

2021년 한국정보보호학회 영남지부 학술대회

2021년 9월 2일(목)

부산 벡스코 제2전시장 3층 회의실(325호, 326호)

Proceedings

주최 한국정보보호학회

주관 한국정보보호학회 영남지부

후원 공공기관: 한국남부발전(주), 한국동서발전(주)

기업: (주)범일정보, (주)포워즈시스템, (주)디아이섹, (주)스마트엠투엠, (주)이글루시큐리티, 쌍용정보통신(주)



한국정보보호학회
Korea Institute of Information Security & Cryptology

한국정보보호학회 영남지부에서는 ‘2021년 영남지부 학술대회’를 온라인 및 오프라인 병행으로 개최합니다. 이에 영남지부 학술대회 프로그램을 다음과 같이 안내드리오니, 많은 관심과 참여 부탁드립니다.

초대의 글

먼저 2021년 정보보호학회 영남지부 학술대회를 부산에서 개최하게 되어 매우 뜻깊게 생각하고 있습니다.

4차 산업혁명으로 시작한 사물인터넷, 빅데이터, 인공지능과 결합한 IT 기술은 이미 인터넷의 가상 공간을 벗어나 가정, 자동차, 공장 등 이미 우리 일상 속에 깊숙이 들어와 있습니다. 더구나 COVID-19로 촉발된 세계적인 팬데믹은 우리의 생활을 온라인 기반의 비대면 생활로 강제로 이행하게 하였습니다. 이에 대한 각종 위협과 공격은 기존 방식과는 새로운 양상으로 전개되면서, 정보보호의 중요성은 더욱 주목받고 있습니다. 향후 예상되는 포스트 코로나의 시대에서는 정보보호와 개인 프라이버시 보안은 반드시 선결되어야 할 최우선 과제라 해도 과언이 아닐 것입니다. 이러한 배경 아래에 우리 학술대회가 정보보호 각 분야에서 축적된 연구 결과를 한 자리에서 토론함으로써 최신 정보보호 분야의 동향과 정보를 교환하는 장이 되기를 기대합니다.

마지막으로 이번 2021년 한국정보보호학회 영남지부 학술대회가 성공적으로 개최될 수 있도록 힘써주신 최윤호 운영위원장님, 김지연 프로그램위원장님, 각 위원님, 좌장님들께 깊은 감사의 인사를 드립니다. 또한, 우수한 연구 논문을 발표해 주신 발표자님들과 학술대회 모든 참석자께도 감사의 말씀을 전합니다.

한국정보보호학회 영남지부
지부장 김창훈

■ 프로그램

▷ 논문발표 (325호)

시간	내용	좌장
09:30 – 10:00	등록 및 온라인 입장	
10:00 – 11:45	Session A : 콘텐츠 보안 및 융합보안	신인철 교수 (부경대)
11:45 – 13:00	점심식사	
13:00 – 14:45	Session B : 컴퓨터 보안 및 네트워크 보안	한성화 교수 (동명대)
14:45 – 15:00	휴식	
15:00 – 16:30	Session C : 블록체인 및 암호응용 기술	최필주 교수 (부경대)
16:30 – 16:40	휴식	
16:40 – 17:20	KIISC 영남지부 정기총회 사 회: 최윤호 교수 (부산대학교) 개회사: 류재철 회장 (한국정보보호학회) 환영사: 김창훈 지부장 (한국정보보호학회 영남지부) 시상식	

▷ 사이버보안 특강 (326호)

시간	트랙: 지자체 및 공공기관 사이버보안 역량 강화를 위한 특강
	[작장: 대구대학교 김창훈 교수]
10:00~10:50(50분)	망 분리 환경에서의 안전한 원격근무 지원을 위한 인프라 구성 방안 (대구대학교 김창훈 교수)
11:00~11:50 (50분)	클라우드 환경에서의 보안 기술 (부산대학교 김호월 교수)
11:50~13:00	점심식사
13:00~13:50 (50분)	미국 공급망 정책 추진 동향 소개 (국가보안기술연구소 김소정 사이버안보정책연구실장)
14:30~15:20 (50분)	성공사례 중심의 네이버 공공 클라우드 보안 서비스 소개 (네이버클라우드 전병선 클라우드보안수석 엔지니어)
15:30~16:20 (50분)	포스트 코로나 시대의 사이버 회복력 (국가보안기술연구소 강정민 사이버안전훈련센터장)
16:30~17:20(50분)	이 순간에도 발생하고 있는 사이버 침해사고 사례와 대응방안 (한국인터넷진흥원 이동연 침해사고 분석단 취약점분석팀장)

■ 등록 안내

▷ COVID-19 바이러스 확산으로 인해 현장 등록 없이 온라인 사전 등록으로 진행함

▷ 사전등록마감일 : 2021년 9월 1일(수)

▷ 등록비 및 등록 방법

회원 및 일반	대학원생	학부생
100,000원	50,000원	면제

* 학부생 논문의 경우 반드시 지도교수 등록 및 kiisc@kiisc.or.kr 로 학생증 사본 송부

- 학회 홈페이지(www.kiisc.or.kr)접속 ►학회행사 ► 사전등록 바로가기 ► 2021 영남지부 학술대회(계좌이체/신용카드 결제 가능)
- 무통장 입금 시
- 사전등록 송금처 : 국민은행 754201-04-135682 (예금주: (사)한국정보보호학회)
- 사전등록 시, 등록비는 위의 계좌로 송금, 입금자가 대리일 경우 통보요망
- 신용카드 결제시 계산서 발급 불가(부가가치세법 시행령 제57조)
- 사전등록 시, (2~3일 이내) 기재해주신 이메일로 청구용 계산서 발행 (영수용 계산서가 필요하신 경우 미리 학회로 연락 바랍니다.)
- 참가확인서는 kiisc@kiisc.or.kr로 행사명, 성명, 소속을 기재하시어 행사 종료 후 요청하시기 바랍니다.
- 학생은 다른 소속이 없는 전일제(학부생/대학원생)에 한합니다.
- 입금명은 필히 등록자 성함으로 기재해 주시기 바랍니다.
- 회사명으로만 기재하시면 입금 시 확인되지 않습니다.

▷ 등록 문의

- 문의처: 한국정보보호학회
- 전화 : 02-564-9333~4 (내선1)
- 이메일 : kiisc@kiisc.or.kr

■ 발표안내

- 이번 한국정보보호학회 영남지부 학술대회는 2021년 9월 1일-3일 온라인으로 진행되는 '네이버사이버보안 컨퍼런스'와 함께 개최됩니다.
모든 발표논문은 온라인 방송으로 실시간 송출되오니 안내에 따라 현장 발표 또는 온라인 발표를 준비해주시기 바랍니다.
(방송링크는 등록자에게 학술대회 개최 전 발송 예정입니다.)

- 현장 발표 : 학술대회 장소에서 배정된 시간에 발표
- 발표시간 : 논문 당 15분 (10분 발표, 5분 질의응답)
- 발표자료 템플릿 : 다운로드 (<https://bit.ly/2WkdNYW>)
※URL 접속 후 : File-Download-Microsoft PowerPoint(.pptx)
- 발표자료 제출 마감 : 2021년 8월 29일(일) ※기한 엄수
- 발표자료 제출처: kiisc2021@gmail.com

- 온라인 발표 : 발표 동영상을 사전 제작하여 제출
- 발표시간 : 12분-15분 분량으로 제작
- 발표자료 템플릿 : 자유양식
- 동영상 포맷: mp4
- 동영상 파일명: 발표자명(소속)
- 동영상 제출 마감 : 2021년 8월 29일(일) ※기한 엄수
- 동영상 제출처: kiisc2021@gmail.com

- 유의사항 : 동영상 녹화 후 소리 등이 재생 잘 되는지 확인 후 제출 부탁드립니다.

■ 행사 및 프로그램 관련 문의

- 한국정보보호학회 영남지부장: 대구대학교 김창훈 교수 (kimch@daegu.ac.kr)
- 운영위원장: 부산대학교 최윤호 교수 (yhchoi@pusan.ac.kr)
 - 운영위원: 권동현(부산대), 김호원(부산대), 신욱(동명대), 이석환(동아대), 백남균(부산외국어대학교), 신상욱(부경대)
- 프로그램위원장: 대구대학교 김지연 교수 (jyk@daegu.ac.kr)
 - 운영위원: 정임영(경북대), 윤종희(영남대), 이경률(대 구가톨릭대), 정기현(경일대)

Session A-콘텐츠 보안 및 융합보안

좌장 : 신인철 교수 (부경대)

9월 2일(목), 10:00 - 11:45

A-1.	무인 이동체에 대한 물리 계층 공격에 관한 연구	1
	강정환, 권동현 (부산대학교)	
A-2.	오픈소스 파일 스캐닝 도구 성능 평가 – 기능 및 메타데이터 기반으로.....	5
	서영민, 정원태, 이경률 (대구가톨릭대학교)	
A-3.	저품질 이미지에서 확장 공간 통계 정보와 텍스처 연산자를 이용한 위조 탐지 연구.....	11
	사우랍 아가왈 (아미티 대학교, 경일대학교), 정기현 (경일대학교)	
A-4.	정보보안 관리실태 평가를 위한 블록체인 기반 통합보안로그 관리 시스템	15
	김도훈, 김호원 (부산대학교)	
A-5.	적대적 예제를 이용한 디지털 워터마킹.....	18
	윤영여, 심준석, 조현진, 강효은, 김호원 (부산대학교)	
A-6.	iOS 부트로더 취약점을 통한 펌웨어 분석 방법 연구.....	21
	김성민, 류재철 (충남대학교)	
A-7.	소셜 네트워크 서비스의 이미지를 통한 개인정보 유출 위험성 연구.....	25
	권희원, 김명주 (서울여자대학교)	
A-8.	안전한 자율협력주행을 위한 C-V2X 시스템 구축 방안 연구	28
	김서연, 오인수, 임강빈 (순천향대학교)	

Session B-컴퓨터 보안 및 네트워크 보안

좌장 : 한성화 교수 (동명대)

9월 2일(목), 13:00 - 14:45

B-1.	클라우드 웹 애플리케이션을 위한 SVM 기반 데이터베이스 이상 탐지 연구.....	31
	조재한, 김지연 (대구대학교)	
B-2.	WLAN IoT 보안을 위한 실시간 이상 탐지 시스템	35
	이승옥, 김지연 (대구대학교)	
B-3.	정보보호 제품의 자가보호 기법 분석.....	40
	이성원, 윤종희 (영남대학교)	
B-4.	코드 재사용 공격 방어를 위한 제어 흐름 무결성 검증 기법 연구	43
	여기수, 권동현 (부산대학교)	
B-5.	접근제어를 지원하는 Tagged Memory Extension 동향	48
	이진재, 데리 프라타마, 권동현, 김호원 (부산대학교)	
B-6.	제로 트러스트관점의 보안체계	52
	고민혁, 이대성 (부산가톨릭대학교)	
B-7.	딥러닝 기반 랜섬웨어 분류를 위한 시각화 기법 연구.....	55
	이수경, 최은정 (서울여자대학교)	
B-8.	망 분리 환경에서의 안전한 원격근무 지원을 위한 인프라 구성 방안.....	58
	박종현, 김창훈 (대구대학교), 권상오, 박성수 (포워즈시스템)	

Session C-블록체인 및 암호응용 기술

좌장 : 최필주 교수 (부경대)

9월 2일(목), 15:00 - 16:30

C-1.	블록체인을 활용한 후불식 톤게이트 지불 시스템 구현.....	63
	박재훈, 권혁동, 서화정 (한성대학교)	
C-2.	RE100 실현을 위한 블록체인 기반 REC 거래 플랫폼	66
	김재석, 황보규민, 최윤호 (부산대학교)	
C-3.	단순 전력 분석 공격에 강인한 타워 곡선 point multiplication 방법.....	69
	최필주 (부경대학교)	
C-4.	부채널 내성 딥러닝 네트워크 동향.....	72
	권혁동, 박재훈, 심민주, 서화정 (한성대학교)	
C-5.	Quantum Information Set Decoding 연구동향.....	75
	장경배, 송경주, 오유진, 서화정 (한성대학교)	
C-6.	블록체인-IoT-클라우드 기반 수산 식품 공급망 모델 제안.....	78
	전미현, 조강우, 신상욱 (부경대학교)	
C-7.	TOR 브라우저에서 비트코인을 사용한 전략물자 불법 거래 내역 추적 방안 연구	81
	선하라, 윤지원 (고려대학교)	
C-8.	DID 지갑을 위한 하드웨어 키 관리 방안 연구.....	84
	Zhuohao Qian, 이경현 (부경대학교)	

무인 이동체에 대한 물리 계층 공격에 관한 연구

강정환*, 권동현†

*부산대학교 (대학원생)

A Survey of Physical Layer Attacks on Unmanned Vehicles

Jeong-Hwan Kang*, Dong-Hyun Kwon†

*Pusan National University(Graduate student)

요약

무인 이동체는 무인 항공기, 무인 자동차, 무인 선박 등으로 분류되며, 사람이 장시간에 걸쳐 수행해야 하는 일을 대신할 수 있어 우리의 삶에 깊이 관여하고 있다. 지능형 시스템을 기반으로 한 무인 이동체는 다양한 사이버 공격에 취약하다. 본 논문에서는 무인 이동체에 대한 세 가지 계층 공격 중 하나인, 물리 계층 공격에 대한 연구들을 분석하였다. 이 중에서도 특히 무인 항공기와 무인 자동차라는 두 가지 대상에 대한 공격에 관한 연구를 중심으로 분석하였다.

I. 서론

인공지능과 무선통신 기술의 발달로 다양한 분야에서 무인 이동체가 연구되고 있다. 크게 무인 항공기, 무인 자동차, 무인 선박 등으로 분류되는 무인 이동체는 사람이 장시간에 걸쳐 수행해야 하는 일을 대신할 수 있어 우리의 일상 생활에 깊이 관여하고 있다. 무인 이동체로 인한 다양한 유형의 사고는 인명피해와 재산적 피해를 유발한다. 따라서 무인 이동체의 보안성과 안전성은 상당히 중요한 문제이다.

특히, 지능형 시스템을 기반으로 한 무인 이동체는 GPS 재밍, 초음파 센서 공격, 카메라 센서 공격과 같은 일반적인 사이버 공격에 취약하다[1].

본 논문에서는 무인 이동체에 대한 세 가지 계층 공격 중 하나인 물리 계층 공격에 대한 연구를 분석하였다. 이 중에서도 특히 무인 항공기와 무인 자동차라는 두 가지 대상에 대한 공격에 관한 연구를 중심으로 분석하였다.

II. 무인 이동체의 공격 모델

일반적인 유인 이동체와 비교했을 때, 무인 이동체는 보안 문제에 직면할 가능성이 더 크다. 악의적인 목적을 위해 무인 이동체의 시스템에 침입하려는 공격은 일반적으로 기술적 공격으로 볼 수 있다. 이러한 기술적 공격은 물리 계층, 통신 계층, 애플리케이션 계층의 세 가지 계층을 통해 구현될 수 있다[2].

물리 계층 공격은 무인 이동체의 센서 및 액추에이터에 중점을 둔 공격이다. 대다수의 무인 이동체는 작업을 탐색하고 수행하기 위해 센서에 의존하며, 액추에이터를 통해 제어 명령을 정확하게 수행한다. 센서 공격은 센서 스푸핑 공격과 물리적 간섭 공격으로 분류할 수 있다. 센서 스푸핑 공격은 시스템이 잘못된 정보를 인식하게 함으로써 잘못된 결정을 내리도록 한다. 물리적 간섭 공격은 센서의 정상적인 기능을 방해하거나 손상시킬 수 있으며, 이는 시스템 충돌로 직접 이어질 수도 있다.

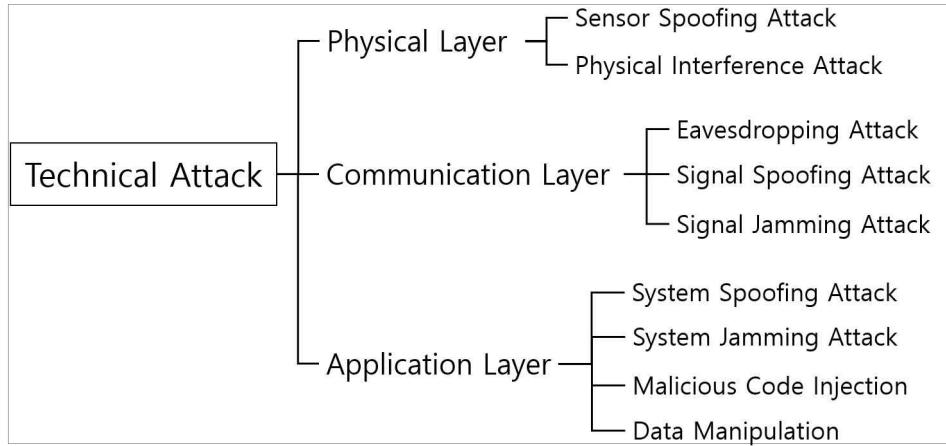


그림 1. 무인 이동체에 대한 계층별 공격[2]

무인 이동체의 통신 계층은 시스템 간의 상호 작용을 보장하지 위해 무선 통신을 사용한다. 그러나 무선 링크의 브로드캐스트 환경으로 인해 무인 이동체의 무선 통신은 공격자에 의해 악용될 가능성이 높다. 따라서 기존의 지상 통신 시스템보다 더 심각한 보안 문제가 발생 한다. 통신 계층 공격은 크게 도청, 신호 스피핑, 신호 재밍으로 분류된다.

무인 이동체의 애플리케이션 계층은 데이터 처리 및 의사 결정을 담당한다. 애플리케이션 계층에 대한 공격 방법에는 시스템 스피핑, 시스템 재밍, 악성코드 주입, 데이터 조작 등이 있다. 통신 계층 공격과 달리 응용 계층에서의 스피핑과 재밍은 특정 애플리케이션에 초점을 맞추고, 통신 계층에서는 이러한 공격은 주로 신호를 전달하는 것을 목표로 한다. 애플리케이션 계층에 대한 공격은 성능 저하 또는 시스템의 악의적인 제어를 유발할 수 있다.

그림 1은 무인 이동체에 대한 기술적 공격을 계층 별로 분류한 것이다.

III. 물리 계층 공격 연구사례

3.1. Sensor Spoofing Attack

무인 항공기에 대한 센서 스피핑 공격으로서 Davidson et al. [3]에서 제안된 광학 흐름 센서 (Optical Flow Sensor)에 대한 입력 스피핑 공격이 있다. 해당 연구에서는 광학 흐름 감지 및 측정을 위해 널리 쓰이는 Lucas-Kanade 기법의 특성을 이용해, 광학 흐름 센서의 입력을 스

피핑하여 광학 흐름을 변조함으로써 무인 항공기를 제어했다. 또한, Random Sample Consensus (RANSAC) 알고리즘을 통해 방어하는 방법을 소개했다.

실제 환경에서 수행된 연구로서 Gluck et al. [4]에서 제안된 초음파 거리 센서(Ultrasonic Distance Sensor)에 대한 스피핑 공격이 있다. 해당 연구에서는 무인 항공기의 초음파 거리 센서에서 발견된 취약점을 악용하여, 입력의 일정한 노이즈에 의해 생성되는 가짜 팰스를 주입함으로써 무인 항공기로 하여금 가짜 장애물을 인식시켰다. 이를 통해 적절한 동작을 방해하고 비행 경로를 제한시켰다. 이때 이들이 제안한 기법은 센서의 상대적 위치나 타이밍 특성에 대한 사전 지식이 필요하지 않다는 장점이 있다.

무인 자동차에 대한 센서 스피핑 공격으로서 Cao et al. [5]에서 제안된 LiDAR 공격이 있다. 해당 연구에서는 기존 연구에서 수행된 LiDAR 스피핑 공격을 재현하고, 이를 통해 기존 연구보다 더욱 정밀하게 제어된 신호를 LiDAR 시스템에 주입하여, 가상의 물체를 인식시키는 정교한 공격을 수행했다. 이들은 최적화 및 글로벌 샘플링을 결합한 알고리즘을 설계하여 공격 성공률을 약 75%까지 향상시켰다.

