

# 2020

한국정보보호학회 하계학술대회

# CISC- S'20

Conference on Information Security  
and Cryptography - Summer 2020

## 07.15<sup>WED</sup>

온라인 컨퍼런스

<http://www.cisc.or.kr/>

주최  
주관



한국정보보호학회  
Korea Institute of Information Security & Cryptology

후원



국가정보원  
National Intelligence Service Korea



과학기술정보통신부



행정안전부



한국인터넷진흥원  
KISA KOREA INTERNET & SECURITY AGENCY

ETRI

한국전자통신연구원  
Electronics and Telecommunications  
Research Institute

NSR

국가보안기술연구소  
National Security Research Institute



아이티센  
ITCen



|                                              |                       |
|----------------------------------------------|-----------------------|
| 적대적 공격에 견고한 Perceptual Ad-Blocking 디자인 ..... | 김민재 · 김보민 · 허준범 / 381 |
|----------------------------------------------|-----------------------|

## AI보안

|                                               |                                   |
|-----------------------------------------------|-----------------------------------|
| DNN 모델 취약점 적대적 예제 생성 기법과 방어기법에 대한 분석 .....    | 서성관 · 문현준 · 윤주범 / 387             |
| 딥러닝 시스템 커버리지 기반 퍼징 프레임워크 최적화 연구 .....         | 유지현 · 문현준 · 윤주범 / 392             |
| Variant AutoEncoder 기반 모바일 맬웨어 클러스터링 기법 ..... | 고은별 · 김진성 · 반영훈 · 이정현 / 396       |
| 딥러닝을 활용한 CCTV 내 폭력 탐지 시스템 개발 .....            | 김태옥 · 박현아 · 임다연 · 최은정 / 400       |
| 머신러닝 기반 네트워크 트래픽 분석 기법 성능 비교 .....            | 황종배 · 하재철 / 404                   |
| 금융데이터를 활용한 연합학습 평가 .....                      | 장진혁 · 최대선 / 409                   |
| CNN을 이용한 스마트 폰 부채널 공격 .....                   | 김동현 · 김지은 · 유건우 · 이승대 · 백유진 / 413 |
| K-means Clustering을 이용한 kNN 속도개선에 관한 연구 ..... | 김도연 · 이용수 · 윤재현 · 이태진 / 417       |

## 신원/인증시스템

|                                           |                                         |
|-------------------------------------------|-----------------------------------------|
| 분산 ID 기반 비대면 데이터 공유 접근 시스템 .....          | 이정륜 · 윤태연 · 이우기 / 423                   |
| 차세대 주민등록정보시스템 구축에 관한 연구 .....             | 이현철 · 김대영 · 김현수 / 427                   |
| 탈 중앙화된 분산 ID 시스템의 특징 및 한계에 대한 동향 조사 ..... | 이길희 · 김석현 · 김은수 · 전소원 · 김봉곤 · 조영섭 / 431 |
| 열 기반 슬라이싱을 이용한 폐기 가능한 홍채 템플릿 생성 방법 .....  | 신영아 · 정재열 · 정익래 / 436                   |
| 안면인식을 통한 USB 파일 복호화 시스템 .....             | 송경주 · 강혜진 · 백민진 · 김현지 · 서화정 / 440       |
| 스마트폰과 스마트워치를 활용한 2차 사용자 인증 방식 제안 .....    | 심민주 · 엄시우 · 김현지 · 서화정 / 444             |

## 암호화폐/양자암호

|                                              |                             |
|----------------------------------------------|-----------------------------|
| 암호화폐 금전가치 분석 동향 .....                        | 박재훈 · 최승주 · 서화정 / 451       |
| 암호화폐 거래 믹싱을 통한 익명화기술 동향조사 .....              | 장설아 · 이경현 / 455             |
| 스마트 컨트랙트의 동적 모니터링을 지원하는 이더리움 가상머신 .....      | 김지수 · 이수연 · 조은선 / 459       |
| NIST 양자내성암호 공모전 부호 기반 암호 구현 동향 .....         | 최승주 · 김현지 · 심민주 · 서화정 / 463 |
| 양자 컴퓨팅 환경에서 안전한 다중 서명 기법 .....               | 이영경 · 박종환 · 이광수 · 이동훈 / 467 |
| $\mu$ -Hope: 오류 정정 부호를 사용한 RLWE 기반 KEM ..... | 이주엽 · 김수리 · 홍석희 / 471       |

