Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.28, NO.5, Oct. 2018 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2018.28.5.1179

웹 기반 디바이스 핑거프린팅을 이용한 온라인사기 및 어뷰징 탐지기술에 관한 연구

장 석 은,[1†] 박 순 태,[2] 이 상 준[1‡]
[1]전남대학교, [2]한국인터넷진흥원

A Study on Online Fraud and Abusing Detection Technology Using Web-Based Device Fingerprinting

Seok-eun Jang,[1†] Soon-tai Park,[2] Sang-joon Lee[1‡]
[1]Chonnam National University, [2]KISA

요 약

최근 PC, 태블릿, 스마트폰 등 다중 접속환경을 통하여 웹 서비스에 대한 다양한 공격이 발생하고 있다. 이런 공격은 웹 서비스의 취약점을 통해 온라인 사기거래, 계정의 탈취 및 도용, 부정로그인, 정보 유출 등 여러 가지 후속 피해를 발생시키고 있다. Fraud 공격을 위한 새로운 가짜 계정의 생성, 계정도용 및 다른 이용자 이름 또는 이메일 주소를 사용하면서 IP를 우회하는 방법 등은 비교적 쉬운 공격 방법임에도 불구하고 이런 공격을 탐지하고 차단하는 것은 쉽지 않다. 본 논문에서는 웹 기반의 디바이스 핑거프린팅을 이용하여 웹 서비스에 접근하는 디바이스를 식별하여 관리함으로써 온라인 사기거래 및 어뷰징을 탐지하는 방법에 대해 연구하였다. 특히 디바이스를 식별하고 이를 스코어링 하여 관리는 것을 제안하였다. 제안 방안의 타당성 확보를 위하여 적용 사례를 분석하였고, 온라인 사기의 적극적인 대응과 이용자 계정에 대한 가시성을 확보할 수 있어 다양한 공격에 효과적으로 방어할 수 있음을 증명하였다.

ABSTRACT

Recently, a variety of attacks on web services have been occurring through a multiple access environment such as PC, tablet, and smartphone. These attacks are causing various subsequent damages such as online fraud transactions, takeovers and theft of accounts, fraudulent logins, and information leakage through web service vulnerabilities. Creating a new fake account for Fraud attacks, hijacking accounts, and bypassing IP while using other usernames or email addresses is a relatively easy attack method, but it is not easy to detect and block these attacks. In this paper, we have studied a method to detect online fraud transaction and obsession by identifying and managing devices accessing web service using web-based device fingerprinting. In particular, it has been proposed to identify devices and to manage them by scoring process. In order to secure the validity of the proposed scheme, we analyzed the application cases and proved that they can effectively defend against various attacks because they actively cope with online fraud and obtain visibility of user accounts.

Keywords: Device Fingerprinting, Online Fraud, Abusing, Web Security, Web Access Control

Received(08. 01. 2018), Modified(09. 28. 2018), Accepted(09. 28. 2018)

† 주저자, dustinjang77@gmail.com
‡ 교신저자, s-lee@jnu.ac.kr(Corresponding author)

I. 서 론

온라인 웹 서비스는 웹 사이트 및 모바일 웹, 앱 기반의 다양한 서비스들에서 발생되는 주요 사이버 범죄행위가 있다. 전통적인 IT시스템인 네트워크, 시스템, 웹 애플리케이션에 발생되는 DDoS, 서버취약점을 이용한 다양한 공격과 웹 애플리케이션에서 부정인증, 인젝션(injection), 크로스 사이트 스크립팅(XSS), 웹셸 공격 등 다양한 공격이 존재한다. 비영리재단인 OWASP가 2017년에 발표한 'OWASP Top 10 - 2017' 보고서에 발표한 내용에 따르면, 취약한 접근 제어와 불충분한 로깅 및 모니터링 등 이용자 계정에 대한 관리이슈가 여전히 웹 애플리케이션의 취약한 부분으로 발표되었다[1].

최근에는 PC, 태블릿, 스마트폰 등 다양한 접속 환경 속에 안전하지 않은 웹 서비스 설계와 지속적인 해킹사고로 인한 정보의 유출을 통해 온라인 사기거래, 계정의 탈취 및 도용, 부정로그인 등 여러 가지 후속피해가 속출하고 있다. 방송통신위원회가 2018년 3월에 발표한 보도자료에 따르면, 최근 인터넷 상에서 실명 아이디뿐만 아니라 비실명 아이디(일명 '유령 아이디')를 대량 생산하여 전문적으로 판매하고, 이를 구매하여 인터넷쇼핑몰 등에서 상품이나 서비스를 거짓으로 평가·홍보하거나, 댓글을 이용한 검색 순위 및 여론조작, 불법도박과 성매매 등 각종 범죄에 악용하는 사례가 빈발하여 인터넷 상 아이디(계정) 불법거래를 집중적으로 단속하기로 하였다. 보도자료 에는 2017년 기준 개인정보 불법거래 게시물은 총 115,522건 탐지되었으며, 이 중 아이디 불법거래 게시물은 8,956건(전년대비 215% 증가)으로 약 8%에 해당된다고 하였다[2]. 2018년 8월에도 OO 마켓 계정이 중국의 최대 오픈마켓인 타오바오에서 거래가 되는 일이 발생되어 한국인터넷진흥원이 이에 대한 조사를 진행하였으며, 해당 전자상거래업체는 중국의 접속을 차단하고 게시물에 대한 삭제를 요청하는 일들이 발생하였다[5]. Lexis Nexis Risk Solutions의 보고서에 따르면 온자인 전자 상거래에서 사기 거래의 성공률은 2017년에 비해 30% 가량 증가했다. 대형 멀티 채널 판매자의 경우 36% 더 높다. 이는 핀테크와 같이 빠르고 편리한 거래에 대한 소비자의 요구가 모바일 및 온라인 전자 상거래 사기의 급격한 증가와 관련 비용으로 이어지고 있음을 보여준다. 설상가상으로 가맹점은 사기로 인한 손실 1 달러당 총 비용 2.94 달러만큼 문제를 해결하기 위해 더 많은 돈을 지불하고 있다. 이는 지불 거절 증가, 상품 재배송, 사기 수사, 법적 기소 및 사이버 보안 운영으로 2017년 대비 6 % 증가한 수치이다.[6]

Fraud 공격을 위한 새로운 가짜 계정의 생성, 계정도용 및 다른 이용자 이름 또는 이메일 주소를 사용하거나 IP를 우회 방법 등은 비교적 쉬운 공격 방법이기 때문에 이를 탐지하고 차단하는 것은 쉽지 않다. 대부분의 웹 서비스에서는 정상적인 접근경로로 탐색이 이루어 진다. 하지만, 공격자가 탐색 및 접근시마다 다른 디바이스를 사용하는 것은 어려운 일이기 때문에 비교적 쉽게 공격자의 디바이스를 특정하여 대응하는 방법론이 필요하다.

본 논문에서는 웹 트래픽을 측정하고 광고에 활용하기 위해 오랫동안 사용되어온 웹 분석 서비스 기술인 디바이스 핑거프린팅(device fingerprinting) 기법을 이용하여 웹 서비스에 접근하는 이용자의 디바이스에 대한 정보를 수집하고, 이를 이용하여 온라인 사기거래, 계정의 탈취 및 도용, 부정로그인 등을 탐지하고 예방하는 방법을 서술한다. 이는 웹 서비스에 접근 및 로그인을 시도하는 이용자의 계정 또는 쿠키(cookies)가 아니라 합법적인 이용자와 악의적인 이용자 접근을 감지하거나 차단, 혹은 인증을 강화하는 도구로 활용될 수 있다.

본 논문의 제 2장에서는 이상거래 탐지시스템과 온라인 행위를 트래킹하는 기술과 보안적인 측면에서의 디바이스 핑거프린팅 기술에 대하여 설명하고 3장에서는 디바이스 핑거프린팅을 기술을 이용하여 데이터를 획득하는 방법에 대해 설명한다. 4장에서는 디바이스 핑거프린팅을 이용하여 데이터를 수집해 보고 그 결과에 대하여 설명한다. 본 논문의 궁극적인 연구목표는 디바이스 핑거프린팅을 이용하여 웹 서비스와 이를 이용하는 이용자를 안전하게 보호할 수 있는 방안을 도출하는 것이다.

II. 관련연구

2.1 이상거래 탐지시스템

금융 산업을 중심으로 시작된 '이상거래 탐지시스템(FDS, Fraud Detect System)'에 대한 연구와 기술의 개발이 두드러지게 나타나고 있다. 2014년 이상금융거래 탐지시스템 기술 가이드와 2017년 이상행위 탐지시스템 기술의 발전 방향에 의하면 이상

거래 탐지시스템은 이용자 혹은 기업의 금전적 손실, 정보 유출 등 이상거래를 통해 발생될 수 있는 악의적인 거래들을 탐지하고 차단하기 위해 고안된 시스템으로 정의하고 있으며, 이와 유사하게는 이상거래 방지시스템(Fraud Prevent System), 위협 관리 시스템(Risk Management System) 등이 존재한다[3],[4]. 이러한 FDS에 대한 연구는 최근 게임 산업과 전자결제 등 다양한 분야에서 확산되는 추세이며 다양한 연구들이 나오고 있는 상태이다.

하지만 최근의 FDS에 대한 주요연구를 정리하면 주로 이상 거래에 대한 연구가 핵심이나 포털, SNS, 전자상거래, 커뮤니티 등 기본적인 웹 서비스 기업에 대한 온라인 사기나 어뷰징을 방지하기 위한 연구는 미진한 실정이다. Table 1은 최근 이루어진 FDS의 주요 연구를 산업군과 주요 기능에 대한 부분을 정리하였다.