B. Nassi et al. [6]에서는 Mobileye 630 PRO와 Tesla Model X, HW 2.5의 ADAS와 Autopilot에서 깊이가 없는 물체(Phantom)를 실제 물체처럼 인지하는 취약점을 악용하였다. 이

Focus	Ref.	Method		Attack Surface	Impact
		Sensor Spoofing	Physical Interference		
UAV	[3]	✓		Optical Flow Sensor	be controlled
	[4]	✓		Ultrasonic Distance Sensor	be controlled
	[7]	✓		MEMS Gyroscope	lose control
AV	[5]	✓		LiDAR	be controlled or freeze
	[6]	✓		ADAS; Autopilot	be controlled
	[8]	✓		Ultrasonic Sensor	inaccuracy in detecting obstacles
	[9]	✓		Anti-lock Braking System	be controlled or disrupted
	[10]	✓	✓	MMW Radars; Ultrasonic Sensors; Cameras	be controlled or jammed
	[11]	✓	✓	LiDAR	be spoofed or blind

표 1. 무인 이동체에 대한 물리 계층 공격의 기준 연구 분석

들은 휴대용 프로젝터가 장착된 드론을 통해 phantom을 도로 위나 가로수 등에 투사시켰고, 위의 시스템이 적용된 무인 자동차는 이를 실제 물체로 인지하였다. 이로 인해 브레이크가 오작동되고 무인 자동차의 주행 차선이 임의로 변조되었다.

3.2. Physical Interference Attack

무인 항공기에 대한 물리적 간섭 공격으로서 Son et al. [7]에서 제안된 MEMS Gyroscopic Sensor에 대한 물리적 간섭 공격이 있다. 해당 연구에서는 의도적인 노이즈를 사용하여 MEMS 자이로스코프를 무력화시켰다. 이들은 15종의 MEMS 자이로스코프를 분석하여, 총 7 개의 공진 주파수를 발견했다. 이들은 실험을 통해 공진 주파수를 사용하여 대상 무인 항공기의 자이로스코프를 무력화시켜 추락시켰다.

무인 자동차에 대한 물리적 간섭 공격으로서 Lim et al. [8]에서 제안된 초음파 센서에 대한 물리적 간섭 공격은 실생활에서 일어날 수 있

는 네 가지 공격 시나리오를 설정하여, 이를 실험 환경에서 수행되었다. 특히, 두 개의 초음파 센서가 마주 보는 시나리오에서의 물리적 간섭은 초음파 센서의 탐지 거리가 실제보다 매우 짧아졌다. 이들은 무인 자동차의 초음파 센서가 상대 차량의 초음파 센서와 마주 보는 경우, 서로의 간섭으로 인해 문제가 발생할 수 있다는 것을 확인하였다.

표 1은 무인 이동체에 대한 물리 계층 공격을 수행한 기준 연구를 분석 및 정리한 것이다. 표에서 UAV(Unmanned Aerial Vehicle)는 무인 항공기, AV(Autonomous Vehicle)은 무인 자동차이다.

IV. 결론

본 논문에서는 무인 이동체의 세 가지 공격 계층 중 하나인, 물리 계층 공격에 대한 기준 연구를 분석하였다. 무인 이동체에 탑재된 센서들이 다양한 취약점으로 인해 물리 계층 공격의 대상이 된다. 또한, 이러한 물리 계층 공격

은 단순히 대상 시스템을 무력화시키는 것뿐만 아니라, 공격자가 의도한 방식으로 대상 시스템의 제어를 탈취할 수 있다는 것을 확인하였다. 무인 이동체가 우리의 삶에 더욱 깊이 관여하고 있으므로 악의적인 공격으로 인한 인명피해 감소와 개인정보 보호를 위해 물리 계층을 비롯한 다양한 계층에서의 취약점 분석과 공격 및 방어 연구가 많이 수행되어야만 한다.

Acknowledgement

"본 논문은 2021년 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임"(IITP-2019-0-01343)

[참고문헌]

- [1] I. Yaqoob, L. U. Khan, S. M. A. Kazmi, M. Imran, N. Guizani and C. S. Hong, "Autonomous Driving Cars in Smart Cities: Recent Advances, Requirements, and Challenges," in IEEE Network, vol. 34, no. 1, pp. 174-181, January/February 2020, doi: 10.1109/MNET.2019.1900120.
- [2] Y. Tan, J. Wang, J. Liu and Y. Zhang, "Unmanned Systems Security: Models, Challenges, and Future Directions," in IEEE Network, vol. 34, no. 4, pp. 291-297, July/August 2020, doi: 10.1109/MNET.001.1900546.
- [3] Davidson, D., Wu, H., Jellinek, R., Singh, V., & Ristenpart, T. (2016). Controlling UAVs with sensor input spoofing attacks. In 10th {USENIX} Workshop on Offensive Technologies ({WOOT} 16).
- [4] Gluck, Tomer, et al. "Spoofing Attack on Ultrasonic Distance Sensors Using a Continuous Signal." Sensors 20.21 (2020): 6157.
- [5] Cao, Yulong, et al. "Adversarial sensor attack on lidar-based perception in autonomous driving." Proceedings of the 2019 ACM SIGSAC conference on computer and communications security. 2019.
- [6] Nassi, B., Nassi, D., Ben-Netanel, R., Mirsky, Y., Drokin, O., & Elovici, Y. (2020). Phantom of the ADAS: Phantom Attacks on Driver-Assistance Systems. IACR Cryptol. ePrint Arch., 2020, 85.
- [7] Son, Yunmok, et al. "Rocking drones with intentional sound noise on gyroscopic sensors." 24th {USENIX} Security Symposium ({USENIX} Security 15). 2015.
- [8] B. S. Lim, S. L. Keoh and V. L. L. Thing, "Autonomous vehicle ultrasonic sensor vulnerability and impact assessment," 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), 2018, pp. 231-236, doi: 10.1109/WF-IoT.2018.8355132.
- [9] Shoukry, Yasser, et al. "Non-invasive spoofing attacks for anti-lock braking systems." International Conference on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, 2013.
- [10] Yan, Chen, Wenyuan Xu, and Jianhao Liu. "Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle." Def Con 24.8 (2016): 109.
- [11] Shin, Hocheol, et al. "Illusion and dazzle: Adversarial optical channel exploits against lidars for automotive applications." International Conference on Cryptographic Hardware and Embedded Systems. Springer, Cham, 2017.

오픈소스 파일 스캐닝 도구 성능 평가: 기능 및 메타데이터 기반으로

서영민¹, 정원태², 이경률^{1,*}

¹대구가톨릭대학교 컴퓨터소프트웨어학부

²대구가톨릭대학교 컴퓨터소프트웨어학과

*교신저자

Performance Analysis of Open Source File Scanning Tools: based on Features and File Metadata

Yeongmin Seo¹, Won-tae Jung², Kyungroul Lee^{1,*}

¹School of Computer Software, Daegu Catholic University

²Department of Computer Software, Daegu Catholic University

*Corresponding author

요약

본 논문에서는 스팸 메일이나 스파이 피싱 등을 통하여 기업이나 일반 사용자들이 사용하는 단말의 자원을 소비하거나 정보를 탈취하는 악성코드를 탐지하기 위하여, 악성코드 탐지에 사용되는 파일 정보 및 메타데이터를 추출하는 오픈소스 기반의 파일 스캐닝 도구인 Strelka, FSF, Laika BOSS, StoQ의 성능을 비교하고 분석한다. 본 논문에서는 악성코드가 실행파일 형태로 배포되는 것을 고려하여, 윈도우즈 실행파일인 EXE 파일을 대상으로, 파일 종류 및 메타데이터 종류에 따른 성능을 분석하였으며, 결과적으로 도구가 제공하는 기능에 따른 성능 분석 결과는 스캐닝하는 파일의 다양성 측면에서 도구를 활용하는 사용자는 StoQ가 적합할 것으로 판단되고, 확장성 측면에서 도구를 활용하는 사용자는 Strelka가 적합할 것으로 판단된다. 마지막으로, 추출 가능한 메타데이터 종류에 따른 성능 분석 결과는 FSF, Strelka, StoQ, Laika Boss 순으로 성능이 좋은 것으로 분석된다.

키워드: 네트워크 IDS, 악성코드, 메타데이터, 파일 스캐닝 도구

I. 서론

최근 4차 산업혁명으로 인하여 정보통신기술이 많이 발전되는 추세이다. 하지만, 그와 동시에 공격 기법 또한 발전하고 다양한 종류의 공격이 등장하였으며 [1], 이에 대한 대비가 필요하다. 다양한 공격 기법의 하나로, 악성코드를 이용하여 피해자들의 PC나 스마트폰과 같은 단말의 자원을 낭비하게 하거나 정보를 탈취하는 공격이 있으며, 이러한 공격을 위하여 공격자는 스팸 메일이나 스파이

피싱 등을 이용하여 악성코드를 메일에 첨부함으로써 피해자 시스템으로 침투한다 [2]. 하지만, 일반적인 네트워크 침입탐지 시스템(IDS, Intrusion Detection System)에서는 메일에 첨부된 파일의 악성 여부를 검증하지 못한다 [3]. 이에 메일에 첨부된 파일의 다양한 정보를 추출함으로써 악성코드를 탐지하는 방안이 요구되었으며, 이를 위한 도구인 파일 스캐닝 도구가 등장하였다. 파일 스캐닝 도구는 메일에 첨부된

파일의 메타데이터를 추출하며, 메타데이터를 기반으로 악성코드 탐지가 가능하다.

상기와 같이 메일에 첨부된 악성코드를 탐지하기 위하여, 오픈소스 기반의 다양한 파일 스캐닝 도구들이 등장하였지만, 각 도구에서 제공하는 기능이 상이한 문제점이 존재한다. 특히, 악성코드를 탐지하기 위한 중요한 정보인 파일 형식(file format) 정보와 메타데이터의 종류가 도구마다 다르며, 이는 악성코드 탐지 성능과 밀접한 관계를 가진다.

따라서, 본 논문에서는 악성코드 탐지와 관련된 성능을 평가하기 위한 목적으로, 파일 종류 및 메타데이터를 기준으로, 대표적인 파일 스캐닝 도구인 Strelka, FSF(File Scanning Framework), Laika BOSS, StoQ의 성능을 분석하고 평가한다.

II. 관련 연구

관련 연구에서는 대표적인 파일 스캐닝 도구인 Strelka, FSF, Laika BOSS, StoQ를 기능 및 특징을 서술한다.

2.1 Strelka

Strelka는 파일 및 메타데이터 추출을 목적으로 개발된 도구이며, 파일 분석 및 파일 식별을 토대로 YARA 시그니처를 활용하여 위협을 탐지한다 [8]. 2018년 9월 26일에 처음 오픈소스로 공개되었으며, 마지막으로 2021년 5월 17일에 업데이트되었다 [4].

2.2 FSF

FSF는 모듈식 파일 검색 솔루션으로, 파일을 스캔하여 메타데이터를 추출하며, YARA 시그니처를 활용하여 분석하는 것이 가능하다 [8]. 2015년 8월 6일에 오픈소스로 공개되었고, 2019년 1월 29일에 마지막으로 업데이트되었다 [5].

2.3 Laika BOSS

Laika BOSS는 YARA [8]와 ZeroMQ [11]와 같은 오픈소스를 기반으로 개발된 파일 스캐너 및 침입 탐지 시스템으로, 다양한 네트워크 프로토콜, 캡슐화 및 난독화를 추상화하여 스캔한다. 2015년 7월 29일 공개 이후로, 2018년 9월 12일에 마지막으로 업데이트되었다 [6].

2.4 StoQ

StoQ는 플러그인 기반의 프레임워크로, 파일을 스캐닝하는 방법을 정의한 플러그인을 기반으로 YARA [8]나 TrID 및 정적 속성을 통하여 파일 정보를 추출한다. 다양한 플러그인 간 통신을 조율하며, 표준화된 결과를 제공한다. 2015년 공개 이후로, 2020년 7월 28일에 마지막으로 업데이트되었다 [7].

III. 오픈소스 파일 스캐닝 도구 성능 평가

본 논문에서는 파일 스캐닝 도구가 제공하는 기능에 따른 성능 평가 결과와 도구에서 추출 가능한 메타데이터 종류에 따른 성능 평가 결과를 서술한다.

3.1 기능별 성능 평가

파일 스캐닝 도구는 도구마다 다른 구현 방법으로 인하여, 추출 가능한 정보도 다르다. 따라서 본 논문에서는 도구에서 제공하는 기능에 따른 성능을 평가하고자 한다. 실험환경은 Intel(R) Core(TM) i5-5200U CPU 2.20GHz와 RAM 16GB가 장착된 컴퓨터에서 Ubuntu 18.04.5 LTS 운영체제에 각 도구를 설치하였다.

우선, Strelka는 YARA hashing을 기반으로 탐지를 수행하며, FSF는 FSF 서버에 정의한 YARA rule을 업로드한 후, 서버에서 스캔할 파일을 지정하여 파일을 스캐닝한다. Laika

BOSS는 자체 프레임워크를 사용하며, 스캐닝하고자 하는 방식에 해당하는 자체 모듈들을 사용하여 파일을 스캐닝한다. StoQ는 확장자를 분석하는 ExifTool이나 TrID와 같이 스캔하려는 목적에 부합하는 플러그인들을 다운로드한 후, 스캔을 수행한다.

악성코드 탐지를 위한 성능 평가를 위하여, 본 논문에서는 각 도구를 비교하는 기능으로, 위장 파일 식별, 사용하는 플러그인, 식별 가능 확장자를 정의하였다. 위장 파일 식별 기능이란, 공격자가 악성코드인 실행파일의 탐지를 우회하기 위한 목적으로 변경된 확장자를 탐지하는 기능이다. 예를 들면, PDF 파일이나 JPEG 파일로 위장한 악성코드인 EXE 파일의 원래 확장자를 탐지하는 기능이다. 식별 가능 확장자는 도구가 식별할 수 있는 확장자의 개수를 의미하며, 사용하는 플러그인은 제공되는 기능과 관련되는 것으로 판단하여 플러그인의 개수를 나타내었다. 기능별 성능 분석 결과를 [표 1]에 나타내었다.

[표 1] 기능별 성능 분석 결과

도구명	위장 파일 식별 여부	식별 가능 확장자 개수	플러그인 (모듈) 개수
Strelka	O	195개	56개
FSF	O	15개	22개
Laika BOSS	O	195개	32개
StoQ	O	14,250개	19개

위장 파일 식별 여부 기능에 따른 성능 분석 결과를 살펴보면, 모든 도구가 위장 파일을 식별하므로, 동일한 성능을 가지는 것으로 판단된다. 식별 가능한 확장자 수는 도구에서 식별할 수 있는 파일의 확장자와 관련된 기능으로, 개수가 많을수록 다양한 파일 형식을 분석할 수 있으며, 이는 높은 성능을 가질 것으로 판단된다. 결과를 살펴보면, FSF가 약 15개로 가장 성능이 낮은 것으로 판단되며,

Strelka와 Laika BOSS가 약 195개로 그다음, StoQ가 약 14,250개로 가장 성능이 높은 것으로 판단된다. Strelka와 Laika BOSS는 파일 형식을 식별하기 위하여 내부적으로 ExifTool을 활용하기 때문에 같은 성능을 가지는 것으로 나타났다. StoQ는 TrID와 ExifTool을 모두 사용함으로써 가장 많은 확장자를 식별할 수 있다. 하지만, TrID는 다수의 사용자로부터 정의된 확장자로, 신뢰성을 확보하기 위한 방안이 요구된다는 단점이 존재한다.

플러그인 개수는 도구에 추가할 수 있는 플러그인의 개수를 의미하며, 활용 가능한 기능 및 확장성과 관련된 성능으로 판단하였다. 결과를 살펴보면, StoQ가 19개로 가장 적은 개수를 가지며, 그다음으로 FSF와 Laika BOSS가 각각 22개, 32개를 가진다. 가장 많은 개수를 가지는 도구인 Strelka는 56개로 기능 추가 및 확장성 측면에서 성능이 높은 것으로 판단된다.

이를 종합할 때, 모든 도구는 위장 파일을 식별하기 때문에 동일한 성능을 제공하지만, 식별 가능한 확장자 수와 플러그인 개수에 따라, FSF가 가장 낮은 성능을 가지는 것으로 나타났으며, 식별 가능한 확장자 개수만을 고려한다면 StoQ, 플러그인 개수만을 고려한다면 Strelka가 가장 성능이 높은 것으로 나타났다. 따라서 스캐닝하는 파일의 다양성 측면에서 도구를 활용하는 사용자는 StoQ 도구가 적합할 것으로 판단되고, 확장성 측면에서 도구를 활용하는 사용자는 Strelka 도구가 적합할 것으로 판단된다.

3.2 메타데이터별 성능 평가

실질적으로 악성코드를 탐지하기 위해서는 파일의 정보인 메타데이터가 필수적으로 요구되며, 이는 악성코드 탐지 성능에도 밀접한 관계가 있으므로 각 도구에서 추출 가능한 메타데이터를 기반으로 성능을 평가하였다. 실험은 실제 악성코드의 파일 형식인 EXE

파일을 대상으로 메타데이터를 추출하였으며, 실험에 활용한 파일은 이미지 파일의 메타데이터를 추출하는 프로그램인 kuso exif viewer 설치파일을 활용하였다.

각 도구에서 제공하는 메타데이터의 종류를 확인하기 위하여, 모든 도구에서 추출되는 메타데이터를 수집하고 각 데이터가 가지는 속성에 따라 [표 2]와 같이 분류하였다.

[표 2]를 살펴보면, 도구에 따라 추출되는

메타데이터의 속성이 다를 뿐만 아니라, 속성에 따른 세부 메타데이터도 상이한 것을 확인할 수 있다. 본 논문에서 메타데이터의 속성 분류는 EXE 파일만을 실험에서 다루었기 때문에, 총 4가지인 PE(Portable Executable) 관련, 파일속성 관련, 개발 관련, 도구 관련으로 분류하였다. PE 관련은 PE Explorer [9]와 PEID [10]에서 나타나는 항목을 포함하며, 파일속성 관련은 일반적으로 파일에 포함되는 정보인 파일 이름이나 확장자 등을 포함한다.

[표 2] 분류한 메타데이터 종류

도구명	메타 데이터 속성	세부 메타데이터	도구명	메타 데이터 속성	세부 메타데이터
Strelka	파일속성 관련	Filename, FileTypeExtension	Laika Boss	파일속성 관련	FileType, Filename, FilePermissions, FileAccessDate, FileModifyDate, FileTypeExtension, FilenodeChangeDate
	PE 관련	EntryPoint, CodeSize, TimeStamp, ImageVersion, LinkerVersion, (Un)initializedContentSize, Subsystem, FileSize		PE 관련	OSVersion, EntryPoint, ImageVersion, TimeStamp, (Un)initializedContentSize, PEType, LinkerVersion, CodeSize, FileSize
	도구 관련	MIMEType		도구 관련	MIMEType, ExifToolVersion
FSF	파일속성 관련	Filename	StoQ	파일속성 관련	FileType, Filename, FileTypeExtension
	PE 관련	ImageBase, EntryPoint		PE 관련	EntryPoint, OSVersion, ImageVersion, TimeStamp, CodeSize, LinkerVersion, (Un)initializedContentSize, PEType, FileSize, Subsystem(Version)
	개발 관련	Compiled, Architecture		개발 관련	MIMEType, ExifToolVersion

개발 관련은 개발 도구와 관련된 항목인 architecture와 compiled를 포함하며, 도구 관련은 파일 스캐닝 도구에서 활용하는 정보인 ExifToolVersion 등을 포함한다.

도구별 결과를 살펴보면, Strelka는 19가지 메타데이터를 추출하였고, FSF는 5가지, Laika Boss는 28가지, StoQ는 21가지의 메타데이터를 추출하였다. 이를 토대로, 가장 적은 메타데이터 종류를 추출하는 FSF가 가장 성능이 낮은 것으로 나타났으며, 가장 많은 메타데이터 종류를 추출하는 Laika Boss가 가장 성능이 높은 것으로 나타났다.

상기와 같이 도구마다 추출하는 메타데이터의 속성과 종류가 다르며, 이는 파일 정보 추출 및 악성코드 탐지의 성능과 밀접한 관계가 있을 것으로 판단된다. 따라서 모든 도구에서 추출하는 메타데이터를 기반으로 각 도구에서 추출하는 메타데이터를 시각화하여 <그림 1>에 나타내었다.

결과를 살펴보면, 추출 가능한 전체 메타데이터에서 각 도구가 추출하는 메타데이터에 색을 부여하였으며, 시각적으로 확인한 결과, FSF, Strelka, StoQ, Laika Boss 순으로 성능이 좋은 것으로 판단된다.

