

2020 춘계학술 발표대회

[논문집]

2020년 5월 22일(금) ~ 23일(토)

장소: 오션스위츠 제주호텔



주최 사단법인 한국인터넷정보학회 주관 제주대학교 SW중심대학사업단

후원 (주)카카오, (주)LGCNS, (주)아이티센, (주)콤텍정보통신, (주)한컴위드, 도서출판 길벗, (주)넷케이티아이, LG히다찌(주)

B-4 ☆ 건설의 객관적 평가 방법 및 시스템 개발	29
문초이, 유현중, 이지원, 김은빈, 김수복, 계슬아, 위예진(순천향대), 백유상(고려대), 이언석(순천향대)	
B-5 2차원 이미지(깊이 카메라)를 통한 실시간 환경에서 물체의 부피를 구하는 알고리즘 연구	31
배성현(성균관대)	
B-6 자동차 번호판 검출을 위한 인조 번호판 생성 시스템	33
김상준, 이유진, 박구만(서울과기대)	
B-7 요구 공학을 이용한 DAQ system의 비기능적 요구사항 추출	35
김경식, 이석원(아주대)	

C. 인터넷보안 (좌장 : 강지원 교수 (세종대))

C-1 안드로이드 디바이스 자원 사용량 분석을 이용한 크립토재킹 탐지 기술	41
이희승, 곽진(아주대)	
C-2 비정형 시나리오 기반 원자력시설의 사이버보안 훈련 평가 체계 강화 방법론 제안	43
김상우, 류진호(한국원자력통제기술원)	
C-3 CLAPP 기반 안드로이드 악성코드 분류 기법	45
정해선, 김득훈, 곽진(아주대)	
C-4 Just Walk Out 서비스에서의 안전한 소비자 그룹화 방법	47
이민경, 곽진(아주대)	
C-5 웹사이트 광고 배너 분석을 이용한 웹툰 침해 의심 사이트 탐지 기술	49
김의진, 최슬기, 곽진(아주대)	
C-6 지멘스 S7 PLC 취약점을 이용한 공격 방법 연구	51
김준원, 김장훈, 이민규, 서정택(순천향대)	
C-7 PLC 제어로직 변조를 통한 DoS 공격 및 대응방안 연구	53
김장훈, 김준원, 정다운, 서정택(순천향대)	

웹사이트 광고 배너 분석을 이용한 웹툰 침해 의심 사이트 탐지 기술

김의진*, 최슬기**, 박진***

****아주대학교 사이버보안학과

**아주대학교 정보통신연구소

e-mail : *dmlws0403@ajou.ac.kr, **skchoi.isaa@gmail.com, ***security@ajou.ac.kr

Detection Technique of Suspected Webtoon Infringement Site Using Web Site Advertisement Banners

* Eui-Jin Kim*, Seul-Ki Choi**, Jin Kwak***

****Department of Cyber Security, Ajou University

**Institute for Information and Communication, Ajou University

요 약

최근 웹툰 시장 규모가 증가함에 따라 다양한 웹툰을 연재하는 사이트가 존재하며, 웹툰을 불법으로 배포하는 침해 사이트도 증가하고 있다. 웹툰 침해를 방지하기 위해서 침해 사이트를 판별해야 하며, 이를 위해 침해가 의심되는 사이트를 선별할 필요가 있다. 웹툰 침해 사이트는 사이트를 유지하기 위해 광고를 이용하는데, 이 중 광고 배너가 약 61%를 차지한다. 이러한 광고 배너는 정식 연재 사이트에서는 게시되지 않기 때문에 광고 배너 특정 분석을 통해 침해 의심 여부를 판단하고 사이트를 선별할 수 있다. 따라서 본 논문에서는 광고 배너의 특징을 분석하기 위해 Google Vision API를 이용하여 웹툰 침해 의심 사이트를 탐지하는 기술을 제안한다.

