류는 <표 3>에서 보여주고 있다.

이러한 매크로의 제작, 수정, 실행 등을 편리하게 제공하는 상용 매크로 프로그램으로는 Insight Software Solutions사에서 개발한 Macro Express[11]가 대표적이며 그 외에도 Robot Task, Macro Magic, 그리고 Journal Macro 등이 있다.

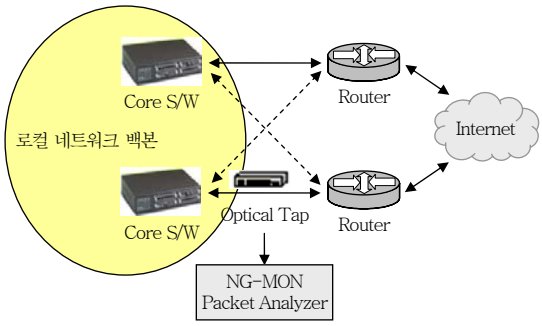
3. 네트워크 모니터링

라우터나 기가스위치에서 네트워크 모니터링을 통해 연결을 차단하거나 불법 배포자를 추적하는 방법이다[4],[12]. 그러기 위해서는 트래픽을 실시간으로 분석하여 콘텐츠가 불법인지 아닌지를 검사해야 하고 불법이라면 연결을 차단하거나 불법배포자 정보를 수집할 수 있어야 한다. (그림 2)와 같이 학교나 기관 등의 로컬네트워크 백본망에 설치하여 운영한다면 밖으로 유출되는 불법콘텐츠를 효과적으로 차단할 수 있다.

불법콘텐츠 검사를 위해서는 핑거프린팅과 특징점 기반 식별 기술이 필요하다. 패킷들을 이용하여 원본콘텐츠를 재조합하고 미리 삽입해 놓은 핑거프린팅 코드가 검출된다면 불법콘텐츠라는 것을 알 수 있을 것이다. 또한 사용자간 주고 받은 패킷들의 복원으로 source IP, destination IP, 로그인 ID 등의 사용자 정보를 검출해 낼 수도 있다.

〈표 3〉 매크로 스크립트 분석

구분	기능 설명	스트립트 명
Dialogs	사용자가 정의하는 임의의 다이얼로그 박스, 옵션 박스를 생성	Text Box Display
Files/Folders	특정 파일 및 폴더에 대한 open, close, delete 등의 기능	Copy, Create, Delete, Rename 등
Keyboard	키보드의 값을 입력 기본적인 문자들의 입력과 특수키의 켜짐/꺼짐, control, alt, shift 키의 down, up을 제어	Repeat Delay, Speed, Text, Alt Key, Control Key 등
Logic	논리적인 연산으로 분기를 하는 프로그래밍적인 부분으로 변수 값의 논리연산이나 분기문, 반목문 지원	Case, If, Else, Switch, Or XOR, And, End If, Repeat 등
Mouse	현재 마우스의 상태/위치에 대한 값을 알아내거나 마우스를 이동시키고, 버튼을 클릭하는 등의 제어를 하는 기능	Get Position, Left Button, Right Button, Wheel, Move 등
Network	네트워크의 연결, 연결 끊기 등의 제어, 네트워크 상태 파악 등의 기능	Connect, Disconnect, Online 등
Timing	키보드나 마우스의 입력 속도 등을 제어, 특정 시간이나 특정 행동이 이루어질 때까지 기다리게 하는 기능	Delay, Pause, Mouse Speed 등
Variables	특정한 값을 받을 수 있는 변수를 정의 수학적 연산이나 문자열의 처리 등의 기능	Set, Get, Modify, Restore, Clear 등
Window Controls	window control의 상태를 판단하고, 사이즈나 위치, caption 등을 알 수 있고 특정 위치의 픽셀의 색 값을 알 수 있음 프로그램의 실행, window 창의 활성화, resize, 최소화, 최대화 등의 제어 기능	Launch, Shut Down, Hide, Maximize, Minimize, Reposition, Resize, Show, Activate, Close 등