IV. 결 론

본 논문은 악성코드 탐지에서 활용되는 오픈소스 기반의 파일 스캐닝 도구인 Strelka, FSF, Laika BOSS, StoQ의 성능을 비교하고 분석하였다. 이러한 파일 스캐닝 도구는 도구마다 다른 구현 방법으로 인하여 제공되는 기능이나 추출 가능한 정보가 다른 문제점이 있으며, 이는 악성코드를 탐지하는데 밀접한 관계를 가질 것으로 판단된다. 이러한 이유로, 본 논문에서는 파일 스캐닝 도구가 제공하는 기능에 따른 성능과 도구에서 추출 가능한 메타데이터 종류에 따른 성능을 비교하고 분석하였다.

감사의 글

1. 이 논문은 ETRI부설연구소의 위탁연구과제 [2021-011]로 수행한 연구결과입니다.
2. 이 논문 내용을 발표하는 때에는 ETRI부설 연구소에서 수행한 위탁결과임을 밝혀야 합니다.

strelka	fsf	Laika boss	stoq
포맷 Metadata 폴더 filetype filename filePermissions fileAccessDate initializedDataSize fileModifyDate filesize machineType fileType uninitializedDataSize imageVersion fileTypeExtension osVersion compiled architecture pType timestamp linkerVersion subSystem version subSystemVersion codeSize fileNodeChangeDate exifToolVersion imageBase ssdeep SHA256 SHA512 SHA1 md5	포맷 Metadata 폴더 filetype filename filePermissions fileAccessDate initializedDataSize fileModifyDate filesize MachineType FileType UninitializedDataSize ImageVersion FileTypeExtension osVersion Compiled Architecture PType Timestamp LinkerVersion SubSystem Version SubSystemVersion CodeSize FileNodeChangeDate ExifToolVersion ImageBase ssdeep SHA256 SHA512 SHA1 md5	포맷 Metadata 폴더 filetype filename filePermissions fileAccessDate initializedDataSize fileModifyDate filesize MachineType FileType UninitializedDataSize ImageVersion FileTypeExtension osVersion Compiled Architecture PType Timestamp LinkerVersion SubSystem Version SubSystemVersion CodeSize FileNodeChangeDate ExifToolVersion ImageBase ssdeep SHA256 SHA512 SHA1 md5	포맷 Metadata 폴더 filetype filename filePermissions fileAccessDate initializedDataSize fileModifyDate filesize MachineType FileType UninitializedDataSize ImageVersion FileTypeExtension osVersion Compiled Architecture PType Timestamp LinkerVersion SubSystem Version SubSystemVersion CodeSize FileNodeChangeDate ExifToolVersion ImageBase ssdeep SHA256 SHA512 SHA1 md5

<그림 1> 도구별 추출된 상세 메타데이터 시각화

[참고문헌]

- [1] 조성혜, 이상진, “MS 오피스 문서 파일 내 비정상 요소 탐지 기법 연구”, 정보처리학회 논문지, 컴퓨터 및 통신시스템, 제6권, 제2호, pp. 87-94, 2017.
- [2] 김목정, 이상진, “DISC 성격 유형과 사이버 보안 위협 간의 상호 연관성에 관한 연구”, 정보보호학회논문지, 제29권, 제1호, pp. 215-223, 2019.
- [3] 이태호, “침입탐지 시스템과 샌드박스를 이용한 악성파일 탐지 시스템 구현”, 충실대학교 정보 과학대학원, 국내석사학위논문, pp. 27-28, 2017.
- [4] Strelka, <https://target.github.io/strelka/#/>
(Retrieved from Jul. 7, 2021)
- [5] File Scanning Framework, <https://github.com/EmersonElectricCo/fsf> (Retrieved from Jul. 7, 2021)
- [6] LaikaBOSS, <https://github.com/lmco/laikaboss>
(Retrieved from Jul. 7, 2021)
- [7] StoQ, <https://stoq-framework.readthedocs.io/en/latest/index.html> (Retrieved from Jul. 7, 2021)
- [8] YARA, <https://yara.readthedocs.io/en/stable/>
(Retrieved from Jul. 7, 2021)
- [9] PE Explore, <http://www.pe-explorer.com/>
(Retrieved from Jul. 7, 2021)
- [10] PEiD, <https://www.softpedia.com/get/Programming/Packers-Crypters-Protectors/PEiD-updated.shtml> (Retrieved from Jul. 7, 2021)
- [11] ZeroMQ, <https://zeromq.org/get-started/>
(Retrieved from Jul. 8, 2021)

저품질 이미지에서 확장 공간 통계 정보와 텍스처 연산자를 이용한 위조 탐지 연구

사우랄 아가왈^{1,2}, 정기현^{2,*}

¹아미티 대학교, ²경일대학교

Forgery Detection using Enhanced Spatial Statistics and Texture Operator in Low Quality Images

Saurabh Agarwal^{1,2}, Ki-Hyun Jung^{2,*}

¹Amity University, ²Kyungil University

*Corresponding author: khanny.jung@gmail.com, +82-53-600-5626

Abstract

In this paper, a robust scheme is discussed to detect median filtering in low quality images. Detection of median filtering assists in overall image forensic. Improved spatial statistical features are extracted from the image to classify pristine and median filtered images. Image array data is rescaled to enhance the spatial statistical information. Features are extracted using Markov model on enhanced spatial statistics. Multiple difference arrays are considered in different directions for robust feature set. Further, texture operator features are combined to increase detection accuracy and SVM binary classifier is applied to train the classification model. Experimental results are promising for low quality factor JPEG compression.

I. INTRODUCTION

Digital images are very popular on social media platforms like WhatsApp, Facebook, Instagram and Twitter. Image is more informative and requires less memory in comparison to other information representation alternatives. The wide availability of digital images makes it vulnerable. Images can be manipulated and shared by anyone using mobile applications and software. The main motive of fake images is to spread some wrong information to gain political, social, business advantages and so on.

Several image processing operations are applied in the process of fake image formation. Some common operations are resampling, rotation, contrast enhancement, blurring, denoising, and filtering, etc. The detection of these operations helped in fake image detection.

In this paper, a new detection of median filtering operation is proposed. Median filter operation is a nonlinear operator and it serves two purposes. First, it is used for denoising and second, it diminishes the artifacts of other operations like resampling, contrast enhancement, etc. Median filter detection is tough in comparison to Gaussian blurring, and mean filtering. Median filter changed the image statistics in nonlinear way. Several attempts have been made to detect median filtering. As JPEG compression is quite popular that make the problem more challenging. JPEG compression suppresses the median filter artifacts. Existing techniques can be improved further to give better accuracy in compressed images. In this paper, one attempt is made to detect median filtering in low quality factor JPEG compression

and small size image blocks. The proposed technique is based on manual feature extraction process. Manual methods computational requirement is very less in contrast to deep learning methods without sacrificing the effectiveness.

Several techniques based on manual feature extraction process are discussed in literature [1]-[9]. In this paper, modified difference arrays are considered for extracting improved Markov features and robust texture operators are utilized for performance boost.

II. THE PROPOSED SCHEME

In this paper, a robust forensic technique for detection of median filtering is proposed to preserve the integrity of images. Markov features are extracted in enhanced rescaled domain and combined with robust texture operator features. Markov features are popular due to their low computational cost and robustness in many classification applications. Second order Markov features proved its worth in compare to first, third and higher order Markov features. Due to this, second order Markov model is considered. Conventionally, second order Markov features can be extracted using following equation.

$$K_{a,b,c} = \Pr(E_{m,n+2} = a | E_{m,n+1} = b, E_{m,n} = c) \quad (1)$$

E is the thresholded difference array in a particular direction and $a, b, c \in \{-H, \dots, H\}$ and H is the threshold value.

The difference array can be defined of a L gray level image $G_{m,n}$ as follows, where $G_{m,n} \in \{0, 1, \dots, L-2, L-1\}$ and $D_{m,n} \in \{-L+1, -L+2, \dots, 0, \dots, L-2, L-1\}$.

$$D_{m,n} = G_{m,n+1} - G_{m,n} \quad (2)$$

Then the thresholded difference array $E_{m,n}$ is

$$E_{m,n} = \begin{cases} -H & \text{if } D_{m,n} < -H \\ D_{m,n} & \text{if } H \geq D_{m,n} \geq -H \\ H & \text{if } D_{m,n} > H \end{cases} \quad (3)$$

However, the thresholding process replaces the $D_{m,n}$ elements with threshold value H without considering its weight. To overcome this gap, first $G_{m,n}$ elements are rescaled in smaller domain and new array named as $GS_{m,n}$, where $GS_{m,n} \in \{-S, -S+1, \dots, 0, \dots, S-1, S\}$. The value of S is decided by experimental analysis. The optimal value is obtained 10 for S . Further, then difference array is calculated as follows, where $GS_{m,n} \in \{-10, \dots, 10\}$ and $DS_{m,n} \in \{-20, \dots, 20\}$.

$$DS_{m,n} = GS_{m,n+1} - GS_{m,n} \quad (4)$$

The enhanced thresholded difference array $ES_{m,n}$ is

$$ES_{m,n} = \begin{cases} -H & \text{if } DS_{m,n} < -H \\ DS_{m,n} & \text{if } H \geq DS_{m,n} \geq -H \\ H & \text{if } DS_{m,n} > H \end{cases} \quad (5)$$

Further, second order Markov model is applied to extract features as follows.

$$KS_{a,b,c} = \Pr(ES_{m,n+2} = a | ES_{m,n+1} = b, ES_{m,n} = c) \quad (6)$$

This process significantly improves the result without increasing feature vector size. The features are extracted for horizontal, vertical, diagonal and minor diagonal difference arrays in forward and backward directions. Further, texture operator based features are extracted.

Texture operators are found effective in many applications like face recognition, object classification, medical imaging, etc. Some of robust texture operators like LBP [10], LBPRIU [11], COALBP [12], RICLBP [13] are also analyzed in median filtering detection. Simple LBP operator [10] provides local statistical information. LBPRIU [11] considers rotation invariant uniform LBP and gives local statistical information. Co-occurrence between LBP's are calculated in COALBP [12] to get global information. RICLBP [13] gives global structural statistical and local statistical details among similar type LBP's. The LBP is treated similar with other LBP if produce similar pattern from one of the rotation angel after rotation, i.e. $0^\circ, 45^\circ, 90^\circ, 135^\circ$ and 180° . Further, co-occurrence of these rotation invariant LBP's give global statistical details. LBP is derived using Fig. 1.

$$\begin{array}{|c|c|c|} \hline 45 & 12 & 12 \\ \hline & 15 & \\ \hline 14 & 19 & \\ \hline \end{array} \rightarrow \begin{array}{|c|c|c|} \hline 1 & & 0 \\ \hline & & \\ \hline 0 & & 1 \\ \hline \end{array} \rightarrow 1011$$

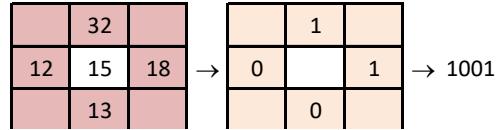


Fig. 1. LBP formation

The center pixel is compared either by horizontal and vertical neighbors or diagonal neighbors. Center pixel is compared with neighbors and if neighbor pixel value less than center pixel value then replaced with 0 otherwise by 1. Four neighbors and anti-clock wise direction is considered for calculating the LBP.

There is noticeable improvement after combining enhanced Markov features and RICLBP texture operator. Experimental results are discussed in next section. Median filter of size 3x3 and 5x5 is considered for experimental analysis. The proposed technique is compared with different texture operators and some median filtering detection techniques.

III. EXPERIMENTAL RESULTS

Median filtering is a nonlinear operation and its detection is challenging in compare to averaging and Gaussian filtering. The proposed scheme is applied on UCID database [14] as many existing techniques are applied on it. UCID database has miscellaneous categories 1,338 images of natural scenes, objects, peoples, etc. Some images of UCID database are shown in Fig. 2.

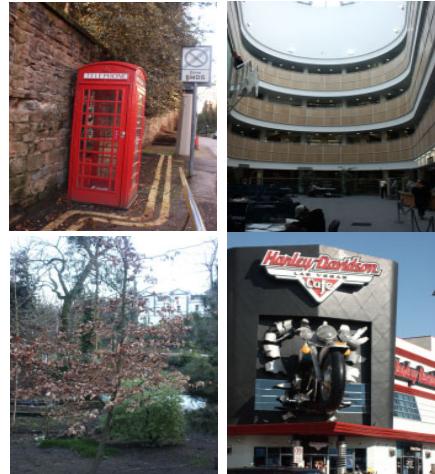


Fig. 2. Example of UCID database images

It is concluded from previous literature and experimental analysis that detection of median filtering is challenging on low quality and small size images. Therefore, experimental results are displayed only for low JPEG quality factors 30, 50 and 70. The small size image blocks of sizes 32x32 and 64x64 pixels are considered. Non filtered images are created by cropping center block of required size after applying JPEG compression. Median filtered images are created by using following steps: apply median filter on the UCID database images in first step, compress images using required quality factor in second step and crop the center block of the compressed median filtered image in the last step.

Padding artifacts of median filter will not arise by following these steps. In experiments, 3x3 and 5x5 window size is considered for median filtering. Training set contains sixty percent images of both the classes and testing set contains remaining forty percent images. Around fifty training and testing pairs are formed for unbiased performance evaluation. The results are shown in terms of detection accuracy, sensitivity, and specificity in percentage, where true positive, true negative, false positive and false negative are abbreviated as TP, TN, FP and FN, respectively.

$$\text{Accuracy} = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) * 100$$

$$\text{Sensitivity} = \left(\frac{TP}{TP + FN} \right) * 100$$

$$\text{Specificity} = \left(\frac{TN}{TN + FP} \right) * 100$$

The proposed technique is compared with COALBP, RICLBP, GDCTF, LBP, LBPRIU, and SPAM. In Fig. 3, average detection accuracies are shown of the proposed and other compared techniques to detect median filter of size 3x3 on block size 64x64. The proposed technique achieves 86.36%, 91.04%, and 93.69% percent detection accuracy for JPEG compression quality factors 30, 50 and 70, respectively. Even for low quality factor Q=30 accuracy is 86.36% of the proposed method. Sensitivity (SE) and specificity (SP) can be seen in Fig. 4. The sensitivity of each method is higher than its corresponding specificity. It means the more number of median filtered images are incorrectly classified as non-filtered image. The difference between sensitivity and specificity values are highest in GDCTF technique.

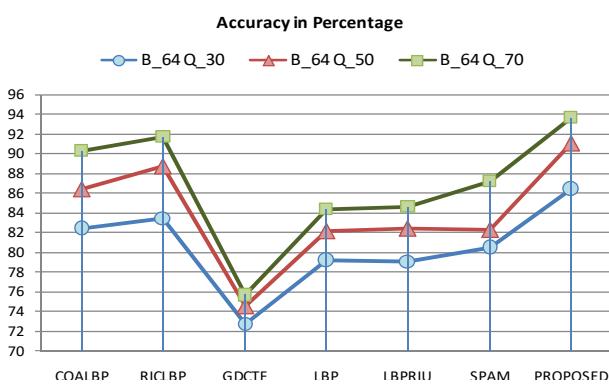


Fig. 3. Detection accuracy for median filter 3x3 and image size 64x64

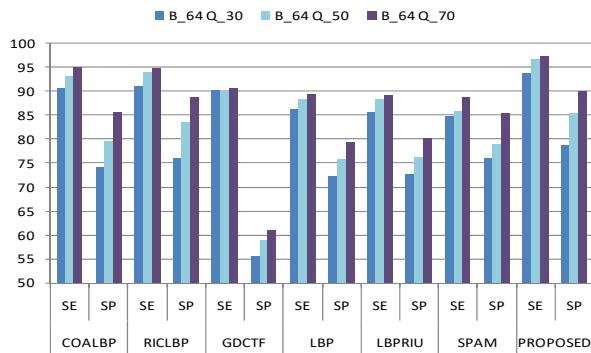


Fig. 4. Sensitivity/specificity for median filter 3x3 and image size 64x64

Now, results are shown for median filtering detection of filter size 3x3 on 32x32 pixels size blocks in Fig. 5. The proposed technique gives 79.38%, 83.34%, and 88.34% percent detection accuracy for quality factors 30, 50 and 70, respectively. The performance gain in compare to other methods of the proposed method for block size 32x32 is comparatively better than block size 64x64. It can be concluded the proposed scheme gives better results in tough scenario. The sensitivity and specificity results are presented in Fig. 6. The similar behavior is followed by sensitivity and specificity as for block size 64x64 in Fig. 4. However, the proposed scheme sensitivity and specificity is the highest in compare to other techniques.

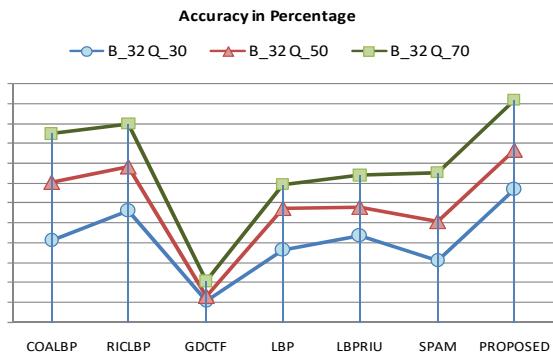


Fig. 5. Detection accuracy for median filter 3x3 and image size 32x32

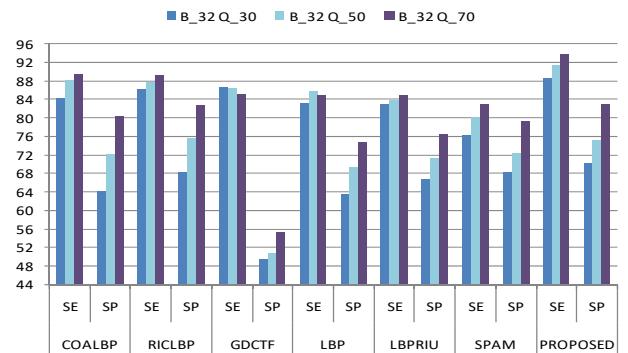


Fig. 6. Sensitivity/specification for median filter 3x3 and image size 32x32

IV. CONCLUSION

Image forensics assured the integrity of the images. Multiple operations have been identified in image forensics. In this paper, a scheme has been proposed for detection of median filtering. Detection of median filtering has been significant due to its nonlinear behavior. Image array is rescaled in optimum domain. The enhanced statistical information has been derived using Markov model. Further, the rotational invariant LBP operator has been applied for additional features. The proposed scheme has given the satisfactory results in experimental analysis. The experiments are performed on small blocks with low quality factor compression.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1A09081842, 2021R1I1A3049788) and Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea (2019H1D3A1A01101687).