Table 1. Main research technology list of Fraud Prevent System for Industry group based

Industry group	Main research
Banking	Billing log analysis and data mining[7] Preventing phishing scams[8] Prevention of illegal transfer accident[9] Improving the rules for detection of abnormality in e-banking transactions[10]
Game	Game payment data analysis [11]
Fintech	Data mining in the mobile payment environment [12] FinTech Transaction Detection Framework [13]

2.1.1 이상거래 탐지시스템의 정보수집 방법 및 특징

일반적으로 이상거래 탐지시스템에서 이용자의 디바이스 정보를 수집하기 위해 별도의 수집프로그램을 이용자 디바이스에 설치하여 정보를 수집하는 방법과, 웹 어플리케이션에서 제공하는 정보를 이용하여 정보를 수집 하는 방식으로 구분 할 수 있다.

Table 2에는 별도의 프로그램(agent)을 이용하여 이용자 시스템에 설치한 후 이용자의 디바이스 정보를 수집 하여 수집서버로 전송하여 동작하는 방식은 비교적 다양한 이용자의 디바이스 정보를 수집하고 전송되는 정보의 수정이나 회피를 위한 변조가 어려운 장점을 가지고 있으나 프로그램을 설치하지 않는 이용자나 다양한 OS, 브라우저 환경에 대한 대응이 어렵다. 웹기반을 이용한 디바이스 정보를 수집하는 경우 이용자 서버 접속정보를 기준으로 디바이스 정보를 수집하는 방식으로 동작되며, 수집되는 정보는 상대적으로 제한이 있으나 다양한 환경에 적용할 수 있으며, 이용자에게 별도의 추가적인 행동을 유발하지 않아 유연하게 적용할 수 있으나 변조가 용이한 단점을 가진다.

Table 2. Feature comparison of user device information collection method

	Advantages	Disadvantages
Agent based	• Various information can be collected • Difficult to manipulate information	• Not available for collection(not installed) • Not supports various system environment
Web based	• Supports various system environment • User convenience	• Easy to manipulate information • Limit information collection

2.2 Non-Agent 기반의 이용자 행위의 트래킹 기술

온라인상에서 별도의 프로그램을 설치하지 않고 이용자의 정보를 수집하여 활용하는 온라인 트래킹 기술이 빅 데이터(big data) 분석 기술을 활용하기 시작하면서 매우 빠르게 발전하고 있다. 다양한 이용자 정보가 서비스 제공 및 광고, 마케팅, 분석 등의 목적으로 방문사이트, 검색어 등 온라인상 이용자의 다양한 활동 정보들을 추적하고 저장 및 활용하고 있다[14]. 이러한 온라인 트래킹 기술은 이용자의 프라이버시(privacy)를 침해할 수 있다는 논란이 있지만, 보안적인 측면에서 이를 활용하여 이용자의 디바이스를 추적하고 관리할 수 있는 기술을 공익적, 온라인 안정성 제공 측면에서 사용할 필요성이 있다.

온라인 트래킹은 사업자가 이용자에게 더 나은 제품과 서비스를 제공하기 위하여 여러 기기에서 유·무선 네트워크를 통해 개인 정보를 수집·저장·공유하고, 이를 평가·분석하는 과정을 거치며, 궁극적으로

이용자의 관심사를 추론하고, 향후 행동을 예측하는 일련의 행위라 할 수 있다[14]. 일반적으로 광고의 대상을 명확하게 지정하기 위해 이용자의 활동에 대한 트래킹으로서 이용자의 검색, 웹 사이트 방문, 웹 콘텐츠 조회, 이메일 콘텐츠, 시청한 동영상, 소셜 네트워크에서의 상호 작용 및 온라인 거래 등을 포함할 수 있다. 이러한 온라인 트래킹 기술은 크게 '쿠키(cookies)'와 '웹비콘(web-beacons)', '디바이스 핑거프린팅(device fingerprinting)', '디바이스 ID 트래킹', '크로스 디바이스 트래킹(cross-device tracking)'등으로 분류할 수 있다[14],[15].

2.2.1 쿠키

쿠키란 웹 서비스가 브라우저에 전송하는 '상태정보'를 저장하기 위한 것으로 이용자가 어떤 웹 사이트를 방문하면 등록한 이름, 방문한 사이트, 열람한 페이지, 특정 사이트에 등록한 비밀번호 등 이용자가 입력한 각종 정보가 저장되는 작은 텍스트 파일로서, 이용자를 식별(identify)하거나 이용자의 디바이스를 인식(recognize)하려는 목적을 가지며, 해당 사이트를 다시 방문했을 때, 저장된 정보를 읽어냄으로써 이용자를 구분(distinguish)할 수 있도록 하는 정보이다[15]. 이러한 쿠키는 지속적으로 발전되고 있으며, 다양한 종류와 방식들의 쿠키 기술들이 등장하고 활용되고 있다.

웹 사이트가 디바이스에 이용행태에 대한 여러 가지 정보를 저장시키는 기술은 이용자의 다양한 정보를 수록하여 활용한다는 측면에서 중요한 부분이다. Table 3에서 정리한 것처럼 쿠키의 저장방식을 중심으로 쿠키 기술들을 간략하게 살펴보면, 크게 HTTP 쿠키(HTTP cookies), 플래시 쿠키(flash cookies), HTML5 스토리지 쿠키(HTML5 storage cookies)로 구분할 수 있다. 대부분의 쿠키는 웹 사이트 정보로서 사이트의 방문 횟수, 기본 설정, 로그인, 탐색 활동 기록, 보았던 페이지와 콘텐츠, 방문시기, 검색 한 내용, 클릭한 광고 등을 저장한다.

Table 3. Key features of HTTP cookies, Flash cookies, and HTML5 storage cookies

	HTTP Cookies	Flash Cookies	HTML5 Storage
size	4Kb	100Kb	5Mb
location	SQL	Non-browser location	SQL
Access	Accessible through a single browser	Accessible through various browsers	Accessible through a single browser

2.2.2 웹비콘

웹 비콘은 웹 버그(web bugs)라고도하며 웹 사이트에서 고객의 행동을 이해할 수 있도록 쿠키와 함께 사용되는 기술이다. 일반적으로 투명한 그래픽 이미지(보통 1픽셀 x 1픽셀)로서 웹 사이트나 전자 메일에 사용하면서 이용자가 일부 콘텐츠에 액세스했는지 여부를 눈에 띄지 않게 추적할 때 사용하는 기술이다. 일반적으로 웹 분석을 위한 전자 메일 추적 및 페이지 태깅이 사용된다[17],[18].

HTML 페이지가 다운로드 되면 브라우저가 그 페이지를 분석하고 화면이나 이미지 등을 표시한다. 이때 필요한 경우 추가 리소스를 찾게 되는데, 여기서 서버에 추적 요청을 생성하는 작은 이미지 파일을 이용하며, 다운받은 각 이미지에 대해서 브라우저가 백그라운드 환경에서 다른 요청을 하는 것을 추적하게 한다. 서버가 요청을 받을 때는 서버의 로드를 모니터링 하는 요청을 기록하고 요청한 이용자가 어디서 이동되어 왔는지에 대한 정보를 기록한다. 이 로그를 분석해서 이용자가 어디로 이동하고 언제 무엇을 했는지를 확인하게 된다.

2.2.3 디바이스 ID 트래킹

스마트폰과 태블릿 PC가 보편화됨에 따라, 멀티 디바이스 환경 하에서 개별 이용자를 모두 식별하는 것은 상당히 어려워졌다. 따라서 사업자들은 기기에 부여된 고유한 ID를 통해 이용자가 아닌 디바이스를 식별하고자 하였다. 디바이스 ID는 전 세계의 모든 스마트 폰 또는 태블릿 PC를 식별하는 숫자 및 문자로 구성된 문자열이다. 디바이스 ID는 모바일 장치에 저장되며, 다운로드 되어 설치된 모든 응용 프로그램에서 검색할 수 있게 된다. 모바일 앱은 일반적으로 디바이스 ID를 검색하여 서버와 통신 할 때 디바이스를 확인하기 위하여 ID를 사용하게 된다. 모바일 광고의 맥락에서 디바이스 ID는 광고업자나 마케팅 담당자 등이 특정한 유형의 디바이스를 추적

할 때 이용된다.

애플(apple)의 iOS를 사용하는 경우 iPhone, iPod touch 및 iPad에는 고유한 장치 ID(UDID, Unique Device ID)라고 하는 고유한 장치 ID 번호가 있다. Apple의 UDID는 문자와 숫자의 40 자리 시퀀스이다. Android 디바이스 ID는 디바이스를 처음 부팅 할 때 결정하게 되는데, 이용자는 Google Play에서 무료 앱을 다운로드하여 기기에서 임의로 생성 한 Android ID에 액세스 할 수 있다. 이러한 장치 식별 번호는 공장 초기화(factory reset)이 수행되지 않는 한 장치의 수명 동안 일정하게 유지된다. 이러한 디바이스의 고유한 식별번호가 개인정보 문제로 확산되어짐에 따라 애플은 'IDFA(Identity For Advertisers)'라는 식별자를 이용하게 하였고, Google도 Android 디바이스에서 GAID(Google Advertising ID)'를 사용하게 하고 있다[20].