(그림 2) 네트워크 모니터링 구조

4. 공개 P2P 프로토콜 조작

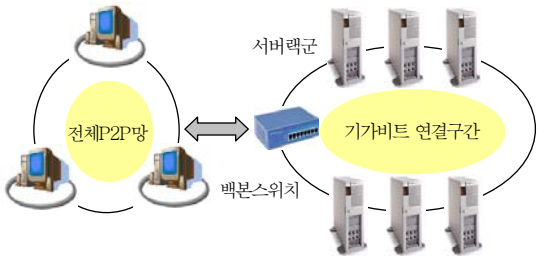
전세계적으로 공개 프로토콜을 사용하는 P2P는 eDonkey, FastTrack, Gnutella, BitTorrent 등이 있다[13]. eDonkey는 현재 전세계에서 가장 많은 사용자를 확보한 프로그램 중 하나로써 한국에서도 많은 사용자들을 확보하고 있다. eDonkey는 여러 개의 소스로부터 동시에 파일을 전송하는 방식인 MFTP을 사용하고 있기 때문에 사용자가 많아질수록 전송속도가 빨라지는 것이 특징이다. eDonkey는 중앙 중재자형(central arbiter type)으로 중앙의 서버가 피어(peer)들의 등록정보를 관리하고 있으면서 어떤 피어가 P2P 네트워크에 접속하게 되면 다른 피어들에게 그 피어의 접속을 알려줌으로써 피어들 간에 통신하는 구조이다. 오픈소스 이뮬(emul)을 활용하여 개발되었으며 같은 부류로 당나귀, 프루나, 오버넷, 동키호테 등이 있다. Gnutella는 오픈 프로토콜을 사용하여 어떤 프로그램이든 이 네트워크에 접속해서 파일을 공유할 수 있도록 공개된 네트워크이다. Gnutella는 순수 분산형(pure distributed type)으로 중앙의 관리 서버 없이 모든 피어들이 서로 대등한 관계로 연결되는 구조를 가진다. 이러한 분산 구조에서는 중앙의 관리 서버가 없기 때문에 트래픽이 한 곳에 집중하지 않고 P2P 네트워크가 견고하게 형성되는 장점은 있으나 서로 간에 데이터 전송이나 피어를 검색하는 기능에 있어서는 효율성이 떨어진다. Gnutella 네트워크 기반 공유 프로그램으로 XoloX, LimeWire, Shareaza 등이 있다. FastTrack은 몇몇 정해진 프로그램만 접속할 수 있는 폐쇄적인 네트워크로서 Kazaa, Grokster, iMesh 등이 있다.

이러한 공개 P2P 프로토콜들은 단순한 프로토콜의 조작만으로 쉽게 원하는 콘텐츠를 검색하고 다운로드할 수 있다. eDonkey의 경우 제공되는 콘텐츠들의 고유 해시 값을 비교하여 같은 콘텐츠의 경우 중복 다운로드를 피할 수 있으며, 부분 다운로드나 간단한 사용자 정보도 수집할 수 있다.

5. 페이크 파일 유포

P2P 서버에 원본파일의 이름과 포맷이 같고 크기도 비슷한 다량의 페이크 파일을 업로드시켜 실제 원본파일이 P2P상에 업로드 되어도 다른 유저들이 원본 파일에 접근하기 힘들게 만드는 방법이다[2]. 페이크 파일을 받은 유저가 다시 이 파일을 공유시킴으로써 페이크 파일이 점차 확산되게 되는 효과도 있다. 사용자들이 불법콘텐츠를 다운로드하기 위해 키워드 검색을 할 경우 페이크 파일에 접근할 확률을 높이기 위해서는 (그림 3)과 같이 국내유선통신사(ISP)의 인터넷 백본망을 직접 연결하여 수 기가바이트(Gigabyte)의 대역폭을 확보하는 것이 효과적이다. 또한 원본파일을 업로드하는 사용자를 검색하여 경고 조치를 취하여 원본파일의 공유를 막음으로써 원본파일의 다운로드 확률을 낮출 수 있다.

이러한 방법은 저작권자의 입장에서 지적소유권 유출로 인해 발생되는 경제적 손실을 사전에 방어할 수 있다는 큰 장점이 있으며, 나아가 네티즌의 입장에서도 고의적인 의도로 지적소유권을 침해하는 소



(그림 3) 페이크 파일 유포망 구조

수의 네티즌들을 제외한 대다수의 일반 사용자들의 의도치 않은 실수로 인한 불법적 행위를 차단하여 네티즌들이 법적 시비 여부에 휘말리지 않도록 도와준다.