REFERENCES

- [1] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," in *Media Forensics and Security II*, 2010, vol. 7541, p. 754110, doi: 10.1117/12.839100.
- [2] Chenglong Chen, Jiangqun Ni, and Jiwei Huang, "Blind Detection of Median Filtering in Digital Images: A Difference Domain Based Approach," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4699–4710, Dec. 2013, doi: 10.1109/TIP.2013.2277814.
- [3] S. Agarwal, S. Chand, and N. Skarbnik, "SPAM revisited for median filtering detection using higher-order difference," *Secur. Commun. Networks*, vol. 9, no. 17, pp. 4089–4102, Nov. 2016, doi: 10.1002/sec.1590.
- [4] A. Peng, S. Luo, H. Zeng, and Y. Wu, "Median filtering forensics using multiple models in residual domain," *IEEE Access*, vol. 7, pp. 28525–28538, 2019, doi: 10.1109/ACCESS.2019.2897761.
- [5] H. Gao and T. Gao, "Detection of median filtering based on ARMA model and pixel-pair histogram feature of difference image," *Multimed. Tools Appl.*, 2020, doi: 10.1007/s11042-019-08340-3.
- [6] H. Gao, T. Gao, and R. Cheng, "Robust detection of median filtering based on data-pair histogram feature and local configuration pattern," *J. Inf. Secur. Appl.*, 2020, doi: 10.1016/j.jisa.2020.102506.
- [7] D. Wang and T. Gao, "Filtered image forensics based on frequency domain features," in *International Conference on Communication Technology Proceedings, ICCT*, 2019, vol. 2019-Octob, pp. 1208–1212, doi: 10.1109/ICCT.2018.8599993.
- [8] A. Peng, G. Yu, Y. Wu, Q. Zhang, and X. Kang, "A universal image forensics of smoothing filtering," *Int. J. Digit. Crime Forensics*, 2019, doi: 10.4018/IJDCF.2019010102.
- [9] A. Gupta and D. Singhal, "A simplistic global median filtering forensics based on frequency domain analysis of image residuals," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 15, no. 3, 2019, doi: 10.1145/3321508.
- [10] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face Recognition with Local Binary Patterns," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3021, 2004, pp. 469–481.
- [11] G. Zhao, T. Ahonen, J. Matas, and M. Pietikäinen, "Rotation-invariant image and video description with local binary pattern features," *IEEE Trans. Image Process.*, 2012, doi: 10.1109/TIP.2011.2175739.
- [12] R. Nosaka, C. H. Suryanto, and K. Fukui, "Rotation invariant co-occurrence among adjacent LBPs," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7728 LNCS, no. PART 1, pp. 15–25, 2013, doi: 10.1007/978-3-642-37410-4_2.
- [13] R. Nosaka and K. Fukui, "HEp-2 cell classification using rotation invariant co-occurrence among local binary patterns," *Pattern Recognit.*, vol. 47, no. 7, pp. 2428–2436, Jul. 2014, doi: 10.1016/j.patcog.2013.09.018.
- [14] G. Schaefer and M. Stich, "UCID - An Uncompressed Colour Image Database," *SPIE, Storage Retr. Methods Appl. Multimed.*, vol. 5307, pp. 472–480, 2003, doi: 10.1117/12.52537

정보보안 관리실태 평가를 위한 블록체인 기반 통합보안로그 관리 시스템

김도훈*, 김호원**

*부산대학교 (대학원생), **부산대학교 (교수)

Blockchain-based Integrated Security Log Management System for Information Security Management Evaluation

Do-Hun Kim*, Ho-Won Kim**

*Pusan National University(Graduate student), **Pusan National University(Professor)

요약

국가 사이버안보를 담당하고 있는 국가정보원은 주기적으로 국가 주요 정보통신망을 점검하고 사이버 모의훈련을 실시하는 등 예방 활동을 수행하고 있으며 사이버 위협 정후를 탐지하여 위험 등급별 경보를 발령하고 있다. 중앙행정기관뿐만 아니라 지방자치단체, 공공기관 등 모든 국가기관에 대해 국가 정보보안 정책 이행실태를 점검하는 정보보안 관리실태 평가를 매년 진행하고 있다. 하지만 기관별 수집되는 정보가 상이하고 관리자에 의해 의도적으로 수정될 가능성을 내포하고 있어 정보시스템 로그에 대한 안전하고 신뢰성 높은 수집·관리 방법이 요구되고 있다. 본 논문은 정보보안 관리실태 평가를 위한 감사증적으로 사용할 수 있는 통합 보안로그를 관리할 수 있는 블록체인 기반 관리 시스템을 제안한다.

I. 서론

최근 한국원자력연구원에 이어 한국항공우주산업 해킹 사태로 인해 내·외부적인 사이버보안의 한계를 지적받으며 긴장감이 고조되고 있다. 해킹사고 발생으로 인해 국민의 국가기관에 대한 신뢰성이 떨어지며 주요 정보 유출을 방지하기 위한 대책 마련이 중요해지고 있다.

국가정보원은 「국가정보원법」 제4조 1항 4호에 따라 국가기관을 대상으로 하는 사이버공격 및 위협에 대한 예방 및 대응 업무를 수행하고 있다. 사이버안보에 관한 국가 차원의 대응을 위해 사이버 관련 정책을 기획하고, 관련 제도지침을 마련하는 등 관리적 보호 방안의 기틀을 마련하며 각 주요 정보통신망을 상시 점검하고, 사이버 모의훈련을 실시하는 등 다양한 정보보호 활동을 수행하고 있다.

국가정보원은 2004년 정보보안 관리수준 평가제도 도입을 통해 각 기관이 국가의 정보보

안 정책 이행수준을 확인하고 체계적인 정보보안 업무 수행을 지원하고 있으며 현재 정보보안 관리실태 평가라는 이름으로 매년 수행하고 있다. 평가 중 결과의 객관성과 공정성을 확보하기 위해 현장실사를 실시하며 민·관 전문가와 함께 평가결과에 대한 적정성을 평가하고 있다. 하지만 현장실사 수행 간 수집할 수 있는 정보시스템 로그에 대한 정보는 해당 제출기관 관리자에 의존할 수밖에 없는 상황으로 고의적인 수정이나 삭제에 대한 대처방안이 부족하다.

본 연구는 정보보안 관리실태 평가를 위한 블록체인 기반 통합보안로그 관리 시스템을 제안하여 통합로그 생성과정에서 자동으로 해시값을 산출하여 블록체인의 온체인에 기록하고, 실제 로그 데이터는 오프체인 저장을 통해 제출되는 데이터의 신뢰성과 무결성을 확보할 방안을 제시한다.

II. 배경지식

2.1 정보보안 관리실태 평가

국가정보원은 국가 정보보안 정책 이행실태 확인을 통해 국가기관이 체계적으로 정보보안 업무를 수행하도록 지원하고 보안의식을 함양하여 사이버안전을 확보한다. 평가는 기관 자체 평가, 현장실사, 결과분석, 결과심의, 결과통보 순으로 진행되며 결과에 대한 공신력 제고를 위해 민·관 전문가로 구성된 평가위원회를 운영한다[1].



그림 1 정보보안 관리실태 평가 수행절차

2.2 보안정보 및 이벤트 관리 시스템(SIEM)

SIEM 소프트웨어는 정보시스템이 운영되는 전 범위에서 로그를 수집, 저장, 분석하며 종합적인 보안 경보와 공격탐지, 차단을 위한 실시간 모니터링 도구이다. SIEM을 사용하여 기관 내 발생하는 모든 정보시스템 구성요소에 대한 정보를 제공받을 수 있다. 장비별 발생하는 보안 이벤트에 대한 통합보안로그를 제공하여 포괄적 방어 체계를 제공하며 다양한 이벤트에 대한 로깅기능을 제공한다.

2.3 블록체인

블록체인은 블록이라고 하는 소규모 데이터 집합을 체인 형태로 연결한 분산 데이터 저장 환경으로써 내부의 데이터를 임의로 수정할 수

없다. 이와 같은 불가변성의 특성은 사용자가 악의적으로 블록체인의 내부 데이터를 변조하는 것을 불가능하게 만들어 데이터 보존에 대한 신뢰도를 향상시킬 수 있다. 또 다른 블록체인의 특성은 블록체인 네트워크에 참여하는 모든 사람은 동일한 원장(저장소)을 가지고 있다는 것으로 사전에 구성원들 간 합의가 끝난 데이터가 저장되어 있어 데이터의 신뢰성을 보장한다[2].

III. 본론

본 논문에서 제안하는 블록체인 기반 통합보안로그 관리 시스템은 온/오프체인의 두 개의 체인으로 구성됩니다. 온체인은 입력과 조회가 빈번하게 발생하는 각급기관의 통합보안로그에 대한 정보와 상세내용의 해시값이 기록되고 오프체인에는 변경이 필요 없는 통합보안로그의 실제 데이터가 저장됩니다. 분리된 체인을 통해 크기가 큰 실제 로그 데이터는 필요할 경우만 열람하며 해시값 비교를 통해 로그가 변경되지 않는지 확인할 수 있다[3].

3.1 시스템 동작절차

제안 시스템은 통합보안로그 내역을 제출하고 감사증적을 소명하려는 각급기관과 통합보안로그 내역을 바탕으로 정보보안 관리실태 평가를 수행하려는 평가기관으로 구성되어 있다. 시스템 동작 절차는 1) 각급기관 및 평가기관의 블록체인 네트워크 참여, 2) 각급기관의 통합보안로그 생성 및 제출, 3) 평가기관의 로그검토 및 평가 순으로 진행된다.

1) 각급기관 및 평가기관 참여 : 각 기관은 블록체인 네트워크에 노드로 참여하기 위해 노드용 서버 또는 PC를 설치하여 로그 제출 및 평가를 위한 준비를 한다.

2) 통합보안로그 생성 및 제출 : 각급기관은 기관에서 보유 중인 SIEM 또는 통합보안관제 시스템 등을 통해 보안이벤트에 대한 통합보안로그를 생성한다. 생성되는 로그는 스마트 컨트

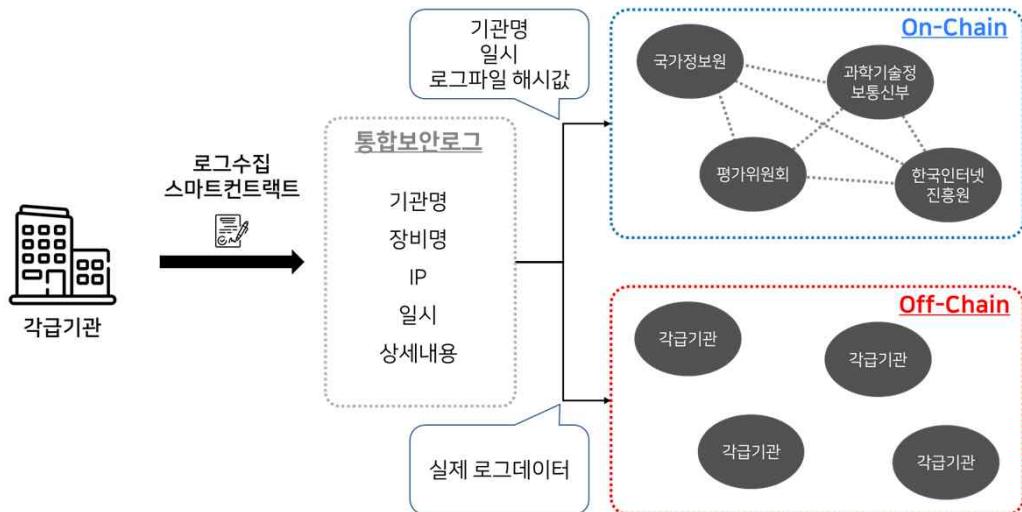


그림 2 블록체인 기반 통합보안로그 관리 시스템 구성도

랙트에 의해 생성 시 자동으로 해시화되어 온 체인에 등록되며, 실제 로그데이터는 각급기관의 오프체인에 저장된다.

3) 로그검토 및 평가 : 국가정보원을 비롯한 평가기관은 보안 수준을 평가함에 있어 감사증적으로 통합보안로그를 사용하며 온체인에 등록된 해시값을 이용해 로그가 변경되지 않았음을 확인한다.

IV. 결론

본 논문에서는 국가기관의 정보보안 수준을 향상하고 주기적으로 점검하기 위한 국가정보원 정보보안 관리실태 평가에 사용되는 감사증적에 대한 신뢰성 확보를 위한 블록체인 기반 통합보안로그 관리 시스템을 제안하였다. 각 기관의 담당자에게 의존적으로 수집해야 했던 로그 정보를 스마트 컨트랙트를 통한 자동수집화하여 기록된 데이터에 대한 무결성과 신뢰성을 확보할 수 있었다. 로그 생성과정에서 데이터를 해시화하여 저장함으로써 임의의 사용자에 의한 수정을 예방할 수 있었다.

향후 실제 시스템을 구현을 통해 데이터 처리 속도 문제와 기존 중앙집중식 플랫폼과의 효율성 비교를 통해 최적화된 시스템 구현을 위한 연구가 필요할 것으로 보인다.

Acknowledgement

이 논문은 국토교통부의 스마트시티 혁신인재 육성사업으로 지원되었습니다.

[참고문헌]

- [1] 국가사이버안전센터, "정보보안 관리실태 평가 소개," 정보보호학회지, 23(5), pp. 9-11, Oct, 2013.
- [2] Dylan Yaga, Peter Mell, Nik Roby and Karen Scarfone, "Blockchain Technology Overview," National Institute of Standards and Technology Internal Report, pp. 68, Oct, 2018.
- [3] Andrew Sutton and Reza Samavi, "Blockchain Enabled Privacy Audit Logs", ISWC 2017: The Semantic Web, pp. 645-660.

적대적 예제를 이용한 디지털 워터마킹

윤영여*, 심준석*, 조현진*, 강효은*, 김호원**

*부산대학교 (대학원생), **부산대학교 (교수)

Digital Watermarking Using Adversarial Examples

Young-Yeo Yun*, Jun-Seok Shim*, Hyun-Jin Cho*, Hyo-Eun Kang*,
Ho-Won Kim**

*Pusan Nataional University([Graduate student](#)), **Pusan National University([Professor](#))

요약

본 논문은 적대적 예제를 이용하여 디지털 워터마크를 생성하는 방법을 제안한다. 적대적 예제는 딥러닝 분류 모델에 이미지를 입력하였을 때 원하는 클래스로 분류하도록 신중하게 조작한 이미지다. 딥러닝 분류 모델에 입력하면 특정 코드로 분류하여 워터마크의 역할을 하도록 적대적 예제를 생성한다. 생성한 워터마크에 대해서 완전성, 견전성 실험 결과를 제시한다. 타겟 모델에 적대적 예제를 입력하여 워터마크를 위한 코드와 동일하게 모델의 출력을 조작하여 완전성을 검증하였다. 또한 워터마크가 없는 이미지를 입력하여 코드 길이 16비트 이상에서 워터마크를 위한 코드와 다르게 출력하는 결과를 통해 견전성을 확인하였다. 이미지의 스케일 변화와 노이즈에 대해 견고성을 실험한 결과, 적대적 예제가 10% 이하 변조되었을 때, 타겟 모델이 워터마크를 위한 코드와 89.1% 이상 동일하게 출력하였다. 워터마크를 위한 적대적 예제 샘플을 보면 워터마크를 위한 코드가 길수록, 견고한 변형의 범위가 넓을수록 적대적 예제와 원본 이미지의 차이를 쉽게 인식할 수 있었고, 코드 길이 16 비트와 변형 범위 10%에 대한 적대적 예제가 효용성 측면과 위화감이 없는 점을 고려하면 워터마크로 사용되기에 적합하다.

I. 서론

컴퓨터 소프트웨어의 발전에 따라 이미지를 처리하는 기술들이 발전하고 있다. 그에 따라 이미지와 같은 디지털 컨텐츠의 저작권 보호 기술에 대한 필요성이 높아지고 있다. 디지털 워터마크는 저작권 보호에 적합한 기술이다. 디지털 워터마크란 사진, 동영상과 같은 원본 미디어에 저작권 정보와 같은 비밀 정보를 삽입하여 관리하는 기술을 말한다.

디지털 워터마크는 준수해야 할 2가지 사항이 존재한다. 첫 번째, 디지털 워터마크는 이미지를 보는 사람의 눈에 띠지 않아야 하며, 디지털 워터마크가 이미지의 품질을 저하시켜서는 안 된다. 두 번째 준수 사항은 이미지 처리에 대한 디지털 워터마크의 견고성(robustness)이다. 이미지 왜곡이 일어나더라도 워터마크가 그 역할

을 해야 하며, 만약 워터마크를 제거할 시 이미지 품질을 저하시키는 수준까지 이미지 왜곡을 진행해야 한다.

본 논문에서는 적대적 예제 기법[4]을 통해 이미지의 픽셀을 수정하여 워터마크를 삽입하는 워터마킹 기법을 제안한다. 이미지의 크기 변화와 노이즈에 견고하고 눈에 띠지 않는 워터마크를 실험을 통해 검증하였다.

II. 관련연구

워터마크에 대한 예시로 least significant bits(LSB)와 같이 이미지에서 중요하지 않은 비트를 워터마크의 비트로 대체하였다[1]. 이미지 픽셀을 변화시키는 방법으로 워터마크를 삽입하는 경우 필터링이나 신호의 변형을 통해 워터마크 정보가 쉽게 제거될 수 있다. [2]는

singular value decomposition (SVD), discrete wavelet transform(DWT)를 통해 주파수 스펙트럼 안에 워터마크를 숨기는 방법을 제안하였으며 높은 보안성과 신뢰성을 달성했다. [3]은 SVD, DWT에 더하여 discrete cosine transform(DCT)까지 적용해 인지할 수 없는 주파수 도메인 기반의 워터마킹 알고리즘을 제안하였다.

III. 워터마크 생성 방법

적대적 예제의 두 가지 특징을 이미지 워터마크를 생성하기 위해 사용한다. 첫 번째 특징은 워터마크와의 공통점으로 사람의 눈에 띠지 않아야 한다는 점이다. 적대적 예제도 워터마크와 마찬가지로 미세한 차이를 통해 원하는 바를 이룬다. 두 번째는 딥러닝 모델이 사용자가 원하는 클래스로 출력하도록 만들 수 있다는 점이다.

Algorithm 1은 적대적 예제를 이용하여 워터마크 생성하는 방법을 나타낸다. n 비트 길이의 2진 코드 \mathbf{b} , 딥러닝 분류 모델 C , 이미지 x 를 입력받는다. 이미지 x 를 n 개로 자르고 조각 이미지 p_i 에 노이즈 r_n 을 추가하고 r_c 만큼 스케일을 변경한다. C 에 변형된 조각 이미지 p_i^t 를 입력한 출력과 b_i 의 차이를 최소화하도록 p_i 를 최적화한다.

Algorithm 1: watermark embedding into image - Random noise r_n , Random scale r_c								
Input: Image x , Model C , Code \mathbf{b}								
Output: x' (image embedded watermark)								
1 $\mathbf{p} = \text{slice}(x, n)$								
2 $\text{for } p_i \text{ in } \mathbf{p} \text{ do}$								
3 for $\text{iter} \leftarrow 1$ to max_iteration do								
4 $p_i^t = \text{clip}(p_i + r_n, 0, 1)$								
5 $p_i^t = \text{resize}(p_i^t, r_c)$								
6 $\text{loss} = L(C(p_i^t), b_i)$								
7 $\text{optimize } \min_{p_i} \text{loss}$								
8 end								
9 end								
10 $x' = \text{assemble}(\mathbf{p}, n)$								

IV. 실험 결과

워터마크를 실험하기 위해 COCO 데이터셋을 모델의 입력으로 활용하였고, ImageNet 데이터를 사전학습한 VGG19[5] 모델의 컨볼루션 파트를 백본으로 크기가 1인 완전연결층을 결합한 타겟 모델을 통해 워터마크를 생성하였다. 또한 최적화를 위해 Adam 알고리즘을 사용하

고 학습률은 10^{-4} , max_iteration 은 10^4 으로 실험하였다.

완전성은 워터마크를 인식할 수 있어야 한다는 것이고 건전성은 워터마크가 없는 이미지에 워터마크가 있다고 하면 안 된다고 하는 것이다. 즉 워터마크가 있는 이미지의 경우에만 입력하였을 때 타겟 모델이 미리 정의한 코드를 출력해야 한다. 샘플링한 이미지 100장에 대해서 **Algorithm 1**의 과정을 거쳐 워터마크를 삽입하고 타겟 모델에 입력하여 같은 코드가 출력되는지 확인하였다(랜덤 스케일 r_c , 랜덤 노이즈 r_n 의 범위는 $\pm 10\%$ 로 수행). 4/16/64 비트 길이의 코드에 대해서 워터마크가 있는 이미지의 경우 워터마크 코드와 동일한 모델 출력으로 완전성을 보장함을 확인하였다. 표 1은 건전성 실험을 위해 워터마크가 없는 100,000장의 이미지를 샘플링해서 모델에 입력한 결과이다. 16 비트 이상의 코드를 사용해야 동일한 코드를 출력하는 이미지가 없음을 확인하였다.

표 1 워터마크 코드와 동일한 모델 출력 비율

코드 길이								
4			16			64		
최소	최대	평균	최소	최대	평균	최소	최대	평균
0.0	1.0	0.5	0.19	0.81	0.62	0.38	0.75	0.62

디지털 이미지는 쉽게 변환될 수 있다. 이는 워터마크도 쉽게 변환될 수 있다는 말이 된다. 그렇기 때문에 워터마크는 이미지가 변환되더라도 인식되어야 한다. 모델이 변환된 워터마크를 인식할 수 있는지 살펴보기 위해 **Algorithm 1**의 랜덤 스케일 r_c , 랜덤 노이즈 r_n 의 범위를 $\pm 5/10/20/40\%$ 로 설정하여 실험하였다. 범위를 $\pm 20\%$ 라고 하였을 때 조각 이미지 \mathbf{p} 의 크기와 동일한 랜덤값을 $[-0.2, 0.2]$ 범위에서 결정하여 \mathbf{p} 에 더하였고, 원래 \mathbf{p} 의 높이/너비 수치를 100%라고 할 때 80~120% 범위로 변경하여 모델에 입력하였다. 표 2는 4/16/64 비트 길이의 코드로 생성한 워터마크가 포함된 이미지 10장을 생성하고 랜덤하게 각 범위 내에서 100번 변형하여 모델에 입력하였을 때, 타겟 모델의 출력과 워터마크 코드가 동일한 비

표 2 변형 범위에 따라 타겟 모델이 워터마크 코드와 동일하게 출력하는 비율

코드 길이	변형 범위											
	$\pm 5\%$			$\pm 10\%$			$\pm 20\%$			$\pm 40\%$		
	최소	최대	평균	최소	최대	평균	최소	최대	평균	최소	최대	평균
4	1.0	1.0	1.0	1.0	1.0	1.0	0.75	1.0	0.995	0.25	1.0	0.945
16	0.938	1.0	0.999	0.938	1.0	0.997	0.813	1.0	0.982	0.625	1.0	0.978
64	0.891	1.0	0.989	0.922	1.0	0.986	0.75	1.0	0.979	0.469	1.0	0.849

율을 나타낸다. $\pm 10\%$ 이하의 변형 범위에서 워터마크 코드와 타겟 모델의 출력이 89.1% 이상 일치하였다.