2.3 디바이스 핑거프린팅

디바이스 핑거프린팅은 웹 사이트를 보기 위해 사용하는 디바이스의 속성정보, 네트워크정보, OS정보, 브라우저의 구성이나 설정 등을 기반으로 시간 경과에 따라 장치를 트래킹하는 방법으로, 쿠키를 사용하지 않고도 이용자를 식별할 수 있는 방법이다. 디바이스와 웹 사이트를 운영하는 서버가 통신을 하는 과정에서 전송되는 기본정보에는 브라우저의 종류, PC의 운영체제(OS), 쿠키 사용여부(브라우저 설정값) 정보가 이용되며, 최근 웹 사이트의 다양한 기능을 구현하기 위해 사용하는 어도비 플래시(Adobe Flash), 자바 가상머신(Java Virtual Machine)을 통해 브라우저 기능을 지원하는 프로그램의 버전, PC의 표준시, 화면 해상도, PC의 보유 글꼴 및 색상, TCP/IP 등의 정보도 수집한다. 이처럼 디바이스가 서버와 통신하는 과정에서 발생하는 정보들이 모아져 디바이스 핑거프린팅이 된다[3].

디바이스 핑거프린팅 정보는 쿠키차단이나 IP주소를 숨기는 프록시를 사용하더라도 수집할 수 있으며, 변경하려면 운영체제 및 소프트웨어의 정보를 변경해야 한다. 전자 프런티어 재단(EFF, Electronic Frontier Foundation)의 연구 결과, 쿠키 없이도 브라우저 핑거프린팅을 통해 웹 사이트 방문자 94%의 정보 수집이 가능한 것으로 나타났다. Panopticlick 사이트[1]를 방문한 참가자들의 브라우저 47만 161개에서 핑거프린팅을 수집한 결과, 어도비 플래시나 자바 가상머신을 사용하는 경우는 방문자의 94.2%에서 브라우저 핑거프린팅을 할 수 있었다. 이처럼 쿠키 설정을 해제하고 IP 주소를 숨기는 프록시를 사용해도 브라우저 핑거프린팅을 이용해 이용자의 정보는 트래킹 될 수 있다[14],[15],[19].

2.3.1 크로스 디바이스 트래킹

크로스 디바이스 트래킹이란 스마트 폰, 태블릿, 스마트 TV 및 PC와 같은 여러 장치에서 이용자를 추적 할 수 있는 기술이다. 일반적으로 이용자는 다양한 디바이스를 이용하여 웹 서비스를 이용하고 있다. 동일한 PC에서도 여러 가지 브라우저를 사용하는 경우에도 이용자를 식별하고 트래킹 할 수 있는 기술을 의미한다[21]. 최근 모바일 사용량이 급격하게 늘어나면서 이른바 멀티스크린(multi-screen)의 시대가 도래하게 되면서, 여러 디바이스 상에서 이용자들의 움직임을 포착하고 정확한 광고에 매칭시키는 것이 매우 중요한 기술로 떠오르게 되었다.

사업자는 이용자에 대한 정보를 그들의 스마트폰, 태블릿, PC, 스마트 TV, 웨어러블 디바이스 등과 같이 서로 연결된 디바이스를 통해 수집하고 이러한 정보를 이용자의 오프라인에서의 습관 및 행태와 결합시키고자 한다. 단순한 방법으로 디바이스 핑거프린팅 정보에 이용자 ID나 IP주소를 매핑시켜 추적하는 기술부터 각각의 디바이스에서 수많은 데이터 포인트를 수집하고, 디바이스 간의 연결성을 식별하기 위하여 복잡한 알고리즘을 이용하기도 한다[14],[21].

2.4 현재 웹 사이트의 보안기술 적용 현황

현재까지의 웹 서비스를 제공하는 많은 웹 사이트들은 보편적인 보안시스템인 방화벽(firewall)과 침입탐지시스템(IPS)등을 이용한 IP기반의 차단이 핵심 이였으며, 추가적으로 브라우저의 헤더(header) 정보와 웹 엑세스 로그(web access log)형태를 파싱(parsing)하여 모니터링하고 통계를 하는 형태로 진행되고 있다.

보다 정교한 탐지를 위하여 쿠키를 활용하여 고유한 ID를 만들어 사용하는 웹 사이트 도 많지만, 웹

1) https://panopticlick.eff.org

서비스에 접근하는 이용자의 디바이스에 대한 정보를 수집하고, 이를 이용하여 온라인 사기거래, 계정의 탈취 및 도용, 부정로그인 등을 탐지하고 예방하는 방법을 적용하는 웹 서비스는 찾아보기 힘든 상황이다. 실제 악성 봇(bot)이나 크롤러(crawler)등은 지속적으로 IP를 변환하거나 쿠키값을 리셋(reset)하여 공격하는 형태를 지니고 있는데, 이러한 기술을 현실적으로 사전에 탐지하고 차단하기는 어려운 상황이다.

2.5 보안영역의 디바이스 핑거프린팅 기술

디바이스 핑거프린팅 기술은 앞서 설명 하였듯이 다양한 이용자 정보가 서비스 제공 및 광고, 마케팅, 분석 등의 목적으로 활용되고 있다. 이러한 기술은 이용자의 개인정보를 침해하고 개인정보의 자기통제권을 위협 할 수 있다. 하지만 보안적인 측면에서는 이러한 기술들을 활용하여 이용자의 디바이스를 추적하고 관리하여 온라인에서의 악성행위를 사전에 차단하거나 추가적인 인증수단으로서 활용하여 온라인 안정성에 기여 할 수 있을 것이다.

2.5.1 디바이스 핑거프린팅 기술구현 현황

디바이스 핑거프린팅 정보는 쿠키차단이나 IP주소를 숨기는 프록시를 사용하더라도 수집할 수 있으며, 변경하려면 운영체제 및 소프트웨어의 정보를 변경해야 하는 강력한 수집기술 구현이 가능하다. 최근에는 웹 사이트에서 온라인 트래킹을 방지하기 위한 안티 트래킹(anti tracking) 기술과 토르(tor)넷이나 토르 브라우저를 이용하는 대안이 제시되고 있지만, 브라우저의 특성뿐만 아니라 디바이스의 고유정보들을 조합하여 관리하는 디바이스 핑거프린트 기술은 회피하기 어렵다. 다수의 이용자가 동일한 디바이스를 소유 할 수 있지만 위치 및 시간대 설정, 운영 체제, 설치된 응용 프로그램 및 플러그인, 폰트, 브라우저 버전 등을 추적하면 신속하게 고유 한 장치를 얻을 수 있다. 이러한 디바이스 핑거프린팅 기술은 TCP/IP, 자바스크립트(javascript), 플래시플레이어(flash player), 실버라이트(silverlight), 웹 GL(WebGL) , 자바애플릿(java applet), 캔버스 핑거프린팅(canvas fingerprinting) 등 종류와 방법도 다양하다. 최근에는 마우스의 휠이나 속도 fingerprinting, CPU Benchmark fingerprinting 기법들도 소개되고 있다.

2.5.2 디바이스 핑거프린팅의 보안 활용사례

해외에서는 디바이스 핑거프린팅 기술을 이용한 디바이스 인텔리전스(device intelligence)의 서비스가 꾸준히 성장하고 있다. 디바이스 핑거프린팅을 이용한 디바이스 평판서비스를 제공하는 Threat Metrix Inc.가 2017년 발표한 자료에 따르면, 자사의 디지털 신원 네트워크(digital identity network)를 통하여 전 세계 4만개의 웹 사이트 및 모바일 앱을 통해 14 억 명의 인식 된 디바이스 고유 ID를 제공하고 있다고 밝혔다[25]. 또한 유사한 서비스로 온라인 사기 예방 및 인증을 제공하는 iovation Inc. 역시 비슷한 규모의 디바이스 인텔리전스 정보를 제공하고 있다[26]. 이들 기업들이 제공하는 기술은 온라인 서비스에서 나타날 수 있는 지불 사기, 계정 탈취, 애플리케이션 사기 및 새로운 계정 사기 등을 모니터링하고, 장치의 식별, 공유 장치 평판, 장치 기반 인증 및 실시간 위험 평가 등의 서비스를 제공하고 있다[23],[24].

2.5.3 디바이스 핑거프린팅의 한계점

웹브라우저를 이용한 디바이스 핑거프린팅은 고유한 핑거프린팅 ID를 만드는 것에 초점을 두고 있다. 하지만 핑거프린팅 기반의 추적을 방지하기 위한 방법도 지속적으로 연구되고 있으며, 이는 크게 식별을 방지하는 방법과, 속성 값을 변조 하는 방법이 존재한다. 최근 브라우저에서 제공하는 privacy모드나 트래킹을 방지하는 기능들과 서드파티 콘텐츠를 차단하는 방법이다. 하지만 이러한 방식은 웹 페이지를 렌더링 하는데 문제를 발생시킬 수 있다. 또한 핑거프린팅 값을 변조하여 정보를 무작위로 생성하는 방법이 있는데, 이는 별도의 브라우저에 플러그인을 설치하여야 가능하고 무작위로 생성 된 정보 때문에 정상적인 브라우저로 타입으로 인식하지 않는 등의 문제가 발생 될 수 있다. 이러한 한계점에 대한 대응을 위해 새로운 디바이스를 스코어링하고 추가적인 인증수단을 거치게 함으로서 보완할 수 있다.

III. 웹 기반 디바이스 핑거프린팅 기술

3.1 웹 브라우저를 이용한 디바이스 정보 수집

과거부터 IP주소와 쿠키를 이용한 이용자를 식별하는 기술은 많은 웹 서비스가 이용해온 기술이다. 하지만 최근에는 이용자의 일반적인 웹 행위에서 이용자를 보다 정확하게 식별 할 수 있도록 다양한 방법을 모색하기 시작하면서 더 많은 이용자의 정보를 수집할 수 있다. 보편화 된 기술들을 이용하여 이용자 디바이스의 200여 가지가 넘는 정보를 식별할 수 있으며, 이 중 핵심적인 컴포넌트를 통하여 이용자의 디바이스를 정확하게 식별하고 추적하는 기술들에 대해 알아보도록 한다. 이 기술은 전자 프런티어 재단의 Panopticlick사이트[2]와 Browserleaks[3], 그리고 fingerprintjs 라이브러리를 제공하는 Github의 개발자 프로젝트인 Fingerprintjs2[4] 등을 참조하였다.