Ⅳ. 불법콘텐츠 추적시스템 아키텍처

온라인상에의 불법콘텐츠들은 P2P/웹하드를 통하여 공유되고 있지만 다양한 종류의 P2P/웹하드 프로그램의 불법콘텐츠 전부를 추적한다는 것은 매우 어렵다. 본 장에서는 불법콘텐츠 추적을 위한 시스템 구축을 위해 필요한 아키텍처를 제안하고자 한다. 1절에서는 불법콘텐츠 추적시스템의 요구사항에 대해서 도출하고 2절에서는 시스템 아키텍처에 대하여 설명한다.

1. 요구사항 분석

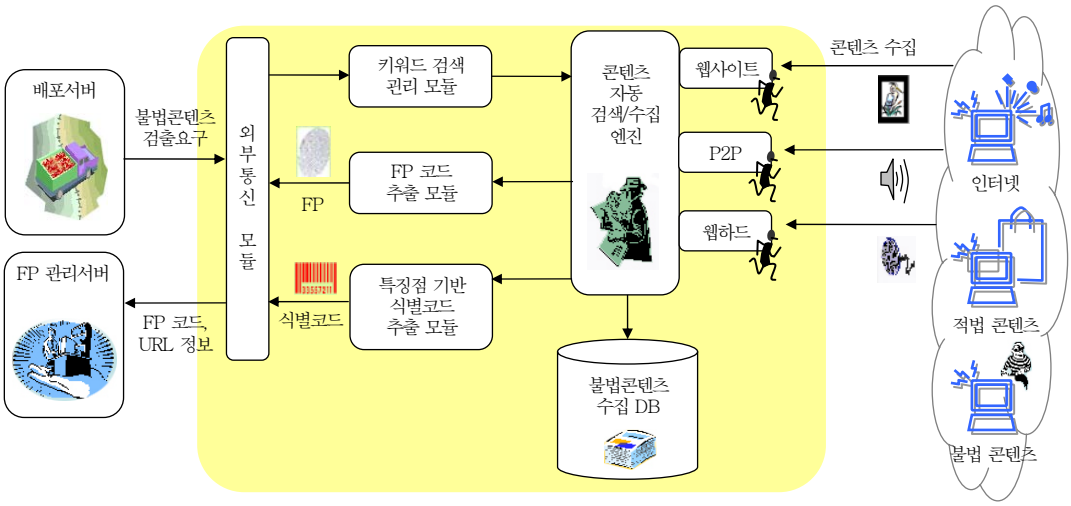
국내에는 수십여 개의 P2P/웹하드가 서비스중에 있으며 모든 프로그램에 적용 가능하도록 확장성을 제공해야 한다. 즉, 업체마다 통신프로토콜과 인터페이스가 다르더라도 에이전트나 프로그램에 의해 자동으로 추적하고 파일을 다운받을 수 있어야 한다. 그리고 수천만 건의 콘텐츠를 전부 다운로드 한다는 것은 네트워크 대역폭과 저장 공간을 고려해 볼 때 현실적으로 불가능하기 때문에 콘텐츠의 일부분만을 다운로드하여야 한다. 같은 파일일 경우 중복 다운로드를 피해야 하며, 다운로드 받은 파일의 진위여부를 확인할 수도 있어야 한다. 최근 P2P 사용자들은 파일 명을 특정 일련번호로 하고 클럽이나 개인 웹사이트에 실제 파일 명을 분리해서 제공하는 경우가 늘고 있다. 이런 경우 단순히 파일 명만으로는 검색할 수 없고 불법배포자도 추적할 수가 없다. 따라서 콘텐츠 특징을 이용한 콘텐츠 식별이 가능해야 한다. 또한 FP 코드가 있다면 FP 코드 추출을 통해서 처음 파일을 유포한 사용자의 정보를 추적할 수 있어야 한다. 불법콘텐츠라고 확인되었을 경우 불법콘텐츠 배포자의 정보를 수집하고 증거자료를 확보할 수 있어야 하며 불법콘텐츠 수집 DB를 통해 관리될 수 있어야 한다.