그림 1은 적대적 예제 생성 기법을 이용하여 만든 워터마크 이미지 샘플이다. 변형 범위 $\pm 5/10/20\%$, 코드 길이 4/16/64 비트에 대한 샘플을 제시하였다. 코드 길이가 길수록, 변형 범위가 클수록 원본과 비교하여 차이가 선명하다.

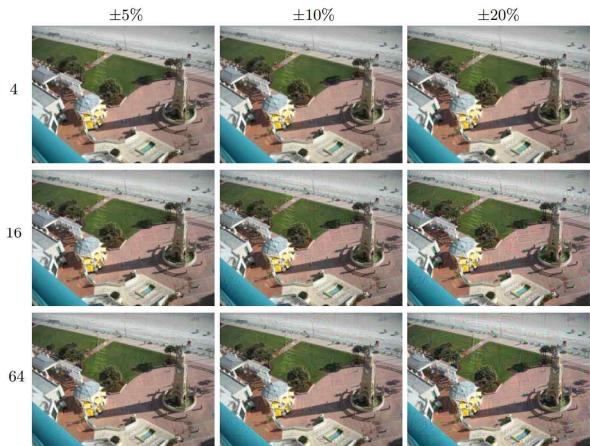


그림 1 워터마크 이미지 샘플
(코드 길이: 행, 변형 범위: 열)

V. 결론

본 논문에서는 적대적 예제 생성 기법을 이용하여 워터마크를 이미지에 삽입하는 방법을 제안하였다. 제안하는 방법으로 생성한 워터마크에 대해서 완전성, 견진성 실험 결과를 제시하였다. 또한 견고성 실험에 대해서 10% 이하 변형 범위에도 타겟 모델이 16 비트 이하 길이의 워터마크 코드와 89.1% 이상 동일하게 출력하였다. 생성된 워터마크 이미지를 보면 코드 길이 16 비트와 변형 범위 10%에서 효용성 측면과 원본과의 차이가 작은 점을 고려했을 때 워터마크로 사용하기 적합하다.

ACKNOWLEDGMENT

이 논문은 국토교통부의 스마트시티 혁신인재육성사업으로 지원되었습니다.

[참고문헌]

- [1] A. Z. Tirkel, G.A. Rankin, R. G. van Schyndel, W. J. Ho, N. Mee, C. F. Osborne.: Electronic Watermark. In: Digital Image Computing, Technology and Applications (1994).
- [2] C. Lai, C. Tsai.: Digital Image Watermarking Using Discrete Wavelet Transform and Singular Value Decomposition. IEEE Transactions on Instrumentation and Measurement 59(11), 3060–3063 (2010).
- [3] D. Singh, S. K. Singh.: DWT-SVD and DCT Based Robust and Blind Watermarking Scheme for Copyright Protection. Multimedia Tools and Applications 76, 13001–13024 (2017).
- [4] I. J. Goodfellow, J. Shlens, C. Szegedy.: Explaining and Harnessing Adversarial Examples. In: International Conference on Learning Representations (2015).
- [5] K. Simonyan, A. Zisserman.: Very Deep Convolutional Networks for Large-scale Image Recognition. In: International Conference on Learning Representations (2015).

iOS 부트로더 취약점을 통한 펌웨어 분석 방법 연구*

김성민* 류재철*

*충남대학교 컴퓨터공학과 (대학원생, 교수)

Study on Firmware Analysis with iOS Bootloader Vulnerabilities

Seong-Min Kim* Jae-Cheol Ryou*

*Department of computer and engineering Chungnam National University
(Graduate student, Professor)

요약

iOS의 부트로더는 여러 단계로 나누어져 있으며, 각 단계가 다음 단계의 무결성을 보장하는 방식으로 설계되어 있다. 만약 일부 부트로더 단계에서 취약점이 발생한다면, 그 이후의 부트 과정이 손상되어 악성 펌웨어가 로드될 수 있다. 따라서 부트로더 구성요소를 분석하여 취약점을 미리 발견하여 제조사가 적절하게 패치할 수 있도록 알리는 것은 매우 중요하다. 본 논문에서는 일반 상용 iOS 장치를 이용하여, 부트로더 취약점과 특수 장비를 결합한 부트로더 펌웨어 분석 방법을 설명하고, 그 과정에서 발생 가능한 문제점과 이를 해결하기 위한 방안을 제시한다. 이는 차후 부트로더 펌웨어 취약점 발견 연구의 기반 기술로 활용할 수 있다.

I. 서론

Apple이 개발한 스마트폰 운영체제인 iOS 시스템은 부트 과정에서 손상된 펌웨어를 로드하지 않도록 설계되어 있다. 이러한 설계는 iOS 보안 부트 체인(Secure Boot Chain)[1]으로 불리며, 그 중 AP(Application Processor)의 가장 첫 번째 부트로더인 SecureROM은 수정이 불가능한 읽기 전용 펌웨어를 포함하고 있다.

SecureROM에서 발생한 취약점을 통해 공격자가 임의 코드를 실행하거나 장치를 제어할 수 있는 권한을 얻게 된다면 이후 부트 과정의 신뢰성이 무너지고, 악성 펌웨어를 로드할 수 있게 된다. 이러한 공격을 방지하기 위해서 펌

웨어를 분석하고, 악용 가능한 취약점을 파악하여 패치할 수 있도록 제조사에 전달할 필요가 있다.

그러나 Apple은 자사의 코드를 공개하지 않으며, Apple Security Research Device Program[2]과 같이 매우 제한적으로 보안 연구를 지원하기 때문에 특수한 장비나 환경이 갖추어지지 않으면 부트로더 분석을 진행하기 어렵다.

본 논문에서는 공개된 SecureROM 취약점과 일반 상용 iOS 장치 및 특수 장비를 활용하여, iOS 부트로더 구성요소들을 분석하는 방법을 보인다. 2장에서는 iOS 부트로더의 전체적인 구성을 대해 간략하게 소개한다. 3장에서는 iOS 부트로더 펌웨어를 정적 및 동적 분석하는 방법과 동적 분석 시 마주할 수 있는 문제점을 소개하고, 해결 방안을 제시한다. 마지막 4장에서는 본 논문에 대한 결론을 내린다.

* 이 논문은 2021년 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2020-0-01840, 스마트폰의 내부데이터 접근 및 보호 기술 분석)

II. 배경 지식

iOS의 AP 부트 단계는 크게 SecureROM, iBoot, Kernel의 세 단계로 분류할 수 있다. SecureROM은 가장 맨 처음에 실행되는 코드로, 읽기 전용이며 하드웨어에 포함되어 있어서 수정할 수 없다. iBoot는 SecureROM 이후에 로드되는 펌웨어로, 코드가 더 길고 복잡하며 큰 메모리 공간에 대해 접근할 수 있으며, Apple의 소프트웨어 업데이트에 의해 패치될 수 있다. 일부 하위 버전의 iOS 장치들은 LLB (Low-Level Bootloader)라는, iBoot과 동일하지만 좀 더 작은 형태의 부트로더가 iBoot 이전에 존재하는 경우도 있다[3]. Kernel은 iBoot에 의해 로드되는 펌웨어로 iOS 시스템의 전반적인 핵심 부분을 구성하며 Kernel이 로드되었다는 것은 곧 부트가 완료되었다는 것을 뜻한다.

각 부트 단계는 다음 부트 단계에 대한 무결성을 검증하기 때문에, 사용자가 악의적으로 편집한 iBoot를 로드하려고 시도하면 이전 단계인 SecureROM이 이를 거부한다. SecureROM의 코드는 읽기 전용이므로 수정할 수 없기 때문에, SecureROM은 전체 부트 과정에서 신뢰의 뿌리(Root of Trust)를 담당하고 있다. 바꾸어 말하면, SecureROM에서 발생한 취약점으로 인해 전체 부트 과정이 크게 손상될 수 있으며 패치를 통해 수정하지도 못하는 매우 위협적인 취약점이라고 할 수 있다.

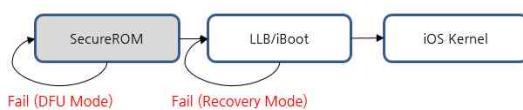


그림 1 iOS 보안 부트 체인 요약

만약 각 부트 단계에서 무결성 확인 등으로 다음 펌웨어를 로드하지 못하는 상태가 되면, 그림 1과 같이 각 단계에서는 새로운 이미지를 다운로드 받기 위한 복원 전용 모드에 진입한다. SecureROM은 이러한 상태를 DFU(Device Firmware Upgrade) Mode, iBoot는 Recovery Mode라고 칭한다. 이를 위한 iBEC, iBSS라고 불리는 펌웨어가 있는데, 이는 iBoot와 유사하

지만 오직 복원만을 위한 펌웨어라는 점이 다르다.

III. 펌웨어 분석 연구

3.1. 정적 분석 방법

iOS 부트로더는 사전에 정의된 메모리 주소에서 실행된다. 따라서 부트로더에는 항상 코드가 위치한 주소를 검사하고, 필요하다면 자기 자신을 옮기는 재배치 코드가 존재한다[4]. 해당 코드를 기준으로 일반적인 바이너리 정적 분석과 동일한 기법으로 펌웨어를 분석할 수 있다.

그러나 Apple은 공식적으로 자사의 부트로더에 대한 어떠한 코드나 심볼 정보도 제공하지 않으므로 이미 알려진 정보나 낮은 버전의 펌웨어에 있는 심볼 정보, 다른 버전의 펌웨어 분석 결과 등을 이용하여 정적 분석을 수행할 수 있다.

```

ADR P      X0, #start@PAGE
ADD      X0, X0, #start@PAGEOFF
LDR      X1, =start
BL       sub_10000073CC
CMP      X1, X0
B.EQ    loc_10000003C
MOV      X30, X1
LDR      X2, =0x180001100
LDR      X3, =start
SUB     X2, X2, X3
SUB
; CODE XREF: start+34↑j
reloc_100000028
LDP      X3, X4, [X0],#0x10
STP      X3, X4, [X1],#0x10
SUBS   X2, X2, #0x10
RET
; -----
loc_10000003C
MSR      #6, #0xF
ADR P      X10, #sub_100000800@PAGE
ADD      X10, X10, #sub_100000800@PAGEOFF
; CODE XREF: start+14↑j
  
```

그림 2 SecureROM 재배치 코드 (T8010)

펌웨어의 재배치 코드 끝 부분에는 일반적으로 그림 2와 같이 main code와 boot trampoline의 주소가 위치한다. main code는 장치의 상태를 읽어 DFU Mode의 동작을 수행하거나 다음 부트 단계를 로드하고 부트 과정으로 진입하는 역할을 담당한다. boot trampoline은 다음 부트 단계로 넘어갈 때 실행되는 코드로, 레지스터 상태를 정리하고 이전 부트 단계의 코드를 삭제하거나 엑세스를 금지하는 역할을 수행하며 handoff trampoline이라고도 불린다.

그림 3과 같이 main code의 일부분, 특히 일

반적으로 끝 부분에는 다음 부트 단계로 넘어가기 위한 `prepare_and_jump` 함수가 호출된다. 만약 부트 단계로 넘어가지 못하는 오류 상태에 빠지면 `platform_reset` 함수를 호출한다.

```

loc_1000007B4      ; CODE XREF: SecureROM:000000010
    MOV     W0, #false
    BL     platform_reset_100009DD4
;
;

boot_loaded_image_1000007BC ; CODE XREF: SecureROM:000000010
    BL     sub_10000B3EC ; security stuff
    BL     sub_10000B470 ; security stuff
    MOV     X1, #0x10000000
    MOVK   X1, #0x8000,LSL#16 ; args[1] -> 1800B000
    MOV     W0, #0 ; args[0] -> BOOT_UNKNOWN
    MOV     X2, #0 ; args[2] -> NULL
    BNE    prepare_and_jump_100009E64

```

그림 3 `prepare_and_jump` 및 `platform_reset` 함수 호출 (T8010)

이러한 방법으로 식별한 함수들을 다른 버전의 펌웨어와 비교하면 Apple이 코드를 어떻게 패치했는지 확인할 수 있으며, 이전 버전의 코드에서 새로운 취약점을 발견할 수도 있다. 반면, 현재까지 식별한 함수들을 기반으로 이후 펌웨어 버전을 분석할 때 심볼을 활용하거나 코드 상태를 유추할 수 있다.

3.2. 동적 분석 방법

`checkm8 exploit`은 DFU Mode에서 발생한 취약점을 악용한다. 공격자는 이를 통해 DFU Mode에서 임의의 코드를 실행하거나, 메모리에 로드된 코드를 변조할 수 있다. 이 취약점을 발견한 axi0mX라는 닉네임의 해커는 ipwndfu라고 불리는 도구를 Github에 공개하여 자유롭게 사용할 수 있도록 게시하였다. 이 도구는 현재 까지 공개된 iOS 부트로더 관련 취약점에 대한 폐이로드와 임의의 코드 실행, 임의의 메모리 읽기 및 쓰기 기능이 포함된 python 프로그램으로, 간단한 코드 몇 줄만으로도 메모리를 읽고 쓰거나 코드를 실행하는 등 다양한 동작을 수행할 수 있다. 이는 펌웨어 동적 분석에 매우 유용하게 활용할 수 있다.

JTAG 및 SWD는 펌웨어 디버깅을 하기 위해 사용할 수 있는 표준 인터페이스이다. 일반적인 iOS 장치들 또한 이러한 인터페이스를 통해서 펌웨어를 디버깅할 수 있다[5]. 그러나 일반적으로 상용 iOS 장치들은 이러한 디버깅 가능 여부를 지정하는 래지스터가 비활성화 되어 있다. 반면에, `checkm8 exploit`에 취약한 기기들

은 ipwndfu 도구가 제공하는 demotion 기능을 활용하면 디버깅 가능 여부를 지정하는 래지스터의 값을 조작하여 JTAG 및 SWD 디버깅이 가능하도록 만들 수 있다.

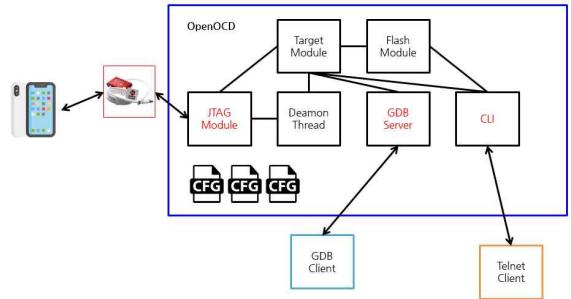


그림 4 Bonobo JTAG/SWD Debug Cable 구성도

LambdaConcept 사에서는 오픈 소스 OpenOCD 드라이버 및 커스텀 펌웨어로 구성된 디버그 전용 케이블인 Bonobo JTAG/SWD Debug Cable을 제공한다. 본 케이블은 lighting cable을 통해 iOS 장치에 연결될 수 있으며, 그림 4에 나타난 것과 같이 OpenOCD와 GDB를 활용하여 펌웨어를 동적 분석할 수 있도록 돕는다.

```

(gdb) x/100i 0x180018000
0x180018000: msr    daifset, #0xf
0x180018004: mov    x29, x0
0x180018008: mov    x18, x1
0x180018010: add    x0, x0, #0x40
0x180018014: cmp    x0, x1
0x180018018: b.cc   0x18001802c // b.lo, b.ul, b.last
0x18001802c: dsb    sy
0x180018040: isb    x3, #0x0
0x180018044: mov    x3, #0x0          // #0

```

그림 5 GDB를 활용한 디버깅 예시

그림 5는 Bonobo JTAG/SWD Debug Cable에 연결된 상태에서 GDB를 통해 CPU 디버깅을 수행하는 모습이다. 이와 같이, CPU 및 래지스터에 엑세스하고, 중단점을 설정하거나 임의의 메모리를 읽고 쓸 수 있다.

부트로더를 동적으로 분석하면 중단점 설정 등으로 오랫동안 CPU가 중단(halt) 상태에 빠질 수 있으며, 이는 일반적으로 올바르지 않은 경우로 간주된다. 따라서 iOS를 포함한 많은 임

베디드 시스템에서는 wdt(watchdog timer)를 도입하여 기기가 오랫동안 멈춰있다고 판단되면, 강제로 재부팅시키는 방법을 사용한다. 오랫동안 동적 분석을 수행하는 도중 기기가 재부팅되는 상황을 막기 위해서, 분석자는 이를 제어할 필요가 있다. iOS 부트로더의 wdt는 memory-mapped register로서 특정 메모리 영역에 맵핑되어 있기 때문에 위치를 찾아내면 이를 비활성화 할 수 있다.

```
platform_reset_100009DD4          ; CODE XREF: SecureROM
                                    ; _panic+DC4p
var_s0    = 0
STP
MOV
BL
dead_loop_100009DE0:           ; CODE XREF: platform_
                                ; dead_loop_100009DE0
B
; End of function platform_reset_100009DD4
```

그림 6 platform_reset 함수 (T8010)

그림 6에 나타난 platform_reset 함수는 wdt_chip_reset 함수를 호출하여 wdt 레지스터가 맵핑된 특정 메모리 영역에 직접 접근하는 동작을 수행한다. 이를 통해 wdt 레지스터의 위치를 특정할 수 있으며, 그림 7과 같이 ipwndfu를 수정하여 wdt를 비활성화 하는 코드를 작성할 수 있다.

```
if 'CPID:8950' in serial_number:
    device.write_memory (0x3F103030+0xC, struct.pack('<I', 0x0))
elif 'CPID:8010' in serial_number:
    device.write_memory (0x2102B0000+0xC, struct.pack('<I', 0x0))
    device.write_memory (0x2102B0010+0xC, struct.pack('<I', 0x0))
elif 'CPID:8015' in serial_number:
    device.write_memory (0x2352B0000+0xC, struct.pack('<I', 0x0))
    device.write_memory (0x2352B0010+0xC, struct.pack('<I', 0x0))
else:
    print ("Not supported")
    sys.exit(-1)
print ("watchdog timer is now disabled")
```

그림 7 wdt를 비활성화하는 수정된 ipwndfu

IV. 결론

본 논문은 iOS 부트로더의 구성을 간략하게 살펴보고, checkm8 exploit과 특수한 JTAG/SWD 장비를 활용한 분석 방법을 설명하였다. 또한 해당 과정에서 찾은 심볼과 코드 정보를 통해 동적 분석에 유용한 wdt 비활성화 방안을 제시하였다.

이러한 분석 기술을 활용하면 다른 버전의 부트로더 분석에 응용할 수 있고, 신규 취약점 발견 등의 보안 연구의 기초 자료로 활용할 수 있을 것으로 기대한다.

[참고문헌]

- [1] Apple Platform Security, May 2021
- [2] Apple Security Research Device Program, <https://developer.apple.com/programs/security-research-device/>
- [3] Introduction to iBoot, Harry Moulton, <https://h3adsh0tzz.com/inside-xnu/iboot/intro>
- [4] iBoot RE (64 bits), <https://github.com/kpwn/iOSRE/blob/master/wiki/iBoot-RE.md>
- [5] KTRW: The journey to build a debuggable iPhone, Project Zero Team, <https://googleprojectzero.blogspot.com/2019/10/ktrw-journey-to-build-debuggable-iphone.html>

소셜 네트워크 서비스 내의 이미지를 통한 개인정보 유출 위험성 연구*

권희원*, 김명주**

*서울여자대학교 정보보호학과(대학원생)

**서울여자대학교 정보보호학과 교수

A Study on the Risk of Personal Information Leakage through Images on Social Network Societies

Hee-Won Kwon*, Myuhng-Joo Kim**

*Department of Information Security, Seoul Women's University
(Graduate student)

*Professor, Department of Information Security, Seoul Women's University

요약

디지털 시대를 맞이하여 소셜 네트워크 서비스(social network service)의 등장은 스마트폰의 보급률 증가와 함께 많은 사람들의 의사소통 수단으로 자리매김하였으며 특히 COVID-19 사태로 인한 언택트 시대를 맞이하여 중요한 소통창구로서 활용되고 있다. SNS에 나타나는 이미지는 자기 지시성, 기록성, 정체성 표현, 간접 경험 4가지 속성을 가질 수 있는데 이 중 기록성과 정체성의 속성은 많은 개인정보 유출을 수반한다. 본 논문에서는 이러한 속성을 강하게 가진 이미지를 통해 어떻게 이름, 생년월일 등 기본적인 개인식별정보는 물론 취향, 관심사, 동선, 행적 등의 민감정보까지 유출하는지를 보임으로써 개인정보 유출 위험성을 제시한다.