3.2 주요 기술을 이용한 정보 수집

3.2.1 IP 주소

일반적으로 디바이스가 웹 서버로 접근하게 되면서 이용자의 신원을 확인할 수 있는 정보가 생성이 된다. 여기에는 IP 주소 및 HTTP 요청 헤더와 같은 기본 기능이 포함되어 있다. 추가적으로 프록시에 대한 탐지를 포함할 수 있게 되는데, 프록시 IP에 대한 국가, 시/도, ISP(Internet Service Provider) 및 ASN(자율 시스템 번호, Autonomous System Number), 현지 시간, 위도 및 경도에 대한 GeoIP 데이터 수집을 수행한다. 또한 Passive TCP/IP OS 핑거프린팅, DNS 및 WebRTC(Web Real-Time Communication)의 노출 테스트와 같은 특수 기능이 사용된다.

3.2.2 JavaScript

브라우저에서 자바스크립트를 실행하는 경우에는 많은 이용자의 디바이스 속성을 얻을 수 있다. 문서 객체 모델(DOM, Document Object Model)은 HTML, XML 문서의 프로그래밍 인터페이스(interface)로서 웹 서비스의 HTML과 같은 문서의 구조화된 표현(structured representation)을 제공하며 자바스크립트 언어가 DOM 구조에 접근할 수 있는 방법을 제공하여 문서 구조, 스타일, 내용 등을 변경할 수 있게 돕는 역할을 수행한다. 이러한 DOM은 웹 페이지를 스크립트 또는 프로그래밍 언어들에서 사용될 수 있게 연결시켜주는 역할을 담당한다[27].

이 기능을 사용하여 웹 브라우저에 대한 민감한 정보인 User-Agent, 아키텍처, OS 언어, 시스템 시간, 화면 해상도 등을 제공한다. 또한 HTML5의 API들을 통하여 브라우저 플러그인 설치정보, 배터리 상태, 터치스크린 지원여부 등 상세한 디바이스의 정보를 수집할 수 있게 된다. 위의 Table 4는 자바스크립트를 이용하여 디바이스에서 추출할 수 있는 주요 항목을 나열하였다.

Table 4. A information item table that can be collected by the device using the javascript

Group	Key Attributes
JavaScript	JavaScript Enabled
	Inline Scripts
	Same-Origin Scripts
	Third-Party Scripts
Screen Info	Screen Resolution
Date	System Time
	Time Zone
Device Info	userAgent
	appVersion
	appName
	appCodeName
	product
	vendor
	platform
	hardwareConcurrency
	deviceMemory
	language
	languages
	doNotTrack
	cookieEnabled
	maxTouchPoints
Battery Status	API Vendor
	Battery Charging Time
	Battery Discharging Time
	Battery Level
etc.	InstalledPlug-Inslist

2) https://panopticlick.eff.org/
3) https://browserleaks.com/
4) https://github.com/Valve/fingerprintjs2

3.2.3 HTML5 캔버스 핑거프린팅

캔버스 핑거프린팅은 브라우저 쿠키 없이 웹 사이트에 접속한 이용자를 식별하여 추적하는 브라우저 핑거프린트 기술 중 하나로서 식별 정확도가 100%가 아니기 때문에, 광고 타겟팅에 주로 사용되었던 기술이다. 이 캔버스 핑거프린팅은 자바스크립트로 스크립팅을 통하여 웹 페이지에 그래픽과 애니메이션을 그리는 데 사용되는 HTML5 API이다. 이용자 디바이스의 그래픽칩셋의 처리 기술 및 브라우저의 랜더링 기술에 따라서 이미지, 문자, 선 등을 처리하는 방법이 차이가 난다. 이것을 이용하여 이용자마다 고유 값을 만드는 방법으로 디바이스를 95%정도의 정확성으로 고유하게 식별할 수 있다[28].

캔버스 핑거프린팅은 동일한 디바이스에서(그래픽 칩셋, OS, CPU, 브라우저 등) 다소 낮은 확률로 디바이스 고유 식별 값이 겹치게 될 수 있으므로 다른 정보들과 결합하여 고유한 디바이스를 식별하는 기술로 활용되고 있다.

3.2.4 Flash Player

브라우저에서 실행되는 플래시 플레이어의 런타임(runtime)속성에서 플래시 버전, 플러그인 유형, 운영 체제, 제조업체, 시스템 언어, 웹 브라우저 아키텍처, 화면 해상도 및 하드웨어 및 멀티미디어 기능 등의 정보를 수집할 수 있다. 하지만 이는 브라우저에서 기본적으로 플래시를 실행 할 수 있는 조건을 가져야 한다.

3.2.5 WebGL 핑거프린팅

WebGL은 플러그인을 사용하지 않고 호환 가능한 모든 웹 브라우저에서 대화식 3D 그래픽을 렌더링하는 자바스크립트 API이다. WebGL 앱은 자바스크립트로 작성된 제어 코드와 디바이스의 GPU에서 실행되는 특수 효과 코드로 구성된다. WebGL은 다른 HTML 요소와 혼합되어 페이지 또는 페이지 배경의 다른 부분과 합성 되어 구현된다. WebGL 브라우저 보고서는 웹 브라우저에서 WebGL 지원을 확인하고 WebGL 장치 핑거 프린팅을 생성하며 다른 WebGL 및 GPU 기능과 관련된 웹 브라우저 식별 정보를 보여준다. 주요 식별 정보는 브라우저의 User-Agent, GL의 버전, 벤더, 렌더 칩셋 정보 등을 활용하여 3D 그래픽을 랜더링 하는 방법의 조합으로 해쉬 값을 만들어 낸다.

주요 브라우저 공급사인 애플, 구글, MS 및 모질라는 기본적으로 WebGL을 지원하고 있다.

IV. 웹기반 디바이스 핑거프린팅 모델 제안

4.1 디바이스 핑거프린팅을 위한 조건

본 논문에서는 이용자 디바이스의 200여 가지가 넘는 정보 중 브라우저의 속성이 비교적 일정하게 나타나고 디바이스 식별을 위한 의미 있는 값을 분류하기 위하여 핵심적인 컴포넌트를 선정하였으며, 이러한 선정조건에는 국내의 인터넷 환경 적용, 구현의 용이성, 디바이스별 공통 식별 값 선정, 안티 핑거프린팅의 회피, 프라이버시 침해 최소화 등의 조건을 통하여 선정하였다. 식별 값의 선정 조건과 추후 제안할 시스템 구성 및 디바이스 탐지방안의 조건에 대하여 간략하게 알아보겠다.

첫째, 국내의 인터넷 환경이 적용되어야 한다. 국내 이용자의 디바이스 정보를 보다 정확하게 판단하기 위해서는 국내의 인터넷 환경에 맞는 가중치를 부여할 필요성이 있다. 국내에서 사용하는 디바이스의 경우 Language, TimeZone 등의 정보는 일치하는 경우가 대부분이기 때문에 디바이스를 식별하는 방법 중에서 해당 속성 값의 가중치를 낮출 필요가 있다. 또한 모바일 디바이스의 경우 접속하는 IP에 대한 가중치를 낮추는 등 유동적인 조건 값 변경이 필요할 수 있다. 이는 각 웹 서비스의 특성에 따라 변경이 가능할 것이다.

둘째, 구현이 용이해야 할 것이다. 웹 서비스에서 자바스크립트나 플래시, 캔버스 핑거프린팅 등 디바이스의 정보를 수집하기 위해 최소한의 코드를 사용하여 이용자의 접근성과 웹 사이트 접근 시 속도저하 등을 최소화 하여야 한다. 실제 방대한 자바스크립트를 구현함으로서 불필요한 정보를 수집하게 되거나, 스크립트의 구현에 따른 웹 사이트의 페이지 랜딩속도의 저하가 필연적으로 발생하게 됨에 따라, 이를 최소화하는 구현방법이 필요할 것이다.

셋째, 디바이스 브러우저에 공통된 식별 값을 선정하여야 한다. 이용자 디바이스의 브라우저에는 식별이 가능한 값과 식별할 수 없는 값들이 존재한다. 사용되는 브라우저별로 수집될 수 있는 항목이 다르기 때문에 정확한 결과와 식별이 될 수 있는 값을 선

정하되, 모바일 디바이스를 공통적으로 적용할 수 있는 기술이 필요하다.

넷째, 안티 핑거프린팅 기술을 회피할 수 있어야 한다. 헤더 값을 숨기는 브라우저, 트래킹을 숨길 수 있는 다양한 툴(tool)들이 프라이버시 침해를 방지하기 위하여 제공되고 있다. 최근에는 토르 네트워크 및 토르 브라우저 등의 이용도 늘고 있는 추세이다. 이러한 안티 트래킹 툴 들이나 프락시 등의 사용으로 디바이스의 속성 값이 지속적으로 변조되는 공격패턴을 감지하고 추가적인 보안절차를 마련할 수 있어야 한다. 이용자 디바이스에서 수집되지 못하는 정보에 대해서 새롭게 추적하는 방법을 추가하거나 이를 검증하는 기술을 적용하여 안티 핑거프린팅을 회피 할 수 있어야 할 것이다.