2. 시스템 아키텍처

불법콘텐츠 추적 시스템에서는 디지털콘텐츠 저작권을 보호하고 P2P와 웹하드에서 불법 복제되어 유통되고 있는 콘텐츠들을 검색/수집하며, 핑거프린팅 기술과의 연동을 통해 불법콘텐츠 첫번째 배포자를 추적할 수 있는 시스템이다. 불법콘텐츠 시스템 아키텍처의 구조는 (그림 4)에서 보여주고 있다.

시스템 동작순서를 살펴보면, 먼저 배포서버에서 원본콘텐츠와 제목, 저작자 등과 같은 키워드 정보들을 전달하고 불법콘텐츠 검출을 요구한다. 불법콘텐츠 자동 추적 엔진은 단순 웹페이지/P2P/웹하드에서 자동으로 콘텐츠들을 검색하고 다운로드 받는다. 검색 및 수집결과들을 불법콘텐츠 수집 DB에 저장하고, 다운로드 받은 콘텐츠는 특징점 기반 식별코드 추출 및 핑거프린팅 코드 추출 모듈을 통해서 특징점 식별코드 및 핑거프린팅 코드 정보를 얻는다. 최종적으로 이러한 핑거프린팅 정보, 특징점 정보, 콘텐츠 수집 메타 정보 등을 FP 관리서버에게 전달하게 된다.

콘텐츠 다운로드 방법은 제Ⅲ장에서 설명한 기술들을 이용하여 P2P/웹하드에서 불법콘텐츠들을 검색하고 다운로드 할 수 있다. P2P의 종류에는 당나귀, 이물과 같은 이동키류 및 그누텔라 등 공개 프로토콜을 사용하는 P2P 프로그램과 비공개 프로토콜 P2P 프로그램으로 나눌 수 있다. 공개 프로토콜의 경우 프로토콜 조작만으로도 원하는 콘텐츠를 검색하고 다운로드 할 수 있지만 비공개 프로토콜의 경우에는 다른 접근방법이 필요하다. 여기서는 매크로 스크립트와 애드온(add-on)을 사용하여 불법콘텐츠를 추적하는 방법을 제안하고자 한다. 매크로 스크립트는 마우스/키보드 및 시스템을 제어할 수 있으며, 변수선언, 반복문 및 조건문 등을 사용하여 복잡하고 반복적인 동작 수행이 가능하다. 이러한 매크로 스크립트는 다양한 P2P/웹하드 프로그램에 쉽



(그림 4) 불법콘텐츠 추적시스템 아키텍처

게 적용 가능하다는 장점을 가진다. 애드온이란 특정한 프로그램의 기능을 보강하기 위해 추가된 프로그램들을 말하며 플러그인과 비슷한 것이다. 애드온 프로그램을 이용하면 클라이언트 프로그램에 종속적으로 동작하지만 매크로 스크립트로 어려운 제어나 사용자 정보 수집 등을 할 수 있다.

불법배포자 추적을 위해서는 다운로드 받은 콘텐츠가 불법으로 판명되었을 경우 사용자 정보를 수집하고 증거자료를 확보해야 한다. 특징점 기반 식별코드는 콘텐츠들을 구별할 수 있는 고유한 값들로서 특징점 서버에서 관리하고 있으며 콘텐츠의 일부분 특징점 코드라도 해당 파일을 정확하게 검출해 낼 수 있다[14]. 이러한 기술은 단순히 파일 명으로만 알기 힘든 콘텐츠를 인식하는 데 있어서 매우 유용하게 사용될 수 있다. 특징점 관리 서버에서는 특징점 코드를 받아 콘텐츠를 확인하고 제목, 저작자 명, 기타 정보를 다시 전송하게 된다. 따라서 이 파일은 불법콘텐츠라는 것을 알 수 있고, 이 경우 불법 사용자 정보를 수집하고 증거자료를 확보한다. 그리고 FP 코드 추출 모듈을 통해 FP 코드가 있는지 확인하고 만약 FP 코드가 존재하면 최초 배포자까지 검출이 가능하다.

불법콘텐츠 추적 시스템에서 수집한 정보들은 불법콘텐츠 수집 DB에 모두 저장되어 관리되고 FP 관리서버로 전송하게 된다. FP 관리서버는 최종적으로 불법 배포자를 판별하고 사이버 수사대에게 관련 정보들을 전달하고 수사 의뢰하게 된다. <표 4>에서는 FP 관리서버에게 보내주는 전송리스트을 보여주고 있으며 서버와의 통신 방식은 XML 데이터 구조로 인코딩되어 SOAP 프로토콜을 사용한다.