I. 서론

우리는 전통적인 아날로그 사회에서 벗어나 디지털 전환(digital transformation)이라는 대변혁의 시대를 맞이하고 있다. 방대하고 다양한 데이터를 핵심 요소로 꿈고 있는 ‘디지털 전환’은 산업 전반의 디지털 생태계 구축은 물론 사람들의 일상생활 및 의사소통에까지 그 영향력이 확산되고 있다[1]. 사람, 사물, 정보가 네트워크를 통해 상호 연결되는 초연결 시대에는 모든 데이터가 디지털화되고 상호 연결된다. 실시간으로 수집되는 개인의 데이터는 개인의 경제적·사회적 성향 등 데이터 총체를 형성하게 되며, 수집·처리·공유되는 데이터 총체는 미래의 활용도가 높은 잠재 자원으로 존재하게 된다.

이처럼 네트워크를 통해 모든 것이 연결되고, 디지털화되는 시대에 개인정보 관련 데이터 침해 위험에 대한 담론은 꾸준히 논의되어져 왔다. 특히 소셜 네트워크 서비스(social network service, 이하 SNS)의 등장은 스마트폰의 보급률 증가와 함께 많은 사람들의 의사소통 수단으로 자리매김하였고 이제는 하나의 문화로서의 역할을 해나가고 있다[2].

대중적이며 대표적인 SNS로는 페이스북(Facebook), 인스타그램(Instagram), 카카오스토리(Kakao Story), 트위터(Twitter), 유튜브(Youtube) 등이 있다. SNS를 통해 커뮤니티를 형성하고 다양한 활동을 하면서 특히 COVID-19 사태로 인한 언택트(untacted) 시대에 중요한 소통창구로서 활용되고 있다.

반면 이로 인하여 개인정보 유출에 관한 문제도 더 두드러지고 있다. 페이스북의 개인정보

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학 지원사업의 연구결과로 수행되었음(2016-0-00022)

유출사건의 경우 피해자 규모가 전 세계 최대 8천 7백만 명, 국내 최대 8만 6천 명에 이른다 [3]. 많은 사람들이 이러한 큰 개인정보 유출 피해사건에는 관심을 두지만 우리가 흔히 SNS에 올리는 이미지들이 얼마나 많은 개인정보를 담고 있는지에 대해서는 특별한 관심과 연구가 이루어져 있지 않다.

따라서 본 논문에서는 SNS 상에서 올라오는 공개정보(Open Source Intelligence, OSINT) [4]중에서 이미지에 내포된 개인정보 유형을 알아보고 관련 개인정보 유출 위험성에 대해 경각심을 가지고자 한다.

II. SNS에 나타난 이미지의 속성

SNS에 나타나는 이미지의 속성은 총 4가지로 정의된다[5]. 이때의 이미지는 사용자가 그리거나 조작한 이미지 보다는 실제 촬영을 통해 SNS에 포스팅한 사진 이미지를 말한다.

2.1 자기 지시성

이미지의 자기 지시성이란, 보조 설명 없이 이미지만으로도 의미를 전달하고 소통할 수 있는 속성을 뜻한다. 즉, 이미지가 문자 언어와 관계없이도 그 자체로 의미를 지니어 음성 언어와 같은 역할을 수행할 수 있는 것을 의미한다.

2.2 기록성

이미지의 기록성이란 이미지가 과거의 사실을 기록해주어 객관적인 증거로 활용될 수 있음을 의미한다. 사용자들은 SNS에 이미지를 업로드하고 이를 보관하는 행위를 반복하는데 특히 “기록성”的 특징을 짚게 가진 이미지들은 해당 사용자가 언제, 어떤 곳을 방문해 무엇을 했는지 등을 타임라인 순대로 파악할 수 있어 사용자의 프로파일을 만들 수 있게 되며 형사 사건의 경우 알리바이에 해당된다. 이는 곧 개인정보 유출과도 직결된다.

2.3 정체성 표현

이미지를 통한 정체성 표현이란 사진을 찍는 주체가 사진이 될 만한 순간을 선택하고 촬영

하는 모든 과정에 걸쳐 자기 자신을 표현하는 것을 의미한다. 즉 스스로 어떤 순간에, 어디에서, 무엇을 촬영할지를 생각하고 결정한 결과가 이미지에 담기게 되기 때문에 이미지에는 그 사람이 보여주고자 하는 자신의 정체성이 표현된다. 특히 이러한 이미지들은 사용자의 취향, 관심사, 개인의 역사 등을 내포하게 된다.

2.4 간접 경험 제공

이미지를 통한 간접 경험이란 이미지를 보고 이미지 속의 세계를 실제와 같이 경험하고 상호작용하는 것을 의미한다. 대표적인 이미지 기반 SNS인 인스타그램의 경우 이미지만 올리는 것이 아니라 해시태그 등을 같이 작성하기 때문에 다른 사람들이 해시태그와 키워드 검색을 통해 관련 해당 이미지들을 접근할 수 있다. 이를 통해 검색된 이미지의 사용자에 대한 경험을 알 수 있게 되고, 이는 곧 해당 사용자의 행적을 비롯한 개인정보를 쉽게 파악할 수 있게 됨을 의미한다.

III. 이미지 속성과 개인정보

SNS에 나타나는 이러한 4가지 이미지 속성 중 개인정보를 가장 많이 나타내는 이미지 속성은 기록성과 정체성 표현이다. 특히 기록성의 경우 출업장, 청첩장, 명함, 상장 등 사용자가 특별하다 생각하는 증명서 등을 추억하기 위해 해당 이미지를 찍어 올리는 경우가 많다. 이러한 이미지의 경우 숨겨진 의미를 찾지 않더라도 이미지 자체에 이름, 생년월일 등의 개인정보가 텍스트 형식으로 남아있는 경우가 많아 각별한 주의가 필요하다.

정체성 표현의 경우 사용자가 구매한 물건, 다녀간 장소, 먹은 음식 등에 대한 이미지가 많다. 이러한 이미지들은 이미지 자체에서 개인정보를 직접적으로 인지해내기는 힘들지만 여러 가지 요소들을 조합하고 유추하면 생각보다 많은 개인정보 특히 민감정보와 알리바이를 파악하는 것이 가능하다. 예를 들어 이미지 내의 배경을 통해 다녀간 장소가 어디인지, 그 장소에 체류했던 시간대가 낮/밤 중 언제인지, 그 장소

에 누구와 함께 있었는지 등을 통해 타임라인별로 프로파일을 만들 수 있다. 한 사용자에 대해 이러한 타임라인별 프로파일이 쌓이게 되면 취향, 관심사, 자주 다녀간 장소, 자주 만나는 사람 등을 알 수 있으며 이를 통해 이후 향후 동선 예측도 가능하다. 프로파일을 만드는 작업은 사람이 수작업으로 진행하는 건 시간적으로나 비용 측면에서 있어서 다소 소모적이지만, 요즈음 많이 활용되는 딥러닝(Deep learning) 기술[6]을 이용하게 되면 이미지에서의 대상 추출, 대상의 특성 추출 등이 자동으로 이루어질 수 있으며 이를 통해 타임라인 상에서 마주치는 사람과 들르는 장소 등을 예측할 수 있게 된다.

이러한 과정을 통하여 본인도 모르는 개인정보 유출이 충분히 발생할 수 있다. 특히 SNS에 올라오는 이미지들은 불법적으로 수집하게 되는 이미지들이 아니라 “공개정보”의 한 요소로 취급되기 때문에 더욱 사용자의 각별한 주의가 필요하다. 유럽연합의 GDPR에서는 이러한 자동프로파일링을 금지하고 있으며[7], 우리나라 개인정보보호법에서도 공개된 정보라도 공개 목적 이외에 사용을 금지하고 있으나[8], 이처럼 이미지를 이용한 프로파일링 그리고 개인정보의 추출은 얼마든지 은밀하게 이루어질 수 있으며 그로 인하여 본인도 모르게 2차 사고로 이어질 수 있어서 그 위험성에 많은 관심을 가져야 한다.

IV. 결론

본 논문에서는 SNS 상에서의 공개된 이미지들의 속성을 알아보고 이미지에 나타나는 개인정보 유출 위험에 대해 알아보았다. 언택트 시대를 맞이하여 SNS 사용이 활발해지는 지금 무엇보다도 사용자가 개인정보 유출에 관한 경각심을 가지고 SNS를 사용하는 것이 중요하다. 또한 개인정보관리정책에서 이러한 공개정보에 대해 개인정보 침해·유출을 막기 위한 방향 제시가 필요하다.

향후 해당 연구를 바탕으로 딥러닝을 이용하여 공개된 이미지에서의 개인정보 유출 위험성

에 대해 연구하고 공개정보에서의 개인정보보호정책 방향성에 대해 논의하고자 한다.

[참고문헌]

- [1] 윤호열, 김민호, 이보라, & 최상옥. (2020). 개인정보침해에 대한 우려와 SNS 이용 행태 간 종단적 관계에 대한 연구: 온라인 정보 프라이버시 이론을 중심으로. 정보통신정책연구, 27(3), 93-119.
- [2] 임태민, & 이형석. (2019). 소셜 네트워크 서비스 (SNS) 에서의 개인정보 제공의도 및 지속적 이용의도에 미치는 영향요인. 인터넷전자상거래연구, 19(1), 17-38.
- [3] 안선희, “페이스북 개인정보 유출 한국 피해자 최대 8만 6천명,” 한겨레, (2018. 04. 06), Available at <http://www.hani.co.kr/article/economy/it/839508.html>
- [4] W. H. Lee. M. W. Yun. & J. S. Park (2013). Intelligence in the Internet Era: Understanding OSINT and Case Analysis. Journal of the Korean Society of Security and Security, (34), 259-278.
- [5] 윤지선, & 류한영. (2019). 이미지 기반 SNS에 나타난 이미지의 속성과 사용자 만족: 인스타그램과 펜터레스트를 중심으로. 한국 HCI 학회 논문지, 14(1), 5-13.
- [6] 김영형, 신기웅, & 이용환. (2015, June). 딥러닝 기반의 미래기술 전망. In Proceedings of KIIT Conference (pp. 219-220).
- [7] Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. Computer law & security review, 34(3), 436-449.
- [8] 개인정보보호법(2020. 8. 5. 법률 제 16930호로 개정 된 것)

안전한 자율협력주행을 위한 C-V2X 시스템 구축 방안

김서연*, 오인수**, 임강빈***

*순천향대학교(대학생), **순천향대학교(대학원생), ***순천향대학교(교수)

Study on C-V2X system for safe autonomous and cooperative driving

Kim Seo Yeon*, Insu Oh**, Kangbin Yim***

*Soonchunhyang University(Undergraduate student)

**Soonchunhyang University(Graduate student)

***Soonchunhyang University(Professor)

요약

안전한 미래 자동차를 위해 차량과 다른 차량, 인프라 및 모바일 기기와 통신하는 기술인 V2X 기술이 발전했고 미국과 유럽에서는 V2X를 위한 통신 표준인 WAVE와 C-V2X를 발표했다. 각 통신 표준에 대한 특징 및 한계점과 차이점을 설명하고 더 넓은 커버리지와 짧은 지연시간으로 효율성이 높은 C-V2X에 대한 보안 요구사항에 대해 설명한다. 또한 국내외 진행중인 V2X 및 C-V2X 사업을 통해 V2X 구축 현황을 살펴보고 실제 상용화된 V2X 시스템에 지원하는 보안 요소인 데이터 암호화와 인증서 관리에 대해 설명하고 V2X 시스템 구축을 위한 추가적인 요구사항에 대해 제안한다.

I. 서론

안전한 미래 자동차를 위한 기반 기술인 V2X(Vehicle to Everything communication)는 차량과 다른 차량, 인프라 및 모바일 기기 등과 무선통신하는 기술로 V2V(Vehicle to Vehicle), V2I(Vehicle to Infrastructure), V2N(Vehicle to Nomadic Device) 등을 총칭한다. V2X 기술은 초고속, 초저지연, 초연결성이 특징으로 차량 간 통신으로 위험한 상황을 알리거나 교통 신호 등의 인프라 정보를 알 수 있고 자율주행차의 핵심 센서인 레이더, 라이다, 카메라의 시야 내에서만 활용할 수 있다는 한계를 넘어 시야 확보가 어려운 사각지대나 기상 악화 상황에서 도 안전한 주행이 가능하다.[1]

본 논문에서는 V2X를 위한 통신 표준인 WAVE와 C-V2X의 특징 및 한계점과 C-V2X에 대한 요구사항을 설명한다. 또 자율협력주행 시스템 구축 동향 및 실제 구축된 C-V2X 시스템을 대상으로 보안 요구사항 검증 방법 및 안전한 보안 요구사항을 제안하고자 한다.

II. 자율협력주행 표준 및 특징

2.1 DSRC(Dedicated Short Range Communication)

DSRC는 무선 주파수를 사용하는 WI-FI 기반의 단거리 통신 방식으로 고속 이동 환경에 특화된 무선랜 기반 기술인 WAVE(Wireless Access in Vehicular Environment) 통신 표준을 기반으로 한다. 5.8Hz 주파수 대역으로 도로에 위치한 고정국인 RSU(Road Side Unit)의 안테나에 의해 형성되는 통신 가능 영역에 차량 내에 탑재된 단말기인 OBU(On-Board Unit)가 있는 차량이 통과할 경우 통신되는 방식의 고속 패킷 통신 시스템이다.[2]

DSRC는 수백 Kbps의 속도로 양방향 근거리 통신이 가능하며 좁은 무선 셀 지역에서 일-대-일, 일-대-다의 통신 방식을 사용할 수 있고 데이터 전송에 대해 무선 채널을 효율적으로 이용이 가능하다. LOS(Line of Sight)를 유지할 수 있는 통신 환경을 가지고, 속도 무선 패킷 데이터 전송이 가능하다.[3]

반면 DSRC는 근접한 경우에만 적용 가능하며 셀 사이의 간섭으로 주파수 재사용율이 저하된다. ASK 변조 방식 사용으로 전송속도가 낮고 짧은 커버리지로 통신 반경마다 설비 구축과 새로운 주파수 대역 사용으로 비용, 시간적으로 비효율적이라는 한계가 있다.[2][4]

2.2 C-V2X(Cellular-V2X)

3GPP-LTE Release 14 이동통신 표준에서 정한 LTE(Long-Term Evolution) 및 5G의 셀룰러 이동통신 기반 V2X 기술로 커버리지가 넓고 보안성이 뛰어나다. 효율적인 자원의 할당으로 높은 트래픽 용량을 제공하며 기지국 SPS 스케줄링 지원 시 혼잡 상황 해소 지원이 가능하기 때문에 차량 트래픽 혼잡 상황을 대응한다. 채널 센싱 기반 최적 전송 자원 선택 및 차등 QoS(Quality-ofService) 지원이 가능하다.[5]

C-V2X는 HARQ(Hybrid Automatic Repeat and request) 방식으로 수신 가능성이 높은 반면 스루풋(throughput)과 최대 통신 가능 차량 수가 절반이며 이동통신 기반으로 데이터를 사용하기 때문에 방대한 비용을 소모한다.[6][7]

2.3 C-V2X와 DSRC 비교[8][9]

	DSRC	C-V2X
표준	3GPP-LTE Release 14	IEEE 802.11 WAVE
통신 방법	Cellular(LTE, 5G)	WiFi
커버리지	LTE 기지국 1~5km(전국망)	250~300m(별 도 기지국)
모빌리티 지원 항목	시속 500km	시속 200km
대역폭	최대 75Mbps(10Mhz 기준)	최대 27Mbps(10Mhz 기준)
지연시간	0.1초 미만(LTE) 0.01초 미만(5G)	0.1초 미만
데이터 전송속도	100Mbps(LTE) 20Gbps(5G)	최대 27Mbps
신뢰성	95~99%(LTE) 99.9~99.999%(5G)	95~99%

III. C-V2X 보안 요구사항

3.1 C-V2X 보안 요구사항

3GPP-LTE Release 14의 TS 33.185에서 보안 요구사항을 제시했다. 발신자 식별로 비인가

차량의 데이터 수신 방지하고 통신에 참여하는 모든 UE(User Equipment)에 대해 인증 표준 설정 및 사용자 인증을 한다. V2X 제어 기능과 UE 사이의 구성 데이터 전송은 무결성 및 기밀성 보호가 필요하며 UE와 PLMN(Public Land Mobile Network)의 상호 인증이 필요하다. 단 말 간 통신 인터페이스인 PC5로 전송된 데이터는 다른 엔티티에 의한 식별과 UE ID 추적을 허용하지 않는다. V2X 메시지의 식별자는 UE 와 사용자 ID 유출 위험을 최소화하며 가명을 사용한 신원 보호와 V2X 메시지의 일부로 전송되는 응용 계층의 UE ID의 도청을 보호한다. OTA(Over-the-air) 방식의 데이터 전송 보호 및 최신 보안 메커니즘 사용과 보안 표준을 사용하는 서로 다른 운영자의 차량 UE간 상호 운용성 유지한다. 또한 UE간 교환되는 메시지는 주기적인 자격 증명이 필요하며 3GPP 인증, 키 승인, 가입자 자격 증명 및 가입자의 ID가 네트워크에 액세스하는데 사용되는 경우 V2X 지원 UE 내의 USIM 또는 UICC에 포함한다.[10]

IV. 안전한 V2X 시스템 구축을 위한 요구사항

4.1. 자율협력주행 시스템 구축 현황

실제 구축된 V2X 인프라는 현재 국내에서는 DSRC를 지원하는 C-ITS 사업을 통해 대전시-세종시에 90.7km 구간에 설치되었으며[11] 미국에서는 Qualcomm과 Commsignia와 함께拉斯베이거스 특정 도로 구간에 C-V2X 인프라를 구축하고 있다. 대표적인 V2X 시스템 개발 제조사 A에서는 각국의 도시에 V2X 시스템을 구축하기 위한 사업을 진행하고 있다.[11]

4.2. 실제 V2X 시스템에 필요한 보안 요소

제조사 A의 V2X 시스템은 C-V2X와 V2X를 모두 지원하며 요구에 따라 미국 표준과 유럽 표준을 사용할 수 있다. DSRC 표준 메시지는 V2X 및 V2I 간 데이터 교환을 위한 메시지, 데이터 프레임 및 데이터 요소와 형식 등 구조를 정의하는 SAE J2735에 따라 전송하며 V2X 메시지는 무선 전송으로 주변 모든 노드들이 수신 및 해석 가능하여 보안이 필요하다.[12]

4.2.1 데이터 암호화

데이터에 대한 기밀성 보장을 위해 암호화를 해야하며 신뢰할 수 있는 사용자만 알고 있는 키를 통해 대칭, 비대칭 암호화를 사용한다. 모든 메시지의 무결성 및 유효성을 메시지 수신 후 암호화 검증을 통해 확인한다.

4.2.2 인증서 관리

V2X 시스템을 통해 통신하는 모든 사용자를 신뢰하기 위해 공개키가 포함된 인증서를 사용하며 SCMS(Security Credential Management System)을 활용하여 V2X 메시지 서명에 사용하는 인증서를 주기적으로 관리한다. 개인정보를 보호하고 사용자를 유추하지 못하기 위해 임의로 선택된 여러개의 인증서를 제공한다.

4.3. C-V2X 시스템 보안 요구사항

시스템 구축을 위해 필요한 요구사항에는 표준 보안 요구사항을 충족하고 있는지에 대한 평가와 데이터 암호화를 위한 대칭키 및 비대칭키 분배 방법과 안전한 채널 주파수를 통한 메시지 전달이 필요하며 검증을 위한 키의 안전한 보관 메커니즘 방법, 인증서 생성 및 폐기, 갱신의 안전한 인증 관리 방법이 있다.