다섯째, 프라이버시 침해를 최소화 하여야 한다. 디바이스를 통하여 이용자 혹은 디바이스 자체만을 식별하는 것 이외의 목적으로 사용되지 않도록 디바이스를 식별하기 위한 최소한의 정보만을 수집하여야 한다. 수집되는 정보 자체가 개인정보침해의 논란이 될 수 있는 상황이 생길 수 있으므로, 법적인 수집내용의 고지와 함께 수집되는 정보를 해쉬(hash)화 하여 저장함으로서 기본적인 프라이버시 보호에 대한 조치가 있어야 할 것이다.

4.2 최적의 정보 수집 모델

앞서 우리는 웹 기반 디바이스 핑거프린팅 기술에서 수집될 수 있는 다양한 값을 확인하였다. 본 논문 연구를 통하여 다양한 디바이스 정보를 수집하고 분석하였으며, 최적의 디바이스 핑거프린팅 조건을 제시하기 위하여 기본적인 모델을 제안하였다. 이 모델은 본 논문 2장에서 조사한 디바이스 핑거프린팅의 보안 활용사례를 통해 해외에서 디바이스 핑거프린팅 기술을 이용한 디바이스 인텔리전스 서비스를 제공하는 두 개의 회사를 포함한 전자 프런티어 재단의 연구 조사결과를 기준으로 각 항목을 동일하게 그룹화 하였다. EFF에 연구 결과에 따르면 모든 디바이스 인텔리전스 회사는 자바스크립트 이외에 이용자 환경을 핑거프린팅하기 위해 플래시 기술을 사용하고 있다[15]. 하지만 플래시 기술은 최근 여러 가지 보안 이슈를 통하여 브라우저 자동실행 플러그인에서 예외되는 양상을 나타내고 있으며, 브라우저에서 기본적으로 플래시를 실행 할 수 없는 조건에서는 의미가 없는 비교항목이 됨으로 본 연구에서는 플래시 기술을 제외하여 제안하였다. Table 5는 연구의 결과로 제안된 최적의 모델이다. 이는 절대적인 값이 아니라 각 웹 서비스의 성격에 따라 수집하여 디바이스 핑거프린팅 하는 항목을 조정할 수 있을 것이다. 특히 IP 주소나 TCP/IP OS 핑거프린팅의 경우에는 네트워크 값의 변경이 많은 경우에는 탐지모드로 사용

Table 5. A Proposal of collection information list for Device Fingerprinting

Fingerprinting Category	Proposal Model	Details	Value
Browser customizations	Plugin enumeration	Installed Plug-Ins	hash
Browser-level user configurations	Cookie Enable	cookieEnabled	value
	DNT Enable	doNotTrack	value
	Memory	deviceMemory	value
	Timezone	Time Zone	value
	Flash enabled	Flash Player Enabled	value
Browser family & version	User Agent	"User-Agent" header	hash
	HTTP Header	Accept, Accept-Language	hash
Operating System & Applications	Font hash	System Fonts	hash
	Platform	Platform	value
	Touch Support	Touch Support	hash
Hardware & Network	Screen Resolution	Screen Size, Color Depth	hash
	Canvas Fingerprinting	Canvas Fingerprinting hash	value
	WebGL Fingerprinting	WebGL Fingerprinting hash	value
	IP Address*	IP address	
	TCP/IP OS Fingerprinting*	Passive, SYN	

* If there are many changes of network value, it can be used in detection mode or differently depending on the nature of website.

하거나 웹 사이트 성격에 따라 다르게 적용하는 것을 제안한다.

4.3 시스템 구성의 방법

위에서 제시된 최적의 정보 수집 모델을 기반으로 실제 웹 서비스에 구현하기 위해서는 Fig. 1과 같이 디바이스 핑거프린팅을 위한 브라우저 수집 스크립트를 웹 서버에서 동작하게 하고 핑거프린팅 서버를 이용하여 특성 값을 조합하여 디바이스 핑거프린팅 ID를 생성한 후 디바이스 핑거프린팅 ID를 중심으로 행동을 추적하고 의심 디바이스에 대한 태깅(tagging)이나 차단, 혹은 서비스의 형태에 따라 ARS인증이나 국내 본인인증과 같은 추가적인 인증을 요구하는 방법으로 관리될 수 있다.

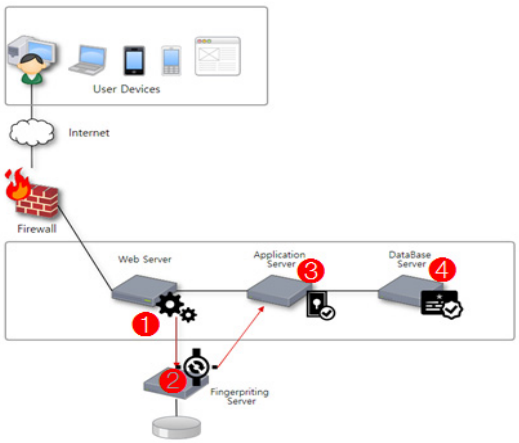

Fig. 1. Proposal model with simplified collection script, analysis system and WEB / WAS interworking system

4.4 온라인 사기 예방을 위한 디바이스 스코어링

기업의 웹서비스들은 온라인 사기와 같은 공격방법은 산업에 따라 다르지만 대부분 계정을 이용하는 웹 서비스들은 동일한 위험에 노출되어 있다. 이러한 공격에 대응하기 위해 전통적인 방어수단을 강구하는 것은 물론 필요한 사항이나, 네트워크에 연결하는 디바이스를 식별하는 것은 웹 서비스에 미칠 수있는 위험을 인식하고 공격을 방어하는 중요한 기술로 평가될 수 있다. 디바이스 기반의 공격 예방은 온라인 환경에서 공격을 효과적으로 방지하고 신뢰할 수 있는 사용자에 안전한 서비스를 제공하는데 큰 역할을 할 수 있을 것이다.

일반적으로 공격자의 디바이스는 하나의 웹 서비스만을 위해 설정되지 않는다. 장치에 대해보다 정확한 데이터를 보유할수록 위험을 평가하고 공격을 중지하는 것이 더 쉬워진다. 디바이스를 식별하는 것은 온라인 공격에 대한 첫 번째 방어선이며 위험성이 높은 행동 양식을 식별하는 강력한 도구가 될 수 있다.

4.5 디바이스 스코어링 모델과 활용

디바이스 핑거프린팅 ID(이하 디바이스 ID)를 식별하고 그 행동을 탐지하는 방법은 현재 사용되고 있는 기술인 쿠키를 이용하여 고유한 ID를 만들어 내고 이에 디바이스 핑거프린팅 ID를 결합한 형태의 식별 방법을 제안한다. 사이트에 따라서는 고유한 시리얼 번호나 이용자의 계정 아이디 등 웹 서비스 회사가 보유한 고유한 값들과 결합한다면 더 가치 있는 탐지 방안이 될 것이다. 이 방법에는 두 개의 고유키(key)가 사용된다.

첫 번째로 사용되는 키는 쿠키 기반의 PUID (Product Unique Identifier)이다. 이는 이용자의 웹브라우저가 최초 웹 서비스에 접속시 쿠키를 기반으로 생성하는 값으로 UUID와 같은 개념이다. 이 키는 이용자의 브라우저가 해당 웹 사이트에 접속했는지에 대한 흔적을 확인하는 데 사용할 수 있다.

두 번째로 사용되는 키는 브라우저의 고유한 값들을 이용해서 생성한 DFID(Device Fingerprint Unique Identifier)이다. 이 키는 앞서 제안한 최적의 정보 수집 모델의 각 값들을 해쉬화 하여 생성할 수 있다. 예를 들어 SHA256으로 해쉬하는 경우에는 64자리의 고유한 키가 만들어 지게 된다. 이러한 키는 이용자 디바이스의 브라우저 Cookies, Local Storage, HTML5 Web SQL 등 가용한 저장공간에 저장해 놓고 웹 사이트 접근 시 비교하는 프로세스를 적용한다. 이는 각 사이트의 탐지수준 및 정책에 따라 상이한 기준을 적용 할 수 있다. Fig. 2는 단말이 웹 사이트에 처음 접근하게 되면 디바이스 핑거프린팅 프로세스가 시작되어 이용자 브라우저의 쿠키 값을 기준으로 PUID를 생성하게 된다. 이 PUID와는 별개로 시스템은 이용자의 디바이스 정보를 획득하는 과정을 통하여 DFID를 생성한다. 이 두 개의 키는 이용자 디바이스와 디바이스 핑거프린팅 시스템 혹은 웹 서비스의 DB에 저장되어 다음