## 데이터식별/추적

|                                                 |                             |
|-------------------------------------------------|-----------------------------|
| 저작권 침해 의심 사이트의 생애주기 단계 자동 탐지 기법 .....           | 정해선 · 곽진 / 477              |
| 링크분석 기반 영상 저작권 침해 스트리밍 사이트 탐지 방안 .....          | 이희승 · 곽진 / 481              |
| 스팸 리뷰 판별을 위한 텍스트 분석 관련 기술 동향 .....              | 임세진 · 김현지 · 서화정 / 485       |
| 음란물 유포 경로 추적을 위한 Tumblr 애플리케이션 메모리 포렌식 분석 ..... | 박진성 · 공성현 · 이창훈 / 489       |
| 안드로이드 어플리케이션의 DRM 기술 현황 분석 :                    |                             |
| 카카오톡, 카카오톡 .....                                | 이나연 · 김도연 · 양승욱 · 김희석 / 493 |
| 메시지 시그니처 분석을 통한 송신자 식별 방안 .....                 | 김동현 · 곽진 / 497              |

# 저작권 침해 의심 사이트의 생애주기 단계 자동 탐지 기법

정해선<sup>1</sup>, 곽진<sup>2</sup>

<sup>1</sup>아주대학교 컴퓨터공학과 정보보호응용및보증연구실

<sup>2</sup>아주대학교 사이버보안학과

## Automatic Detection Method for Life Cycle Step of Suspected Copyright Infringement Site

Hae-Seon Jeong<sup>1</sup>, Jin Kwak<sup>2</sup>

<sup>1</sup>ISAA Lab., Department of Computer Engineering, Ajou University

<sup>2</sup>Department of Cyber Security, Ajou University

### 요 약

저작권 침해 의심 사이트의 시간 흐름에 따른 대응 단계를 정의한 생애주기를 기반으로, 불법 행위를 수행하는 저작권 침해 의심 사이트에 대응할 수 있다. 그러나 다수의 저작권 침해 의심 사이트에 대하여 수동으로 생애주기 단계를 탐지하는 것은 비효율적이며, 현재 생애주기의 각 단계를 자동으로 탐지하기 위한 기법에 대한 연구는 미비한 실정이다. 이에 따라 본 논문에서는 저작권 침해 의심 사이트 모니터링을 효율적으로 수행할 수 있도록 생애주기 단계 자동 탐지 기법을 제안한다.

## I. 서론

저작권 침해 의심 사이트는 저작물을 사용자에게 불법적으로 무료 제공함에 따라 높은 수요가 있으며, 이로 인해 저작권 피해 규모 또한 지속적으로 증가하고 있다. 따라서 콘텐츠 생성자들의 저작권을 보호하기 위해 침해 의심 사이트를 모니터링하며 대응해야 한다.

이때 저작권 침해 의심 사이트의 생애주기 모델은 시간 흐름에 따른 저작권 침해 의심 사이트의 대응 단계를 정의하고, 생애주기 단계를 예측함으로써 대응할 수 있도록 한다. 하지만 저작권 침해 의심 사이트의 생애주기가 짧아지고, 사이트 수가 증가함에 따라 수동으로 생애주기 모델을 모니터링하는 것에는 한계가

있다. 따라서 본 논문에서는 저작권 침해 의심 사이트의 생애주기 단계를 효율적으로 모니터링하기 위한 자동 탐지 기법을 제안하였다. 2장 관련 연구에서 생애주기 단계와 이를 자동 탐지하기 위해 필요한 관련 개념을 설명하고, 3장 제안 기법에서 자동 탐지 기법을 제안한다. 그리고 4장 실험 결과에서 제안 기법의 실험 결과를 분석하며 5장에서 결론을 맺는다.

## II. 관련 연구

### 2.1 저작권 침해 의심 사이트의 생애주기

생애주기는 생성, 운영, 대응, 변경, 폐쇄의 5개 단계로 구성되어 있다[1].