V. 결론

V2X 표준인 C-V2X와 DSRC의 표준화가 완료되고 상용화를 위해 준비중이다. V2X 관련 표준 및 상용화된 V2X 시스템 분석을 통해 인증서, 데이터 암호화가 구현된 것을 확인할 수 있다. 그러나 안전한 V2X를 위해 주파수 등의 무선 통신을 사용하기 때문에 전파 방해 공격, 도청, 개인정보유출 등에 대해 추가적으로 요구 사항을 적용하여 적절히 보호할 필요가 있으며 안전한 자율주행 구현에 적용할 수 있다.

Acknowledgement

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2019-0-01343, 융합보안핵심인재양성), 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2021R1A4A2001810).

[참고문헌]

- [1]"V2X 통신의 이해." 수정, 2021-08-13 접속, <https://www.fescaro.com/ko/archives/467>.
- [2] 차종환, “자율주행 상용화 앞두고 국가표준 채택 ‘고심’”, 정보통신신문, 2021-08-17 접속, <https://www.koit.co.kr/news/articleView.html?idxno=77313>.
- [3] 장재득, 김태중, 자율주행차량을 위한 밀리미터파 무선통신 기반 차량 사물 통신 기술, 2017.
- [4] 문형돈, 이재환, 이동일, 국내외 DSRC 기술 및 표준화 동향, 2009.
- [5] 정성훈, C-V2X 서비스 프레임워크 – 네트워크 아키텍처와 통신 절차 2019.
- [6] 교통과학연구원, 자율주행차 핵심기술 : V2X, 2016.
- [7] 차종환, “국토부도 5G-V2X…ITS 시장 중기 입지 ‘흔들’”, 정보통신신문, 2021-08-17 접속, <https://www.koit.co.kr/news/articleView.html?idxno=82371>.
- [8] kimley-horn, “DSRC 및 C-V2X: 커넥티드 차량의 유사점, 차이점 및 미래”, Douglas G et t m a n , 2 0 2 1 - 0 8 - 1 7 접속, <https://www.kimley-horn.com/dsrc-cv2x-comparison-future-connected-vehicles/>
- [9] 김일규, 3GPP 5G NR V2X 표준화 동향, 2019
- [10] 3GPP. Security aspect for LTE support of Vehicle-to-Everything (V2X) services (3GPP TS 33.185 version 14.0.0 Release 14). n.p.: ETSI, 2017.
- [11] C-ITS, <https://www.c-its.kr/>, 2021-08-03 접속
- [11] THEELEC, “미국 라스베이거스, 웰컴파 C-V2X 인프라 구축”, 2021-08-17 접속, <http://www.thelec.kr/news/articleView.html?idxno=549>

클라우드 웹 애플리케이션을 위한 SVM 기반 데이터베이스 공격 탐지 연구

조재한*, 김지연*[†] (교신저자)

*대구대학교 컴퓨터정보공학부 컴퓨터공학전공 (학부생)

*[†] 대구대학교 컴퓨터정보공학부 컴퓨터공학전공 (교수)

SVM-based Detection of Database Attacks
for Cloud Web Applications

Jae-Han Cho*, Jiyeon Kim*[†]

*Dept. of Computer Engineering, Daegu University
(Undergraduate student)

*[†] Dept. of Computer Engineering, Daegu University (Professor)

요약

클라우드 컴퓨팅 시장이 성장하면서 클라우드 서비스를 활용하여 웹 애플리케이션을 운영하는 사용자가 증가하고 있다. 클라우드 웹 애플리케이션은 일반적으로 N-tier 구조로 설계되기 때문에 안전한 애플리케이션 운영을 위해서는 각 서버의 상태를 실시간으로 모니터링하고, 사이버 공격 발생 시, 신속히 탐지 및 복구하는 것이 필요하다. 본 논문에서는 웹 서버에서 별도의 공격 탐지 솔루션 없이 각 서버의 공격을 탐지할 수 있는 새로운 공격 탐지 모델을 제안한다. 제안된 모델은 웹 서버의 성능 메트릭을 머신러닝 모델 중, SVM(Support Vector Machine) 기반으로 학습하고, 학습된 모델을 활용하여 데이터베이스 서버에 발생하는 공격을 실시간으로 탐지하게 된다. 특히, 클라우드 웹 애플리케이션 유형별로 경험 데이터를 학습하여 정상 및 공격 모델을 정의하기 때문에 기존의 규칙기반 탐지 기술에 비해 오탐율을 낮출 수 있다는 장점을 가진다.

I. 서론

2021년 5월 발표된 IDC 시장 조사 결과에 따르면, 클라우드 컴퓨팅 시장은 전년 대비 24% 이상 성장하였다[1]. 클라우드 서비스를 도입하는 사용자가 빠르게 증가하면서 온 프레미스 환경(on-premise)에서 운영되던 수많은 웹 애플리케이션이 클라우드

환경으로 이전하고 있으며 동시에 이러한 클라우드 웹 애플리케이션을 대상으로 하는 사이버 공격 또한 증가하고 있다.

클라우드 웹 애플리케이션은 일반적으로 웹 서버, 데이터베이스 서버 등 특정 기능을 담당하는 서버들이 독립적으로 동작하는 N-tier 구조로 설계된다. 본 논문에서는 N-tier를 구성하는 각 서버에 개별 보안 솔

루션을 운영하지 않아도 클라이언트의 서비스 요청을 수신하는 웹 서버에서 다른 서버에 발생한 공격을 신속히 탐지할 수 있는 클라우드 웹 애플리케이션 공격탐지 모델을 제안한다.

본 논문에서는 클라우드 웹 애플리케이션을 구축하고, 웹 서버 성능 메트릭을 정상적인 운영상태 및 데이터베이스 삭제 공격 상태에서 수집한다. 수집된 메트릭은 머신러닝 모델 중, SVM(Support Vector Machine) 기반으로 학습하여 데이터베이스 삭제 공격을 위한 실시간 탐지 모델로 개발한다. 또한, 규칙기반 탐지 기술과의 융합을 비교함으로써 본 논문에서 제시한 머신러닝 모델의 성능을 검증한다.

본 논문의 구성은 다음과 같다. 2장에서는 최신 데이터베이스 취약점 및 데이터베이스 공격방법에 대해 살펴보고, 3장에서는 제안된 SVM 기반의 데이터베이스 공격 탐지 모델을 개발한다. 4장에서는 개발된 모델의 성능을 검증하기 위해 공격 탐지 정확성을 분석하고, 5장에서 결론 및 향후연구 계획을 제시한다.

II . 관련 연구

본 장에서는 2021년에 발견된 데이터베이스 취약점 및 이러한 취약점을 악용하는 주요 데이터베이스 공격 방법을 살펴본다.

2021년 1월부터 7월까지 데이터베이스 취약점은 약 90개 이상 발견되었다[2]. 이러한 취약점을 악용한 데이터베이스 공격은 권한 공격 및 SQL 삽입 공격이 대표적이며 SQL 삽입 공격은 권한 탈취를 위한 수단으로도 이용된다. 권한 및 SQL 삽입 공격과 관련된 취약점으로는 CVE-2021-2336[3], CVE-2021-38159[4] 등 권한 및 SQL 삽입 공격과 관련한 많은 취약점이 발견되고 있다. 이는 데이터베이스 취약점을 악용한 공격이 빈번히 지속적으로 발생하고 있으며 이러한 공격에 대응하기 위한 데이터보안 기술 마련의 중요성을 보여준다.

N-tier 구조의 웹 애플리케이션에서 SQL 삽입 공격은 웹 서버, 애플리케이션 서버, 데이터베이스 서버에 걸쳐 이루어진다 [5]. SQL 삽입 공격 유형에는 논리적으로 잘못된 SQL 문을 삽입하여 기본 오류 페이지를 반환하게하거나, 인증을 우회하거나 데이터를 추출하기 위해서 데이터베이스의 항목에 대하여 항상 참으로 평가되는 쿼리를 삽입하는 동어 반복, UNION SELECT를 삽입하여 의도한 것과 다른 테이블에서 데이터를 반환하도록 속이는 유니온 쿼리 삽입 공격, 데이터 추출, 수정, 서비스 거부 또는 원격 실행을 위해 원본 쿼리에 추가 쿼리를 삽입하여 공격하는 피기백(Piggy-Back)공격 등이 존재한다[6]. 이러한 SQL 삽입 공격을 탐지하기 위해서는 필터링을 통해 특수문자를 제거하는 방법, 웹 애플리케이션의 SQL 쿼리문을 정적으로 분석하는 방법, 웹 애플리케이션을 스캔한 후 응답을 분석하여 공격을 탐지하는 동적 해석 방법, 정적 분석 방법과 동적 해석 방법의 장점을 결합하여 공격을 탐지하는 방법, 머신러닝을 이용하여 공격을 탐지하는 방법 등이 존재한다[7]. 그러나 이러한 방법들은 공격 패턴을 사전에 정의해야 하는 절차가 선행되어야 한다는 단점이 존재할 뿐 아니라, 알려지지 않은 취약점을 악용한 재로데이 공격 또는 우회 공격에 취약하다. 따라서 본 논문에서는 데이터베이스 서버가 아닌, 웹 서버의 성능 메트릭을 활용하여 데이터베이스 공격을 탐지하는 방법을 제시하고자 한다.

III . 머신러닝 기반 데이터베이스 공격탐지 모델 개발

본 장에서는 웹 서버의 성능 메트릭을 머신러닝 기반으로 학습하여 데이터베이스 공격 탐지 모델로 개발하기 위한 실험 환경 및 데이터셋 수집 방법을 설명한다.

3.1 실험 구성

웹 애플리케이션 환경을 구축하기 위하여 클라이언트, 웹 서버, 데이터베이스 서버로 구성되어 있는 클라우드 웹 게시판 애플리케이션인 RUBBoS[8]를 사용하였다. 데이터베이스 공격으로 인하여 데이터베이스가 삭제된 환경을 공격 시나리오로 하여 실험을 진행하였고, 다양한 클라우드 애플리케이션의 워크로드 하에서 실험을 하기 위해서 [표 1]과 같이 게시판에 접근하는 클라이언트 수와 게시글을 등록하는 Authors 수를 다양하게 조합하여 실험을 진행하였다.

Num. of Clients	Num. of Authors		
	250	400	500
30	c30_250 benign	c30_400 benign	c30_500 benign
	c30_250 attack	c30_400 attack	c30_500 attack
50	c50_250 benign	c50_400 benign	c50_500 benign
	c50_250 attack	c50_400 attack	c50_500 attack
100	c100_250 benign	c100_400 benign	c100_500 benign
	c100_250 attack	c100_400 attack	c100_500 attack

(표 1) Client 수 및 Authors 수를 고려한 워크로드

3.2 실험 데이터

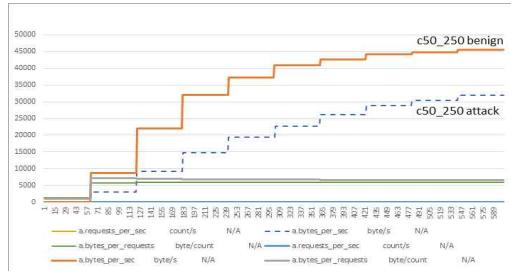
본 실험은 구축한 클라우드 웹 애플리케이션 환경에서 웹 서버의 성능 메트릭을 수집하였다. 웹 서버에서의 성능 메트릭을 수집하기 위해서 넷플릭스에서 개발한 Performance Co-Pilot (PCP)를 이용하였고, 총 17개의 메트릭을 실시간 수집하였다. 또한, 정상인 환경에서의 메트릭과 공격 환경에서의 메트릭에 대하여 머신러닝의 학습을 위해 [표 2]와 같이 데이터를 나누어 진행하였다.

Dataset	Benign	Attack	Total
Training set	420	420	840
Testing set	180	180	360

(표 2) SVM 기반 학습에 사용된 데이터 샘플 수

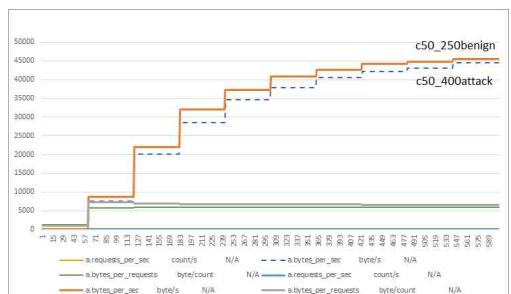
3.3 실험 결과

본 실험에서는 다양한 워크로드 환경에서 실험을 진행하기 위하여 글을 저장하는 사용자인 Authors 수와 실제 데이터베이스 사용자 수를 나타내는 클라이언트 수를 변경해가며 수집을 진행하였다. [그림 1]은 클라이언트 수와 Authors 수가 동일한 환경에서의 정상 데이터와 공격 데이터와 두 데이터의 비교 결과에 대해 확인할 수 있다. [그림 1] 그래프에서 확인할 수 있듯이 Authors 수가 동일한 환경에서는 정상 데이터와 공격 데이터를 규칙 기반 탐지 방법을 사용하여 탐지하여도 충분히 탐지 가능한 형태를 보이고 있다.



(그림 1) Client 수 50명, Author 수 250명 워크로드에서의 정상, 공격 비교 그래프

이후 클라이언트 수와 Authors 수를 변경해 성능 메트릭을 수집하고 비교하는 절차를 진행하였다. 이때 정상 데이터와 공격 데이터가 매우 비슷하게 수집되는 경우를 발견할 수 있었으며 [그림 2]는 그 예시를 보여주고 있다.



(그림 2) Client 50aud, Author 수 250명(정상), 400명(공격) 데이터 비교 그래프

이와 같이 성능 메트릭이 비슷한 수치로 수집이 되는 경우 규칙 기반 탐지 방법으로는 데이터베이스 공격 탐지에 어려움이 생긴다. 이에 본 실험에서는 이와 같은 상황을 대비하여 두 성능 메트릭을 머신러닝 알고리즘의 한 종류인 SVM을 활용하여 비교하는 실험을 진행하였다. 모델의 성능을 측정하기 위해 정밀도(Precision), 재현율(Recall), F1-score를 활용하였다. 위와 같이 SVM을 이용하여 비슷한 수치를 보이는 두 성능 메트릭에 대해 학습시킨 결과 [표 3]과 같은 지표를 얻을 수 있었다.

Performance	SVM
Precision	0.97
Recall	0.97
F1 Score	0.97

[표 3] SVM 기반 공격 탐지 정확도

위의 지표 결과로 성능 메트릭의 수치가 비슷하게 수집될 경우 SVM을 활용하여 공격을 탐지하는 방법이 기존 규칙 기반 탐지 방법에 비해 더 높은 탐지율을 보인다.

IV . 결론

본 논문에서는 클라우드 환경에서의 웹 애플리케이션 공격탐지 방법에 대해 제시했다. 웹 애플리케이션은 서로 통신을 하는 N-tier로 구성되어 있으며 이러한 웹 애플리케이션 특징상 각 계층에 공격이 발생했을 경우 원활한 서비스를 제공하지 못하게 된다. 따라서 웹과 데이터베이스의 보안이 중요해진다. 본 논문에서는 넷플릭스에서 개발한 웹 성능 메트릭 수집 도구인 PCP(Performance Co-Pilot)를 이용하여 웹 애플리케이션의 웹 서버에서의 성능 메트릭을 수집하여 머신러닝 모델 중, 하나인 SVM으로 학습시켜 데이터베이스 공격을 탐지하는 방법을 제시하였다. 실험 결과에서 볼 수 있듯이 제안하는 탐지 방법은 다양한 워크로드에서 기존에 사용되던 탐지 방법의 사전 정의가 없어도 데이터베이스 공격에 대하여 높은 탐지율을 보였다. 향후에는 SVM 뿐 아니라, 다양한 머신러닝 모델을 기

반으로 클라우드 웹 애플리케이션의 데이터베이스 공격을 웹 서버의 성능 메트릭을 활용하여 탐지하는 연구를 수행할 예정이다.

Acknowledgement

이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2018R1D1A1B07050543)

[참고문헌]

- [1] [https://www.idc.com/getdoc.jsp?
containerId=prUS47685521](https://www.idc.com/getdoc.jsp?containerId=prUS47685521)
- [2] <https://cve.mitr.org>
- [3] [https://cve.mitre.org/cgi-bin/
cvename.cgi?name=CVE-2021-2336](https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2021-2336)
- [4] [https://cve.mitre.org/cgi-bin/
cvename.cgi?name=CVE-2021-38159](https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2021-38159)
- [5] Inyong Lee, Soonki Jeong, Sangsoo Yeo, Jongsub Moon,A novel method for SQL injection attack detection based on removing SQL query attribute values, Mathematical and Computer Modelling, 58–68, January 2012
- [6] Harefa, J., Prajena, G., Alexander, A. M., Dewa, E. V. S., & Yuliandry, S. SEA WAF: The Prevention of SQL Injection Attacks on Web Applications, March 2021
- [7] Sivasangari, A., Jyotsna, J., & Pravalika, K. SQL Injection Attack Detection using Machine Learning Algorithm. In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) (pp.1166–1169). IEEE, June 2021
- [8] [https://github.com/michaelmior
/RUBBoS](https://github.com/michaelmior/RUBBoS)

WLAN IoT 보안을 위한

실시간 이상탐지 시스템

이승욱*, 김지연*[†] (교신저자)

* 대구대학교 컴퓨터정보공학부 컴퓨터공학전공 (학부생)

*[†] 대구대학교 컴퓨터정보공학부 컴퓨터공학전공 (교수)

Seung-wook Lee*, Jiyeon Kim*[†]

Live Anomaly Detection System

for WLAN IoT Security

*Dept. of Computer Engineering, Daegu University

(Undergraduate student)

*[†] Dept. of Computer Engineering, Daegu University (Professor)

요약

스마트 홈, 스마트 헬스케어, 자율 자동차 등 4차산업 혁명시대 핵심기술들은 사물인터넷 기반으로 동작한다. 사물인터넷은 다양한 유형의 사물들이 인터넷에 연결된 네트워크로서 인터넷 통신을 통해 연산 및 사물 제어를 수행한다. 따라서 안전한 사물인터넷 운영을 위해서는 사물인터넷 환경에서 발생할 수 있는 다양한 네트워크 공격을 사전에 예측하고, 이를 방어하기 위한 대응기술을 마련하는 것이 필요하다. 본 논문에서는 WLAN(Wireless Local Area Network)에 속한 사물들의 안전한 통신을 위하여 IoT 환경에서 발생할 수 있는 주요 공격인 분산서비스거부 공격 및 중간자 공격을 실시간 탐지할 수 있는 이상탐지 시스템을 개발한다.

I. 서론

4차산업 혁명시대를 이끌어가는 스마트 홈, 스마트 오피스, 스마트 헬스케어 등의 기술은 사물인터넷(Internet of Things, 이하 IoT)을 기반으로 동작한다. IoT 환경에서 각 IoT 기기들은 인터넷으로 연결된 서버와의 지속적인 통신을 통해 사물에서 수집된 데이터를 가공 및 연산하고, 연산 결과를 반영하여 IoT 기기의 상태를 제어할 수 있다. 최근에는 다양한 스마트 기기가 등장하면서 개인도 쉽게 IoT 환경을 구축 할 수 있지만, 안전한 IoT 환경 구축을 위한 보안 솔루션의 보급은 미비한 실정이다.

개인이 구축하는 홈 IoT의 경우, 무선 공유

기를 활용한 WLAN(Wireless Local Area Network) 환경에서 IoT 기기들이 운영되는데 이러한 소규모의 IoT 환경에서도 무선 단말의 설정 취약점, IoT 기기 취약점 등을 악용한 다양한 네트워크 공격들이 발생할 수 있다. 예를 들어, 무선 공유기의 초기 설정을 변경하지 않는 경우, WLAN 내의 트래픽을 위변조하여 잘못된 IoT 기기 제어를 수행할 수 있고, 보안 솔루션이 탑재되기 어려운 경량의 IoT 기기들의 취약점을 악용하여 기기를 해킹하거나 파손하는 것이 가능하다.

본 논문에서는 WLAN 기반의 IoT 환경에 발생할 수 있는 주요 네트워크 공격인 분산서

비스거부(Distributed Denial of Service, 이하 DDoS) 공격, 중간자(Man-in-the-Middle, 이하 MITM) 공격을 실시간 탐지하고, 사용자들이 신속히 공격을 인지하고 대응할 수 있도록 웹 기반의 모니터링 인터페이스를 제공하는 실시간 이상탐지 시스템을 개발한다. 시스템에 탑재되는 DDoS 공격 탐지 모델은 딥 러닝 모델 중, LSTM(Long-Short Term Memory) 모델을 사용하여 개발되고, MITM 공격 탐지 모델은 ARP 패킷 분석을 통해 위변조 여부를 판단하여 공격을 탐지한다. 제안된 모델의 성능을 검증하기 위하여 본 논문에서는 가상의 IoT 환경을 구축하고, 실제 DDoS 공격 및 MITM 공격을 주입하면서 이상탐지 시스템의 탐지 결과를 관찰한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존에 수행된 딥 러닝 기반 IoT 연구를 살펴보고, 3장에서는 본 연구에서 제안하는 DDoS 공격 탐지 모델 및 MITM 공격 탐지 모델을 개발한다. 4장에서는 두 모델을 탑재한 이상탐지 시스템을 구현하고, 실제 공격 주입을 통해 시스템의 공격 탐지 결과를 분석한다. 마지막으로 5장에서는 결론 및 향후연구를 제시한다.