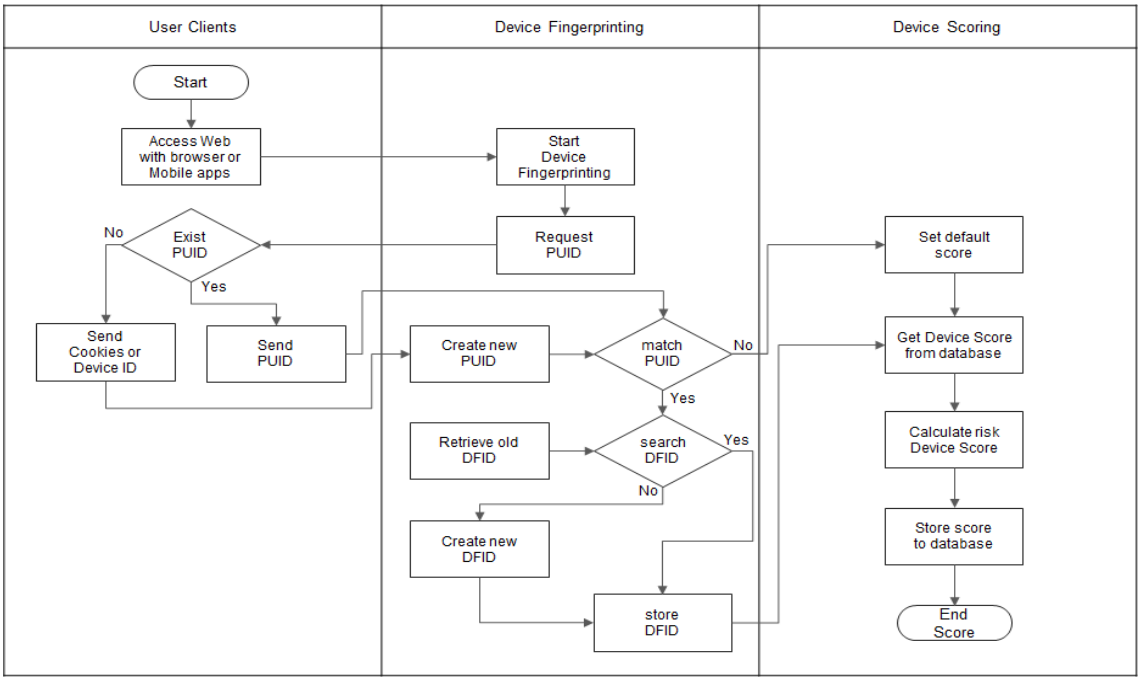

Fig. 2. Proposal model with two unique keys matching for Device Scoring

비교 시 활용 된다. 이용자의 PUID가 존재하는 상황에서 DFID가 없는 경우나 일치하지 않는 경우에는 새롭게 생성하여 그 변경을 기록하고, 이 디바이스에 대해서는 태깅이나 차단 등으로 모니터링을 강화하고, 서비스의 형태에 따라 ARS인증이나 국내 본인인증과 같은 추가적인 인증수단을 이용하여 이용자를 다시 한번 더 식별하는 프로세스를 이용하면 효과적일 것이다. 이 PUID와 DFID 두 개의 키를 이용하여 보안 관리를 위해 이미 사용 중인 디바이스의 접근 IP주소관리 나 로그인하는 계정의 이력관리와 연동하여 적절히 사용할 수 있다.

4.6 디바이스와 계정의 연결

기업 내 웹서비스에 적용하기 위해 디바이스와 계정정보를 연결하고 이 연결을 추적하면 함께 작동하는 사이버 공격자를 발견하는 유용한 무기로 만들 수 있다. 조직되고 정교한 해킹단체의 경우 종종 여러 지역에 위치한 서로 다른 유형의 장치를 사용하는데, 동일한 계정에 로그인하면 이러한 연결정보를 추적관리 하면 IP 주소기반으로 탐지하는 현재의 방어방법보다 훨씬 더 정교한 탐지가 가능하다. 예를 들어 동일한 모바일 게이트웨이 IP를 사용하는 경우는 그 활동에 대한 탐지나 차단이 불가능하나, 하나의 디바이스가 여러 계정을 차례로 만들거나 여러 디바이스가 모두 동일한 계정을 사용하는 경우에도 탐지가 가능하며 이러한 디바이스 및 사용자 계정은 별도의 스코어링 관리를 통하여 위험관리를 할 수 있다.

V. 실험 및 검증

5.1 국내의 인터넷 환경의 적용

앞서 우리는 디바이스 핑거프린팅이 국내의 인터넷 환경이 적용되어야 한다고 제안하였다. 다시 말해 국내 이용자의 경우는 국내의 인터넷 사용 환경에 맞는 가중치를 부여할 필요성이 있다. 이 부분을 검증하기 위해 오픈 소스이며 브라우저의 주요 특성에 대한 간략한 요약뿐만 아니라 글로벌 통계를 정보를 제공하고 있는 amiunique[5]의 데이터와 국내의 A커머스 기업의 브라우저 정보를 비교하였다. Table 6은 국내의 한 PC에서 amiunique에 접속하여 디바이스 핑거프린팅을 한 결과 샘플을 나타낸다. 68만개의 브라우저 테스트 결과이며, 샘플 데이터에서 유

5) https://amiunique.org/

의깊게 봐야 할 내용은, 각 항목에 대한 유사도 부분이다. 특히 Contentlanguage, Language 등의 항목은 테스트된 해외 데이터에서 0.1%미만의 결과를 나타낸다.

amiunique의 글로벌의 데이터 기준으로 OS는 윈도우가 56.66%, 리눅스가 14.53%가 그 뒤를 이었다. 브라우저는 파이어폭스가 42.7%, 크롬이 39.9%를 기록하였으며, 시스템 언어는 영어가 62.5%, 프랑스어가 11.1% 순으로 나타났다. 타임존의 경우 UTC+1이 20.8%, UTC+2가 19.1%로 나왔다[29].

본 테스트는 브라우저 샘플 데이터의 편향성이 보인다. 기본적으로 해외의 데이터이며, 핑거프린팅에 대한 동의 후 데이터를 수집한 결과이다. 특히 이 데이터는 프라이버시에 관심 있는 인터넷 이용자의 인구를 대표 할 가능성이 크며, 웹 사이트의 트래킹을 회피하기 위한 이용자들이 주로 이용한 결과로 예측이 된다.

다음으로 국내의 인터넷 이용자의 디바이스 속성을 비교해 보기 위하여 우리는 구글 애널리틱스(google analytics)가 설치되어 운영 중인 국내의 A 커머스 기업의 데이터를 이용하였다. 구글 애널리틱스는 전 세계적으로 가장 많이 활용되는 무료 웹 로그 분석 도구로서 웹 사이트와 모바일 앱 데이터를 수집, 측정, 분석하고 다양한 리포트를 제공하는 기능을 수행하고 있다. 일반적으로 웹 사이트와 모바일 앱, 그리고 온라인 광고에 대한 효과성을 분석하는데 활용되고 있다.

우리는 구글 애널리틱스의 자바스크립트[6] 를 분석하여 구글 애널리틱스가 이용자 디바이스의 대부분의 속성 정보를 수집하는 것을 확인할 수 있었다. 물론 구글이 이 디바이스 핑거프린팅 기술을 어떻게 적용하고 있는지에 대한 분석은 불가하였지만, 기본적으로 이용자의 디바이스를 추적하거나 혹은 크로스 디바이스 트래킹을 이용하여 이용자를 추적하는 것을 추론할 수 있었고, 실제 애널리틱스 리포트에서는 브라우저가 아닌 이용자 단위의 디바이스로 리포트되는 것을 확인하였다.

A커머스 기업의 2018년 1월부터 3월까지의 데이터, 약 4천 8백만개 이상의 디바이스 이용자로 구분하여 구글 애널리틱스를 이용해 분석해 보았다. 구글 애널리틱스의 특성 상 PC 웹와 모바일 웹 및 안드로이드, iOS 앱이 각각 다른 경로로 데이터를 수집하였기 때문에, 이번 분석에서는 PC웹과 모바일 웹의 애널리틱스 두 가지만 분석하였다. GeoIP를 기준으로 국내 접속이 97% 이상으로 나타났으며, OS는 PC 웹의 경우 윈도우가 81%, 모바일 웹은 97.8%가 안드로이드와 iOS가 차지하였다. 브라우저는 PC 웹에서 인터넷 익스플로어가 64.6%, 모바일 웹에서 안드로이드 웹뷰(android webview)가 38%의 빈도를 보였다. 시스템 언어는 PC 웹의 경우 98.3%, 모바일 웹의 경우 94.4%가 한국어를 사용하는 것으로 나타났다.

해외의 데이터와 국내의 데이터를 비교해 본 결과 국가적인 특성이 분명하게 나타났다. 물론 국내에서 해외를 대상으로 서비스하고 있는 형태가 아닌 국내 중심의 A커머스 데이터 만을 비교하였지만, 비교적 방대한 데이터를 분석해 본 결과, 디바이스의 속성

Table 6. Sample of similarity result of global device attribute data by amiunique project.

Attribute	Similarity ratio	Value
Useragent	0.25%	"Mozilla/5.0(WindowsNT1..."
Accept	0.81%	"text/html, application..."
Contentencoding	29.88%	"gzip, deflate"
Contentlanguage	<0.1%	"ko-KR"
Listofplugins	<0.1%	"Plugin1:Flash 29.0.0.140:..."
Platform	41.13%	"Win32"
Cookiesenabled	79.83%	"yes"
DoNotTrack	50.49%	"NC"
Timezone	1.22%	"-540"
Screenresolution	3.72%	"1920x1200x24"
Useoflocalstorage	76.97%	"yes"
Useofsessionstorage	76.97%	"yes"
Canvas Fingerprinting	0.22%	
WebGLVendor		"Microsoft"
WebGLRenderer		"NVIDIA GeForce..."
Listoffonts	<0.1%	elision
Screenresolution	1.36%	"1920x1200"
Language	<0.1%	"ko"
Platform	3.89%	"Windows 10"
UseofAdBlock	49.10%	"no"

6) http://www.google-analytics.com/analytics.js

값은 국가의 인터넷 사용 환경에 맞도록 혹은 웹 서비스의 서비스 대상의 성격에 따라 적절한 가중치 부여는 의미가 있다. 국내환경에서 국내를 대상으로 서비스하는 웹 사이트에 디바이스의 속성이 국내가 아닌 제 3국의 디바이스가 접속한다면 스코어링 모델에 따라 처리하는 것이 효과적일 것이다.

5.2 구현의 용이성

웹 서비스에서 디바이스의 정보를 수집하기 위한 핑거프린팅은 최소한의 코드를 사용하여 이용자의 접근성과 웹 사이트 접근 시 속도저하를 최소화 하여야 한다. 아래 Fig. 3은 테스트 웹사이트를 구축하여 테스트한 결과로 50여회의 테스트에서 동일한 디바이스에서 같은 브라우저를 사용하였을 시 모두 동일한 핑거프린트 값이 추출되었으며 결과속도도 평균 184.2ms였다. 이는 충분히 일반적인 웹 서비스에 구현할 수 있을 것으로 기대한다.