1. 서 론

최근 세계적으로 한국의 문화가 확산되면서, 웹툰 기반 한국 문화 사업이 해외 시장으로 진출하고 있다. 웹툰을 원작으로 한 K-드라마, K-영화의 해외 시장으로 진출과 함께 한국 웹툰에 대한 인기가 높아지고 있으며, 웹툰 시장의 규모가 증가하고 있다[1]. 그러나, 웹툰 시장이 증가함에 따라 웹툰 불법 복제로 인한 피해도 증가하였다. 2017년 기준 웹툰 불법 복제 유통 서비스로 인한 피해 규모는 약 9,939억 원이며, 이를 막기 위해 민·관 협력의 노력으로 단속이 이뤄지고 있지만, 웹툰 침해 사이트는 지속해서 등장하고 있다[2]. 따라서 웹툰 침해 사이트를 판단하여 피해를 방지하기 위해, 침해 의심 사이트를 선별할 필요가 있다.

웹툰 침해 사이트는 저작물을 불법으로 무료 배포하지만, 서버가 무료로 운영되는 것이 아니기 때문에 수익 창출을 필요로 한다. 웹툰 침해 사이트에서 수익 창출의 대표적인 수단은 광고 배너이다. 음란물, 도박과 같은 불법 사업자의 경우, 광고 철회에 대한 강제성이 존재하지 않기 때문에 지속해서 광고 배너를 노출하지만, 합법적인 사업자의 경우 기업의 이미지를 위해서라도 대부분 음란물, 도박 등의 광고를 진행하지 않는다. 또한, 웹툰 침해 사이트에서 이용되는 불법 광고 배너는 프레임 링크가 적용되지만, 정적 페이지 소스 분석을 통해서만 광고 배너 내부 정보를 분석하기 어렵다[3].

따라서 Google Vision API를 이용하여 광고 배너 내부 정보를 분석한다. 본 논문에서는 2장 관련 연구에서 Google Vision API, 침해 사이트 내 광고 배너에 대하여 설명하고, 3장에서 제안 사항을 설명한 뒤 4장에서 실험 결과를 설명하고, 5장에서 결론을 맺는다.

2. 관련 연구

2.1 Google Vision API

Google Vision API는 선행 학습된 머신러닝 모델을 제공하며, 이미지에 라벨을 할당하고 사전 정의된 수백만 개의 카테고리도 빠르게 분류할 수 있다. Vision API는 객체 감지, 얼굴 감지, 라벨 감지, 로고 식별, 이미지 속성 감지, 유해성 콘텐츠 감지 기능 등을 포함하고 있으며, Vision API의 OCR(Optical Character Recognition) 기능은 다양한 파일 형식과 이미지 내의 50개가 넘는 언어로 된 텍스트를 감지하고 추출할 수 있다[4].

2.2 침해 사이트 내 광고 배너

침해 사이트의 수익 창출은 주로 광고를 통해 이루어진다. 광고형태는 메인페이지에 AD 광고와 광고 배너를 게시하는 형태로 수익을 창출하는데, 침해 사이트에서의 광고형태는 광고 배너가 2018년 기준 약 61%를 차지한다. 광고 배너는 페이지 내의 위치와 기간을 설정하여, 그에 따른 비용을 지불하는 특징을 갖고 있다. 이와 같은 특징 때문에, 침해 사이트의 특정 프레임 내에 여러 개의 광고 배너가 게시된다. 또한, 광고 내 텍스트 영역과 이미지 영역의 경계가 모호하여 텍스트 추출이 어려우며, 이러한 광고 배너는 애드블록에도 탐지가 되지 않는 경우가 있다. 침해 사이트의 일일 방문자 수, 일일 페이지뷰, 사이트 내 위치에 따라 가격이 다르며, 이러한 광고 배너의 종류로는 도박, 성인용품, 만남 및 영상 등 다양한 모습을 나타내고 있다[3].

3. 제안 사항

본 장에서는 침해 의심 사이트를 판단하는 방법을 제안하고자 하며, 제안 방법은 그림 1과 같다. 해당 기법은 페이지 소스 분석으로는 알 수 없는 광고 배너 내부 정보를 분석할 수 있으며, 동작 과정은 다음과 같다.