〈표 4〉 핑거프린팅 관리서버 전송리스트

구분	내용
콘텐츠	이미지: jpg, gif, bmp 등 오디오: mp3, ogg, wav, wma 등 비디오: avi, mpg, wmv, asf 등
사용자 정보	웹사이트 주소 또는 서비스 명, 사용자 IP, 사용자 ID, 최초 업로드 ID 등
콘텐츠 수집 메타 정보	파일 명, 파일 위치, 링크 명, 링크설명, 검색 시간, 파일 타입, 콘텐츠 사이즈, 화면 캡처 이미지 등
특징점 정보	콘텐츠 제목, 저작자 명, 기타 정보
핑거프린팅 코드 정보	핑거프린팅 코드

V. 결론

국내외 불법콘텐츠 시장현황과 검색 및 추적을 위한 업체 기술 동향을 알아보았고 불법콘텐츠 추적

시스템 아키텍처에 대하여 살펴보았다. 다양한 P2P/웹하드에서 불법콘텐츠를 다운로드하고 사용자정보를 수집하는 방법과 불법콘텐츠 추적의 핵심 기술로 핑거프린팅과 특징점을 이용한 식별 기술의 역할에 대해서도 알아보았다.

전세계적으로 디지털 콘텐츠 저작권 보호를 위해 워터마킹(WM), 핑거프린팅(FP), 디지털 저작권 관리(DRM), 디지털콘텐츠 식별시스템(DOI) 등의 기술들이 활발하게 연구중에 있으며, 이러한 기술들이 상용화되어 디지털 문화에 정착된다면 불법복제로 인한 피해가 상당히 줄어들 것으로 예상된다. 네티즌들의 저작권 침해에 대한 인식이 바뀌지 않는 한 불법복제의 완전한 근절은 어려울 것이지만 이러한 불법콘텐츠 추적 시스템이 성공적으로 개발되어 상용화된다면 저작권 침해로 난관을 겪고 있는 디지털 콘텐츠 산업에 활기를 불어넣을 수 있을 것으로 기대한다.

약어 정리

DRM	Digital Right Management
WM	Watermarking
FP	Fingerprinting
P2P	Peer To Peer
CPS	Copyright Protection System
DOI	Digital Object Identifier
MFTP	Multi-Source File Transmission Protocol
IFPI	International Federation for the Photographic Industry
MPA	Motion Picture Association
USTR	United States Trade Representive
PWL	Priority Watch List
ISP	Internet Service Provider
SOAP	Simple Object Access Protocol

참 고 문 헌

[1] 정보통신윤리위원회, http://www.icec.or.kr/

[2] 노프리, http://www.nofree.co.kr

[3] 지란지교소프트, http://www.jiran.com/

[4] Philips, http://www.philips.com/index.htm

[5] Digimark, http://www.digimark.com/

[6] BayTSP, http://www.baytsp.com/

[7] I. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Trans. Image Processing*, Vol.6, Dec. 1997, pp.1673-1687.

[8] D. Boneh and J. Shaw, "Collusion-secure Fingerprinting for Digital Data," *In Advances in Cryptology - CRYPTO'95,* Lecture Notes in Computer Science, Vol.963, Springer-Verlag, Heidelberg, New York, 1995, pp.453-465.

[9] A. Fiat and T. Tassa,, "Dynamic Traitor Tracing," *J. Cryptol.*, Vol.14, No.3, 2001, pp.211-223.

[10] R. Safavi-Naini and Y. Wang, "Sequential Traitor Tracing," *IEEE Transactions on Information Theory*, Vol.49, No.5, May 2003.

[11] Macro Express, http://www.macros.com/

[12] 김명섭, 강훈정, 홍원기, "Flow Grouping을 통한 P2P 트래픽 분석 방법에 관한 연구," *Proc. of KNOM 2003 Conference*, Daejeon, Korea, May 22-23, 2003, pp.210-218.

[13] P2P 사용자모임, http://p2p.pe.kr/

[14] Job Oostveen, Ton Kalker, and Jaap Haitsma, "Feature Extraction and a Database Strategy for Video Fingerprinting," *VISUAL 2002, LNCS 2314*, 2002, pp.117-128.