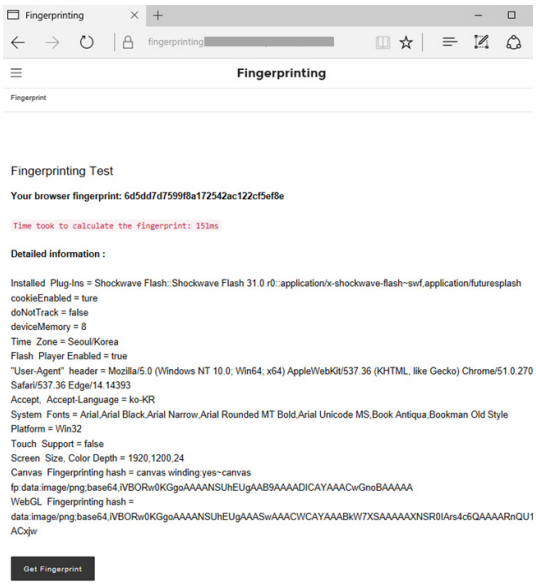

Fig. 3. Test results made with a simple implementation

5.3 브라우저별 공통된 식별 값 검증

우리는 디바이스나 브라우저에 공통된 식별 값을 선정하기 위하여 국내에서 흔히 접할 수 있는 다양한 브라우저를 이용하여 그 결과값을 추출하였다. Table 7은 디바이스 핑거프린팅을 위해 제안된 모델의 디바이스 속성 값을 다양한 브라우저에서 식별할 수 있는지에 대한 테스트를 실행한 결과이다.

PC에서 자주 사용되고 있는 대표적인 브라우저인 인터넷 익스플로어(IE), 파이어폭스(FF), 크롬(CH), 오페라(OP) 브라우저를 테스트 하였으며, 국내의 포털사에서 제작하여 배포하고 있는 웨일(WH) 브라우저도 테스트에 포함하였다. 모바일 디바이스의 테스트를 위해 몇 가지 종류의 스마트폰을 이용하여 모바일 브라우저들을 테스트 하였으며, 이색적인 테스트를 위하여 중국 샤오미(xaomi)의 미(MI, Miui) 브라우저도 테스트 해 보았으며, 네이버의 네이버 앱(NA)을 통한 핑거프린팅도 시도해 보았다. 테스트한 모든 브라우저에서 제시된 항목의 속성값을 일정하게 획득하는 것을 확인할 수 있었다.

다음으로 디바이스 핑거프린팅의 고유성을 검증해 보기 위하여 정보 엔트로피(entropy)를 측정하였다. 엔트로피는 누군가의 정체성을 유일하게 드러내는 데 얼마나 가까운지를 측정 할 수 있게 해주는 수학적 양으로서 단위로 비트(bit)를 사용한다. 정보는 발생 가능한 사건이나 메시지의 확률분포의 음의 로그로 정의할 수 있는데, 수식(1)은 우리가 한 속성에서 새로운 사실을 알게 된다면 그 사실은 불확실성의 엔트로피를 어느 정도 줄여 주게 되는지를 수식

Table 7. Verification result of effective value per browser

Attributes	Tested Results						
	IE	FF	CH	OP	WH	MI	NA
Plugin enumeration	O	O	O	O	O	O	O
Cookie Enable	O	O	O	O	O	O	O
DNT Enable	O	O	O	O	O	O	O
Memory	O	O	O	O	O	O	O
Timezone	O	O	O	O	O	O	O
Flash enabled	O	O	O	O	O	O	O
User Agent	O	O	O	O	O	O	O
HTTP Header	O	O	O	O	O	O	O
Font hash	O	O	O	O	O	O	O
Platform	O	O	O	O	O	O	O
Touch Support	O	O	O	O	O	O	O
Screen Resolution	O	O	O	O	O	O	O
Canvas Fingerprinting	O	O	O	O	O	O	O
WebGL Fingerprinting	O	O	O	O	O	O	O

Table 8. Entropy measurement result by proposed attribute

Fingerprinting Category	Attribute	Details	bits of identifying	Value
Browser customizations	Plugin enumeration	Installed Plug-Ins	5.68	hash
Browser-level user configurations	Cookie Enable	cookieEnabled	0.22	value
	DNT Enable	doNotTrack	1.25	value
	Memory	deviceMemory	2.5	value
	Timezone	Time Zone	7.21	value
	Flash enabled	Flash Player Enabled	0.9	value
Browser family & version	User Agent	"User-Agent" header	7.3	hash
	HTTP Header	Accept, Accept-Language	13	hash
Operating System & Applications	Font hash	System Fonts	4.71	hash
	Platform	Platform	1.52	value
	Touch Support	Touch Support	0.59	hash
Hardware & Network	Screen Resolution	Screen Size, Color Depth	5.34	hash
	Canvas Fingerprinting	Canvas Fingerprinting hash	18.88	value
	WebGL Fingerprinting	WebGL Fingerprinting hash	8.89	value

화하는 공식이다. S는 비트 단위로 측정 된 엔트로피의 감소량이고 $Pr(X=x)$는 임의의 항목을 알게 될 확률이다.

$$S=-\log_2 \Pr(X=x) \quad (1)$$

세부적인 엔트로피를 표한하기 위한 계산으로는 아래 수식[2]를 이용하여 계산 할 수 있다.

$$S=-\log_2(\frac{x}{X}) \quad (2)$$

단순한 계산으로 동전던지기와 같이 2가지의 가능성이 있는 경우는 1bit, 4가지의 가능성이 있는 경우 2bit가 된다. 예를 들어 IPv4 환경에서 IP는 전 세계 기준으로 약 42억개가 사용될 수 있다. 이중 특정 하나의 IP가 식별될 수 있는 엔트로피는 아래와 같이 풀이 될 수 있다.

$$S=-\log_2(\frac{1}{4{,}200{,}000{,}000})=31.96bit$$

이용자의 디바이스 속성값 중 시스템 언어가 ko-KR인 경우는, 윈도우 7기준으로 36개의 시스템 언어를 지원함으로 아래와 같이 계산될 수 있다.

$$S=-\log_2(\frac{1}{36})=5.16bit$$

즉, 시스템 언어가 ko-KR이면서 특정 IP를 가지는 시스템의 엔트로피는 37.12bit가 된다.

모든 디바이스의 속성을 분석하고 엔트로피를 측정 하는 것은 충분한 시간과 방대한 데이터를 분석할 필요가 있기 때문에, 본 연구에서는 전자 프런티어 재단의 Panopticlick과 오픈소스 amiunique의 유사도와 엔트로피 검증결과를 토대로 제안한 최적의 디바이스 핑거프린팅 조건에 각각 환산한 엔트로피를 Table 8과 같이 측정을 하였다. 검증결과 각 항목의 평균은 5.57bit이며 제안된 항목의 합계는 77.99bit로 검증되었다. 위 제안된 최적화 모델을 기준으로 디바이스 핑거프린팅을 수행 하였을 경우, 디바이스 ID의 고유성은 1.20×10^{24} 의 확률로 중복 될 수 있는 결과이다.

5.4 안티 핑거프린팅의 회피

안티 핑거프린팅 기술을 회피하기 위해서는 안티 핑거프린팅 툴(프라이버시 침해차단 소프트웨어)이나 토르 네트워크 및 토르 브라우저 등을 이용하게 되는데, 아쉽게도 본 연구에서 안티 핑거프린팅 기술을 회피하지는 못하였다. 우리는 안티 트래킹 조건하에서는 User Agent, Time Zone과 같은 디바이스의 속성 값이 일부 변경 되는 것을 확인 하였으며 이는 디바이스 ID를 지속적으로 변경시켰다. 하지만 안티 트래킹에 대한 대응에 대한 방법은 위에서 제시한 디바이스 스코어링 방안을 통하여 제안했듯이, DFID가 변경되었을 경우의 프로세스를 이용하여

대응하는 것이 필요하다.

5.5 프라이버시 침해의 최소화

디바이스의 특성을 통하여 이용자를 식별하고 디바이스 ID를 이용자 계정과 연결하는 등 가시성을 확보하기 위한 노력이 프라이버시 침해와는 상반되는 주장이 된다. 제안된 디바이스 핑거프린팅은 이용자의 디바이스와 행동에 대한 정보를 수집하게 되는데, 비록 개인을 특정하기가 어렵다 하더라도 이용자의 계정정보와 연계하면 개인 정보로 이어지기 때문에 핑거프린팅 정보를 수집하는 웹 서비스는 철저하게 보안적인 목적 및 관리용도로 사용되어야 할 것이다. 특히 개인정보 처리방침 및 이용 약관에 관련한 내용을 포함하여 법적인 분쟁의 소지를 해결함과 동시에 이용자의 개인정보 자기결정권을 보장하여야 한다. 또 디바이스를 식별하기 위한 최소한의 정보만을 수집하고, 수집되는 정보는 암호화 및 해쉬화 하는 것이 필요하다.