#### □ 생성 단계

도메인 등록을 통해 저작권 침해 의심 사이트가 최초 생성된 경우이다.

본 연구는 문화체육관광부 및 한국저작권위원회의 2020년도 저작권연구개발사업의 연구결과로 수행되었음(2019-PF-9500).

#### □ 운영 단계

우회하지 않아도 최신 도메인에 접속 가능한 경우이다.

#### □ 대응 단계

최신 도메인에 대하여 정상 접속은 차단되지만 우회 접속은 가능할 경우, 또는 이전 도메인에 대하여 우회 접속만 가능하며 다른 도메인으로 리다이렉트되지 않는 경우이다.

#### □ 변경 단계

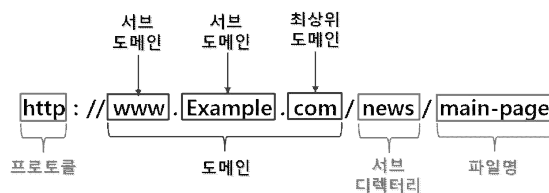
최신 도메인에 대하여 정상 접속이 가능하지만 다른 도메인으로 리다이렉트될 경우, 또는 이전 도메인에 대하여 정상 접속이 불가능하지만 우회 접속은 가능하며 다른 도메인으로 리다이렉트될 경우이다.

#### □ 폐쇄 단계

모든 도메인에 대하여 정상 접속 및 우회 접속 모두 불가능한 경우이다.

### 2.2 도메인 구조

저작권 침해 의심 사이트는 신규 도메인을 생성하며 사이트 차단에 대응하는데, 이때 도메인의 구조적 특성을 이용한다. URL은 (Fig. 1)과 같이 프로토콜, 도메인, 디렉터리, 파일명으로 구성되어 있고, 프로토콜 다음에 도메인이 위치한다. 도메인 내에서 구분자 “.”를 기준으로 오른쪽은 최상위 도메인이고, 왼쪽에 있는 것이 하위의 서브 도메인에 해당하게 된다. (Fig. 1)을 기준으로, “www”는 “Example”의 왼쪽에 위치하므로, “www”가 “Example”의 서브 도메인이다. 저작권 침해 의심 사이트는 서브 도메인 또는 최상위 도메인을 변경하며 신규 도메인을 생성한다[2].



(Fig. 1) Structure of URL and Domain

### 2.3 응답 코드

응답 코드는 길이 3의 정수로, 응답 코드를 통해 서버로부터의 응답을 파악하고 사이트의 생애주기 단계를 유추할 수 있다. 응답 코드의 첫 번째 정수는 <Table 1>과 같이 응답 유형을 정의하고, 뒤의 2개 정수가 응답 코드를 세부 정의한다[3].

<Table 1> Status Code by First Digit

| Status Code | Description            |
|-------------|------------------------|
| 1**         | Informational Response |
| 2**         | Successful             |
| 3**         | Redirection            |
| 4**         | Client Error           |
| 5**         | Server Error           |

## III. 제안 기법

생애주기는 하나의 사이트의 시간 흐름에 따른 대응 단계이므로, 신규 도메인이 속한 사이트를 탐지하여야 도메인이 속한 사이트의 시간 흐름을 고려한 생애주기를 탐지할 수 있다. 따라서 사이트 탐지 기법과 생애주기 단계 자동 탐지 기법을 함께 제안한다.

### 3.1 사이트 탐지

사이트별 (“최신 도메인”, “도메인 변경 기록”, “최신 도메인의 HTML 태그 순서”)를 수집하는 “사이트 DB”를 활용하여 다음 과정으로 신규 도메인의 사이트를 탐지한다.

#### Step 1. “키워드 도메인” 비교

저작권 침해 의심 사이트는 다양한 방식으로 신규 도메인을 생성한다. 하지만 대부분 해당 사이트를 식별할 수 있는 “키워드 도메인”에 다양한 규칙을 적용하며 신규 도메인을 생성한다. 서브 도메인이 1개일 경우, 서브 도메인이 “키워드 도메인”이며, “키워드 도메인” 끝에 숫자를 추가하거나 최상위 도메인을 변경하거나 새로운 서브 도메인을 도메인의 오른쪽에 추가하며 업데이트한다. 서브 도메인이 2개일 경우, “키워드 도메인”은 오른쪽 서브 도메인에 해당하며, 왼쪽의 서브

도메인을 변경 또는 삭제하거나, 키워드 도메인의 끝에 숫자를 추가하거나, 최상위 도메인을 변경하며 도메인을 업데이트한다.

따라서 신규 도메인의 서브 도메인이 2개일 경우에는 오른쪽 서브 도메인, 1개일 경우는 서브 도메인에 대하여 정규식을 매칭하여 마지막에 숫자가 오는 경우 이를 제외하고 “키워드 도메인”을 생성한다. 해당 “키워드 도메인”이 “사이트 DB” 내 도메인들의 “키워드 도메인”과 일치할 경우 기존 사이트의 최신 도메인으로 탐지한다. 동일한 “키워드 도메인”이 없을 경우 다음 단계를 진행한다.

### Step 2. HTML 태그 순서 비교

신규 “키워드 도메인”으로 신규 도메인을 생성할 경우 Step 1으로 탐지할 수 없기 때문에 HTML 태그 순서를 기반으로 사이트의 신규 여부를 식별한다. HTML 태그 순서는 사이트의 기본 구조이므로, 사이트 탐지 기준으로 사용할 수 있다. 파이썬의 Html\_similarity 라이브러리를 사용하여 신규 도메인과 “도메인 DB” 내 사이트 사이 HTML 태그 순서 유사도를 생성한 결과, 90% 이상일 경우 기존 사이트의 최신 도메인으로 탐지한다. 만약 모든 사이트와 HTML 태그 순서 유사도가 90% 미만일 경우에는 신규 사이트로 탐지한다.

## 3.2 생애주기 단계 자동 탐지

### Step 1. 우회 접속 필요 여부 확인

모니터링을 우회하기 위해 국내 사용자들의 접속을 차단할 경우 우회 접속이 필요하며 파이썬 Requests 라이브러리에서는 접속 요청 시 접속이 차단될 경우 Connection Error가 발생한다. 따라서 (Fig. 2)의 예외처리를 통해 우회 접속 필요 사이트 목록을 생성한다.

```
def get_page(url):
    try:
        # 접속 시 우회가 필요한 사이트
        res = requests.get(url, Verify=False)

        # 접속 시 우회가 필요한 사이트
    except requests.ConnectionError as e:

        # 접속 시 우회가 필요한 사이트 목록 생성
        blocked_urls.append(url)
        pass
```

(Fig. 2) Pseudo Code for Exception Handling

### Step 2. 응답 코드 확인

응답 코드 404는 해당 도메인의 서버가 존재하지 않음을 의미하므로, 404 응답 코드를 수신할 경우 “폐쇄” 단계라고 탐지한다. 500 응답 코드 또한 서버 에러로 인해 도메인 서버가 정상 운영되지 않으므로 “폐쇄” 단계라고 탐지한다. 추가적으로 403 응답 코드의 경우, 도메인의 사이트가 정상 운영되는 경우도 있지만 응답 페이지에 “Forbidden” 키워드가 포함된 경우에는 모든 사용자 접근을 차단하며 사이트를 운영하지 않는 경우로 “폐쇄” 단계로 탐지한다.

### Step 3. 응답 도메인 확인

접속 시 다른 도메인으로 리다이렉트될 경우 응답 도메인이 변경되므로, 파이썬의 Requests 라이브러리의 url 속성을 통해 리다이렉트 여부를 확인할 수 있다.

이때 방송통신심의위원회로부터 차단된 모든 사이트는 “http://warning.or.kr”으로 리다이렉트되기 때문에 응답 도메인이 이에 해당하면 차단된 것으로 판단한다. 차단된 경우는 우회 접속 필요 사이트 목록에 해당 도메인을 추가한다.