VI. 활용 방안

본 연구에서는 우리는 웹 기반의 디바이스 핑거프린팅을 이용하여 웹 서비스에 접근하는 디바이스를 식별하여 관리함으로서 온라인 사기거래 및 어뷰징을 탐지하는 방법에 대해 확인하였다. 이용자의 온라인 행위를 트래킹하는 기술과 보안적인 측면에서의 디바이스 핑거프린팅 기술에 대하여 설명하였고, 디바이스 핑거프린팅 기술을 이용하여 데이터를 획득하는 방법에 대해 알아보았다. 또한 디바이스 핑거프린팅을 이용하여 데이터를 수집하고 그 결과에 대하여 검증하였다. 본 연구의 궁극적인 목표인 디바이스 핑거프린팅을 이용하여 웹 서비스와 이를 이용하는 이용자를 안전하게 보호할 수 있는 방안을 도출하였고 몇 가지 유의미한 결과와 함께 그 활용방법을 제시할 수 있다.

첫째, 온라인 사기의 적극적인 대응이 가능하다. 이용자 디바이스를 식별함으로서 정상적인 이용자를 보호하고 디바이스를 추적하여 차단하거나 관리할 수 있다. 이 기술은 서버에 별도의 에이전트(agent)를 설치하지 않고 식별이 가능하며 기존의 IP중심의 차단방식을 보완하여 활용 할 수 있다. 각 서비스의 특성에 따라 이용자별 구매 횟수, 새로운 이용자 계정 생성에 대한 제한 등 특정행위에 대한 유연한 통제가 가능하며, 모바일에 대한 제어도 가능하다.

둘째, 이용자 계정에 대한 가시성을 확보 할 수 있다. 디바이스에서 사용되는 계정정보를 추적함으로서 각 서비스별로 유효한 계정을 확인하는 행동들에 대한 차단이 가능하다. 또한 부정 사용되는 계정에 대한 통계 및 처리가 가능하여 궁극적으로 계정 중심의 관리가 아닌 디바이스 중심의 보안이 가능하다.

셋째, 다양한 공격에 효과적으로 방어할 수 있는 기반을 마련할 수 있다. 일반적으로 웹 사이트 해킹을 위해 사전탐색에 대한 모니터링을 강화하는 수단으로 사용할 수 있고, 특이 행동과 이상 거래 행위에 대한 탐지도 가능하다. 웹 크롤러나 봇과 같은 공격에는 더욱 효과적으로 사용될 수 있으며, 반복적인 특정 행위에 대해 관리할 수 있다.

마지막으로 전 세계적으로 많이 활용되는 웹 로그 분석 도구나 마케팅 툴에서 그 기능이 활용되고 있듯이, 핑거프린팅 자체가 웹 서버의 부하나 장애로부터도 비교적 자유롭고 안전한 방식으로 운영될 수 있는 장점을 지니고 있다.

VII. 결 론

웹 서비스를 안전하게 보호하기 위하여 여러 가지의 보안 기술들이 활용되고 있다. 그러나 이러한 기술들만 이용해서는 다양해지고 복잡해지는 웹 공격을 효과적으로 방어하는데 한계가 있다. 이에 본 논문에서 웹 기반의 디바이스 핑거프린팅을 이용하여 다양한 웹 공격의 위협으로부터 효과적으로 사용될 수 있는 것을 확인하였다.

다만 본 연구에서 국내 인터넷 사용 환경에 맞는 엔트로피를 추가적으로 검증할 필요가 있으며, 웹 기반 디바이스 핑거프린팅을 이용하여 실제 사이트에서 거래사기 및 어뷰징 차단기술을 적용해 좀 더 세밀한 프로세스를 만들어 낼 필요가 있다. 또한 디바이스 핑거프린팅을 탐지하는 기술들에 대응하기 위하여 자바스크립트 난독화 등의 기술도 접목할 필요가 있으며, 이는 향후 추가적으로 연구할 계획이다.

References

[1] OWASP, “OWASP top ten project, OWASP top 10 - 2017”, https://www.owasp.org/index.php/Category:OWASP_Top

_Ten_Project, 2017.
〔2〕 Korea Communications Commission, "Press releases - Concentrated illegal transactions such as ID on the Internet", Mar. 2018.
〔3〕 Lim Hyungjin, "Development direction of abnormal behavior detection system technology," The Journal of The Korean Institute of Communication Sciences, Vol. 34, No. 3, pp. 37~46, Feb, 2017.
〔4〕 Financial Security Research Institute, "Overseas Financial Transaction Detection System Technical Guide", Aug. 2014.
〔5〕 ZDNet Korea, "http://www.zdnet.co.kr/news/news_view.asp?artice_id=20180822173414&type=det&re=zdk".
〔6〕 LexisNexis, "2018 True Cost of Fraud: Retail Edition". July 2018.
〔7〕 Seong Hoon Jeong, Hana Kim, Youngsang Shin, Taejin Lee and Huy Kang Kim, "A Survey of Fraud Detection Research based on Transaction Analysis and Data Mining Technique," Journal of the Korea Institute of Information Security & Cryptology, Vol. 25, No. 6, pp. 1525~1540, Dec, 2015.
〔8〕 Kim Jung Sun, "Anomaly Detection Analysis Method for Preventing Phishing Fraud," REVIEW OF KIISC, Vol. 23, No. 6, pp. 41~48, Dec, 2013.
〔9〕 Si-wan Yoo, "Study on a Real Time Based Suspicious Transaction Detection and Analysis Model to Prevent Illegal Money Transfer Through E-Banking Channels," Journal of the Korea Institute of Information Security & Cryptology, Vol. 26, No. 6, pp. 1513~1526, Dec, 2016.
〔10〕 Eui-soon Choi and Kyung-ho Lee, "A Study on Improvement of Effectiveness Using Anomaly Analysis rule modification in Electronic Finance Trading," Journal of the Korea Institute of Information Security & Cryptology, Vol. 25, No. 3, pp. 615~625, Jun, 2015.
〔11〕 Jiyoung Woo, Hana Kim, Byung Il Kwak and Huy Kang Kim, "온An abnormal transaction detection model based on online game payment data analysis", REVIEW OF KIISC, Vol. 26, No. 3, pp. 38~44, Jun, 2016.
〔12〕 Hee Chan Han, Hana Kim and Huy Kang Kim, "Fraud Detection System in Mobile Payment Service Using Data Mining," Journal of the Korea Institute of Information Security & Cryptology, Vol. 26, No. 6, pp. 1527~1537, Dec, 2016.
〔13〕 Woo Young Moon and Soo Dong Kim, "Adaptive Framework for Detecting FinTech Frauds," KIISE Transactions on Computing Practices, Vol. 24, No. 7, pp. 337~344, Jul, 2018.
〔14〕 KISA, "Consumer protection in the era of online tracking: Suggestions for improving legislative framework", Aug. 2017.
〔15〕 Nick Nikiforakis, Alexandros Kapravelos "Cookieless monster : Exploring the ecosystem of web-based device fingerprinting", 2013 IEEE Symposium on Security and Privacy, pp.541-555, 2013.
〔16〕 Wikipedia, "HTTP cookie", https://en.wikipedia.org/wiki/HTTP_cookie
〔17〕 Wikipedia, "Web beacon", https://en.wikipedia.org/wiki/Web_beacon
〔18〕 Stefanie Olsen, "Nearly undetectable tracking device raises concern", CNET News, JAN. 2002.
〔19〕 Peter Eckersley, "How Unique Is Your Web Browser?", Electronic Frontier Foundation, 2010.

〔20〕 Mitchell Reichgut, "Advertiser ID Tracking And What It Means For You", Forbes, May. 2016.
〔21〕 Wikipedia, "Cross-device tracking", https://en.wikipedia.org/wiki/Cross-device_tracking
〔22〕 Allaboutcookies, "Mobile Technology Tracking Methods other than cookies", http://www.allaboutcookies.org
〔23〕 Threat Metrix, "Device fingerprinting and fraud protection whitepaper", ThreatMetrix.com
〔24〕 Iovation, "The power of device intelligence", Iovation.com
〔25〕 Threat Metrix, "ThreatMetrix ushers in the new era of trust and identity with ThreatMetrix ID - Digital Identity Summit 2017", ThreatMetrix.com, Sep. 2017.
〔26〕 Financesonline.com, "https://reviews.financesonline.com/p/iovation/" Financesonline.
〔27〕 Mozilla, "Introduction to the DOM", https://developer.mozilla.org/en-US/docs/Web/API/Document_Object_Model/Introduction
〔28〕 Gunes Acar & Christian Eubank, "The Web Never Forgets : Persistent Tracking Mechanisms in the Wild", CCS '14, pp.674-689, Nov 2014.
〔29〕 amiunique.org, "https://amiunique.org/stats", amiunique

〈저자소개〉

장 석 은 (Seok-Eun Jang) 정회원
2001년 2월: 성균관대학교 스포츠과학과 졸업
2005년 8월: 세종대학교 정보통신대학원 정보보호학과 석사
2018년 8월: 전남대학교 정보보안협동과정 박사과정 수료
〈관심분야〉 전자상거래보안, 모바일보안, 사이버프로파일링

박 순 태 (Soon-Tai Hong) 정회원
1992년 2월: 단국대학교 전자계산학과(학사)
1998년 8월: 국민대학교 정보과학대학원 정보통신학과(석사)
2010년 8월: 전남대학교 정보보안협동과정 박사
1994년 7월~1999년 9월: 육군 전산장교
2000년 4월~현재: 한국인터넷진흥원 정보보안운영팀
〈관심분야〉 정보보안 정책과 실무, 정보통신기반 보호

이 상 준 (Sang-Joon Lee) 정회원
1991년 전남대학교 전산통계학과(이학사)
1993년 전남대학교 전산통계학과(이학석사)
1999년 전남대학교 전산통계학과(이학박사)
1995년~2005년 서남대학교 경영전산정보학과 조교수
2005년~2007년 신경대학교 인터넷정보통신학과 조교수
2007년~현재 전남대학교 경영학부 교수
〈관심분야〉 경영정보시스템, 스마트컴퓨팅, 소프트웨어공학, 정보보호