응답 도메인과 입력 도메인이 같은 경우 생애주기의 “운영” 단계, 다른 경우는 “변경” 단계로 탐지한다. VPN 사용 후 응답 도메인과 입력 도메인이 같은 경우는 생애주기의 “대응” 단계, 다른 경우는 “변경” 단계로 탐지한다.

```
def check_redirect(res, url):
    # 방송통신심의위원회에서 차단한 경우
    if res.url == "http://warning.or.kr":
        # 접속 시 우회가 필요한 사이트
        blocked_domains.append(url)

    # 리다이렉트되지 않은 경우
    elif res.url == url:
        # VPN 우회 접속을 한 경우 -> 대응
        # VPN 우회 접속을 하지 않은 경우 -> 운영

    # 리다이렉트된 경우 -> 변경
    elif res.url != url:
```

(Fig. 3) Pseudo Code for Response Domain Check

**Step 4. VPN 작동**

Step 1, Step 3 수행 결과 우회 접속 필요 사이트 목록이 생성된다. 호스트 OS에서 VPN 애플리케이션을 작동한 후 우회 접속 필요 사이트에 대하여 Step 2와 Step 3을 수행하여 생애주기 단계를 식별한다. 이때 VPN 작동 후에도 Connection Error가 발생하는 경우는 생애주기의 “폐쇄” 단계로 탐지한다.

**IV. 실험 결과**

본 논문의 제안 기법의 실험 환경은 아래 <Table 2>와 같으며, 저작권보호원의 모니터링 대상인 총 282개 사이트를 대상으로 실험하였다.

&lt;Table 2&gt; Experimental Setup

| Component       | Description                                              |
|-----------------|----------------------------------------------------------|
| OS              | Windows 10 Pro 64-bit                                    |
| SW              | Python 3.8.2                                             |
| VPN Application | Touch VPN                                                |
| Dataset         | Sites Monitored by the Korea Copyright Protection Agency |

**4.1 사이트 탐지 결과**

폐쇄된 도메인에는 접속할 수 없어 HTML 태그 순서를 받아올 수 없기 때문에 “키워드 도메인” 비교 기반 사이트 탐지를 실험하였다. 전체 사이트의 도메인 변경 기록 총 358개 중 220개의 변경 기록이 속한 사이트를 탐지하였다. 이는 67%의 도메인에 대해서는 도메인만으로 동일 사이트를 식별하며 탐지하고, 그 외의 약 33%에 대해 Step 3를 수행하면 동일 사이트를 탐지할 수 있음을 의미한다.

**4.2 생애주기 단계 자동 탐지 결과**

생성 단계는 도메인 등록 순간이므로 그 외 4개 단계를 대상으로 생애주기 단계를 탐지하였다. 저작권 보호원의 282개 모니터링 대상 사이트에 대하여 <Table 3>과 같이 운영, 대응, 변경, 폐쇄 단계를 모두 탐지하였다.

&lt;Table 3&gt; Detection Result

| Lifecycle        | Number of Sites (Number) |
|------------------|--------------------------|
| Operation        | 123                      |
| Action           | 56                       |
| Change           | 42                       |
| Closed           | 61                       |
| Total / Detected | 282 / 282                |

**V. 결론**

본 논문에서는 저작권 침해 의심 사이트 탐지 및 생애주기 단계 자동 탐지 기법을 제안하였다. 본 논문이 제안하는 탐지 기법을 적용함으로써 최신화된 생애주기 정보를 효율적으로 수집하며, 저작권 침해에 대응할 수 있다.

향후에는 사이트별 HTML을 수집하여 HTML 태그 순서 기반 사이트 탐지에 대해 실험하고, 생애주기 단계에 대한 다양한 특징을 추가 도출하여 기존 데이터셋으로 발견하지 못한 경우에 대비하기 위한 연구를 진행할 예정이다. 또한 해당 제안 기법을 사용하여 수집한 생애주기 데이터를 가공 및 분석하여 생애주기 모델의 정확도와 예측률을 향상시킬 계획이다.

**[참고문헌]**

- [1] 김득훈, 정해선, 광진, “저작권 침해 사이트 생애 주기 모델 개발”, 한국정보보호학회 정보보호학회논문지, Vol. 30, No. 1, pp.101-121, Mar. 2020.
- [2] Huawei Cloud, “Domain Name Service Overview”, Issue. 14, Feb. 2020.
- [3] DELL, “REST API Programmer’s Guide”, Jun. 2019.