II. 관련 연구

IoT 공격을 탐지하기 위한 기존 연구로서 가정용 IoT 보안을 위해 라즈베리와 하둡을 이용하여 빅데이터 기반의 이상탐지 모델을 제안한 연구[1], ISCX2012 데이터셋을 이용하여 딥 러닝 기반의 DDoS 공격 탐지를 수행한 연구[2], N-BaIoT 데이터셋을 활용하여 딥 러닝 모델들의 IoT 공격 탐지 성능을 비교한 연구[3] 등이 존재한다. 그러나 공개 데이터셋을 활용한 기존 연구들은 제안된 모델의 성능을 IoT 환경에서 검증하고 있지 않기 때문에 IoT 보안 솔루션으로서의 활용성을 검증하기 어렵다.

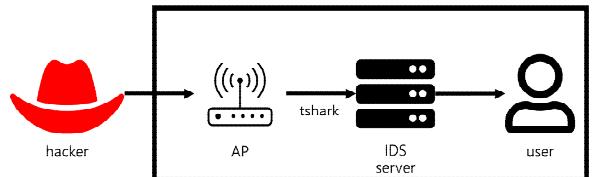
본 연구에서는 제안된 이상탐지 시스템이 IoT 환경에서 잘 동작하는지 검증하기 위하여 가상의 IoT 환경을 구축하고, 실시간 트래픽을 수집하여 시스템 성능을 검증한다는 점에서 기존 연구들과의 차별점이 있다.

III. 실시간 이상탐지 시스템 설계

본 장에서는 WLAN IoT 환경에서 발생 가능한 DDoS 공격 및 MITM 공격을 탐지하기 위한 모델을 개발한다.

3.1 이상탐지 시스템 개요

본 논문에서 제안하는 실시간 이상탐지 시스템은 <그림 1>과 같이 구성된다. WLAN IoT 환경에는 AP(Access Point) 단말과 IDS 서버, 그리고 실시간으로 IoT 환경의 상태를 웹 기반으로 모니터링하는 사용자가 존재한다. IDS 서버에는 본 논문에서 개발하는 DDoS 공격 탐지 모델 및 MITM 공격 탐지 모델이 탑재된다.



<그림 1> 이상탐지 시스템 동작 원리

AP에서는 패킷 캡쳐 도구인 tshark을 활용하여 IoT 기기 간에 송수신되는 패킷을 실시간 수집하고, 패킷의 정보를 가공하여 CSV 파일로 저장한다. IDS 서버는 이 CSV 파일을 실시간 읽어들여 DDoS 및 MITM 공격 여부를 판단하게 되고, 공격이 탐지된 경우에는 웹 인터페이스를 통해 사용자에게 실시간 알림을 준다.

3.2 LSTM 기반 DDoS 공격 탐지 모델

본 논문에서는 DDoS 공격 탐지 모델을 개발하기 위하여 DDoS 공격 데이터셋인 CIC-DDoS2019와 4000명 이상의 사용자 워크로드로 구성된 정상 데이터셋 Mendeley[4]를 LSTM 모델로 학습하였다. CIC-DDoS2019는 Canadian Institute of Cybersecurity UNB에서 개발한 데이터셋으로 총 85개의 특징(feature)을 갖는다. 본 논문에서는 85개 특징 중, 본 논문에서 구축한 가상 IoT 환경에서 수집 가능한 10개의 특징을 선별 학습함으로써 제안된 DDoS 공격 탐지 모델의 성능을 IoT 운영환경에서 검증하고자 하였다.

<표 1>은 DDoS 공격 탐지를 위해 LSTM 기반으로 학습한 10개의 패킷 특징을 보여준다.

ID	Feature
1	Timestamp
2	Total length of forward packets
3	Total length of backward packets
4	Forward packet length mean
5	Backward packet length mean
6	Source IP
7	Source port
8	Destination IP
9	Destination port
10	Protocol

<표 1> LSTM 기반의 DDoS 공격 탐지모델 개발을 위한 10개의 패킷 특징

제안된 LSTM 기반 DDoS 탐지 모델은 Yuan의 논문[2]의 모델을 참조하여 개발하였다. 정상 패킷 5만개와 DDoS 공격 패킷 5만개 샘플을 LSTM 기반으로 학습한 결과, <그림 2>와 같이 96% 이상의 정확도로 정상 및 공격을 구분해내는 모델이 개발된 것을 볼 수 있다.

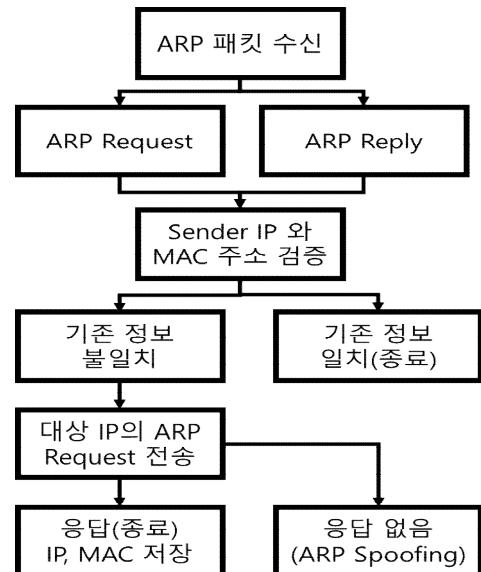


<그림 2> LSTM 기반 DDoS 공격 탐지모델 생성 과정

3.3 MITM 공격 탐지 모델

본 논문의 이상탐지 시스템은 ARP(Address Resolution Protocol) 통신 과정에서 발생할 수 있는 MITM 공격인 ARP 스푸핑(spoofing) 공격 탐지 기능도 포함한다. ARP Request 및 Reply 패킷이 수신되면, 기존의 ARP Table에 저장된 Sender의 IP 및 MAC 주소와 비교하여 기존 정보와 일치하는 경우에는 정상적인 요청으로 처리하고, 불일치하는 경우 즉, MAC 주소

가 변경된 경우에는 ARP Request를 재전송함으로써 공격 여부를 탐지한다. <그림 3>은 제안된 ARP 스푸핑 기반 MITM 공격 탐지 알고리즘을 순서도를 통해 표현한 것이다.



<그림 3> MITM 공격 탐지 알고리즘

IV. 실시간 이상탐지 시스템 구현

본 장에서는 앞에서 설계한 두 공격 탐지 모델을 이상탐지 시스템에 탑재하고, 실제 공격을 주입하면서 개발된 시스템의 설계 및 동작을 검증하고자 한다.

4.1 시스템 개발 환경

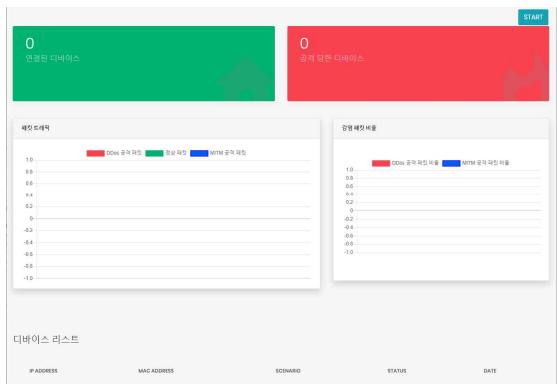
실시간 이상탐지 시스템은 우분투 가상머신에 개발하고, 접근성을 높이기 위하여 웹 기반의 사용자 인터페이스를 제공한다. 사용자는 제공된 웹 인터페이스를 통해 실시간 네트워크 상태 모니터링을 수행하고, 공격 발생 시, 실시간 알림을 수신할 수 있다. <표 2>는 본 연구에서 제안하는 실시간 이상탐지 시스템 개발 환경을 보여준다.

Type	Specification
OS	Ubuntu 18.04
CPU	1core
Memory	2GB
HDD	20GB
Python	3.8

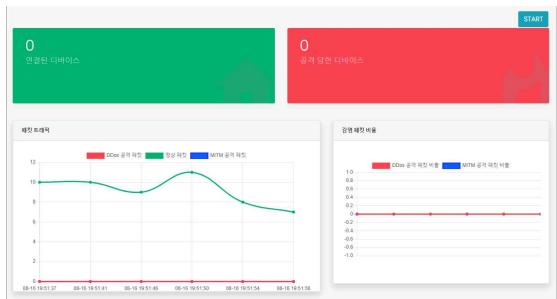
<표 2> 실시간 이상탐지 시스템 개발환경

4.2 웹 인터페이스 기반 시스템 동작 검증

사용자에 의해 이상탐지 시스템이 시작되면 연결된 기기 목록 및 수집된 패킷의 상태를 확인 가능하다. 만일, 공격이 탐지되면 웹 브라우저를 통해 실시간 사용자 알림을 전송할 수 있으며 MITM 공격의 경우에는 공격당한 기기의 정보 또한 확인 가능하다. 실시간 이상탐지 시스템의 메인 화면은 <그림 4>와 같으며 우측 상단의 파란색 Start 버튼을 누르면 <그림 5>와 같이 실시간 탐지가 시작된다.

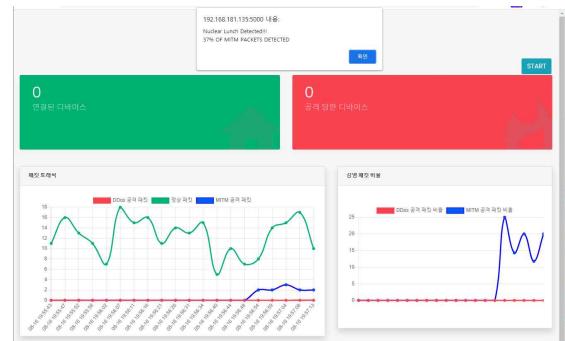


<그림 4> 실시간 이상탐지 시스템
메인 화면



<그림 5> 실시간 이상탐지 시스템
모니터링 화면

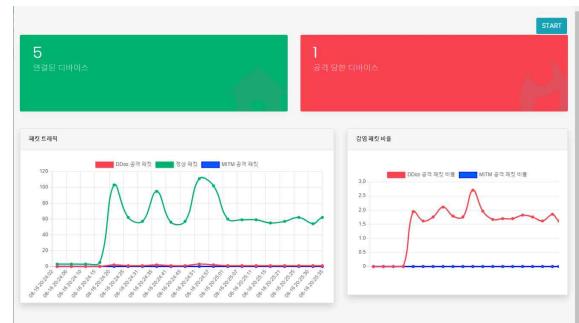
웹 인터페이스는 매 3초마다 수집된 패킷을 처리 및 분석하여 좌측에 정상 패킷(초록색), DDoS 공격 패킷(빨간색), MITM 공격 패킷(파란색)을 그래프로 보여준다. 그래프는 패킷 유형별 수집된 패킷 수를 의미한다. 우측 화면에는 전체 패킷 중, DDoS 공격 패킷 비율, MITM 공격 패킷 비율이 그래프로 보여진다.



<그림 6> 실시간 이상탐지 시스템의
MITM 공격 탐지 화면

<그림 6>은 Kali 리눅스의 ARP spoof를 이용하여 MITM 공격을 시도한 후, 실시간 이상탐지 시스템에서 관찰한 결과 화면이다.

공격 시도 전에는 좌측에 초록색의 정상패킷만 보여졌지만, 공격이 시도된 후에는 MITM 공격을 나타내는 파란색 그래프가 보여지는 것을 양쪽 그래프에서 모두 확인 가능하다.



<그림 7> 실시간 이상탐지 시스템의
DDoS 공격 탐지 화면

<그림 7>은 Kali 리눅스에 설치된 hping3 도구를 이용하여 SYN Flooding 공격을 시도한 후, 실시간 이상탐지 시스템에서 관찰한 결과화면이다. 우측 화면을 보면, 공격 주입 후, DDoS 공격 패킷 비율을 나타내는 빨간색 그래프가 급격히 증가하는 것을 볼 수 있다.

위 실험결과와 같이 본 논문에서 구현한 이상탐지 시스템은 DDoS 공격 및 MITM 공격을 개발된 탐지 모델을 기반으로 신속하고도 정확하게 탐지하고 있음을 확인할 수 있다.

V. 결론 및 향후연구 계획

본 논문에서는 WLAN IoT 환경에서 발생할 수 있는 DDoS 공격 및 MITM 공격을 실시간 탐지할 수 있는 이상탐지 시스템을 개발하였다.

시스템에 탑재된 DDoS 공격 탐지 모델은 LSTM 기반으로 총 10만개의 정상 및 공격 샘플을 학습하여 생성하였고, ARP 스푸핑 기반의 MITM 공격 탐지 모델은 ARP table과 실시간 ARP 패킷을 비교·분석하여 공격을 탐지하도록 개발하였다. 두 모델의 성능을 검증하기 위하여 본 논문에서는 가상 IoT 환경을 구축하고, AP로부터 캡처한 패킷을 두 모델을 기반으로 실시간 검사하도록 하였다. 실제 DDoS 및 MITM 공격을 주입하여 관찰한 결과, 이상탐지 시스템은 신속하고도 정확하게 두 공격을 탐지해내는 것을 확인하였다.

본 논문에서 개발한 시스템은 경량으로 동작 할 뿐 아니라, 웹 인터페이스를 통해 쉽게 사용이 가능하기 때문에 소규모의 IoT 환경을 운영하는 개인 및 기업이 쉽게 도입할 수 있다는 장점이 있다.

향후에는 DDoS 공격 및 MITM 공격 외에도 다양한 IoT 공격을 탐지할 수 있는 모델을 개발하여 시스템의 기능을 확장하고, 가상 IoT 환경이 아닌 실제 IoT 기기를 활용한 성능평가를 통해 제안된 시스템의 성능을 개선할 계획이다.

Acknowledgement

이 논문은 2018년도 정부(교육부)의 재원으로
한국연구재단의 지원을 받아 수행된
기초연구사업임 (No.2018R1D1A1B07050543)

[참고문현]

- [1] 박연진, 오주혜, 이근호, 전유부. n.d. 가정용 IoT 네트워크에서의 이상 정후 탐지 솔루션 제안. n.p.: 한국정보처리학회 2016년도 추계 학술발표대회 2016 Oct. 27
- [2] X. Yuan, C. Li and X. Li, "DeepDefense: Identifying DDoS Attack via Deep Learning," 2017 IEEE International Conference on Smart Computing, 2017.

- [3] J. Kim et al., "Intelligent detection of iot botnets using machine learning and deep learning," Applied Sciences, Vol 10(19), pp. 7709(1-23), 2020.
- [4] Singh, Manmeet; Singh, Maninder; Kaur, Sanmeet (2019), "10 Days DNS Network Traffic from April-May, 2016", Mendeley Data, V2, doi: 10.17632/zh3wnddzxy.2

정보보호 제품의 자가보호 기법 분석

이성원, 윤종희

영남대학교 컴퓨터공학과

Analysis of self-protection techniques of information security products

Sung-Won Lee, Jong-Hee Youn

Computer Engineering, Yeungnam University.

요약

본 논문에서는 정보보호 프로그램의 자가 보호 분석을 진행했다. 그 대표적인 백신 프로그램을 주 대상으로 진행하였으며, 백신 프로그램은 공격하는 악성 소프트웨어로부터 방어하기 위한 자가 보호 기능이 제공한다. 선행 작업과 정적 분석, 동적 분석을 통해 다양한 암티바이러스 제품에 대하여 분석을 진행하였고, 관련된 API 와 흐름에 대하여 파악할 수 있었다. 윈도우에서 권장하는 방식인 드라이버를 통한 보호는 커널 레벨에서 작동하기 때문에 기존의 유저레벨에서의 공격은 모두 방어해준다. 결국 자가 보호 기능은 드라이버를 통해 구현한 것이 높은 보안성을 가지므로 해당 방법을 통해 자가 보호를 구현하는 것을 권장한다.

I. 서론

사물 인터넷, 클라우드 컴퓨팅, 빅데이터, 모바일 기술 등의 발달과 확산으로 IT 산업 인프라로써의 정보보호의 중요성이 급증했다. 내부 보안 역량이 부족한 중소기업을 중심으로 랜섬웨어 등의 외부로부터의 공격을 피하기 위함 보안 솔루션 및 서비스에 대한 수요가 증가하고 있고, 개인 사용자들 또한 보안 의식의 성장과 개인 정보의 중요성, 다양한 개인정보보호 침해 사고 등의 이유로 정보보호 제품에 대한 수요가 증가하고 있다. 정보 가치의 중요성과 수요가 증가하고 정보가 다양하게 가공되어 처리되는 만큼 이를 보호하기 위한 정보 보호 제품들 또한 넓은 스펙트럼의 다양한 제품들이 등장하고 있다.

정보보호 제품의 대표적인 백신 프로그램의 경우 기존의 지정된 시간이나 작동을 통해 악성 소프트웨어를 탐지하거나 삭제하는 단순한 작업에서 에이전트 개념을 도입함으로써 보안

기능의 향상뿐만 아니라 복합적인 일들을 수행 가능하게 되었다. 정보보호 제품은 사용 목적에 따라 가장 대중적인 백신뿐만 아니라 보안 관리 시스템, 파일 배포 관리 시스템, 매체 보안 솔루션, 패치관리 시스템 등 다방면으로 활용하고 있다.

본 논문은 기본적으로 제공되고 있는 윈도우 시스템의 보안 기법에 대해 확인하고, 이와 다른 정보보호 제품의 자가 보호 기법에 대해 분석한다[1][2]. 본문에서는 정보보호 제품에 대한 조사와 적용 가능한 자가 보호 기법의 분석을 진행한다.

II. 자가보호 기법 분석

자가 보호 기능 대부분 파일, 프로세스, 레지스트리 등 프로그램 무결성이 보장되어야 하는 파일을 대상으로 별이고 있으며, 안랩의 V3에서는 랜섬웨어처럼 특수한 바이러스에 대해 보호하기 위해 볼륨 보호 기능도 제공하고 있다.

2.1 자가보호 기법 정적 분석

자가 보호 기법에 대하여 분석하고 검증하기 위한 첫 번째 방법은 정적 분석이다. 정적 분석 방법은 별도 대상 프로그램 실행 없이 진행할 수 있는 분석 방법으로서, 전체적인 구조나 기능에 대하여 파악하는데 용이하다.

NirSoft OpenSource DriverView 프로그램을 이용하여 드라이버 목록에 대하여 간편히 확인할 수 있다. 아래 그림 1의 경우 안티바이러스 제품인 Avast를 대상으로 드라이버 실행 목록을 확인한 그림이다. 간단하게 관련된 시스템 드라이버의 목록을 확인할 수 있다. 이러한 방법을 통해 자가 보호 방법을 분석하기 위한 대상을 특정지어 우선 분석 진행 할 수 있다.

aswMonFlt.sys	System Driver	Avast File System Minifilter for Wind...
aswRvrt.sys	System Driver	Avast Revert
aswSnx.sys	System Driver	Avast Virtualization Driver
aswSP.sys	System Driver	Avast self protection module
aswVmm.sys	System Driver	Avast VM Monitor

그림 1 Avast 제품 관련 드라이버

그림 1의 경우 확인할 수 있는 것처럼, 실제 제품의 드라이버에 대해 제조사 측에서 남겨놓은 설명이 존재할 수도 있다. 그림 14에서는 Avast 제품의 드라이버 중 aswSP 드라이버를 대상으로 ‘Self Protection module’ 이란 설명을 확인하였고, 이를 통해 해당 드라이버가 자가 보호와 관련된 드라이버라는 것을 추측할 수 있다.

일련의 과정을 통해 분석하고자 하는 시스템 드라이버 파일의 (.sys 파일) 우선순위를 정하였고, 본격적으로 정적 분석을 진행한다. 정적 분석의 여러 방법 중에서도 파일 디버깅을 통해 내부 구조를 파악 한다. 내부적으로 호출되고 실행되는 API 들은 다양하며, 이 중 자가 보호 기능과 연관성이 있는 API 들도 상당히 확인할 수 있다. 레지스트리에 관련된 API(ZwOpenKey, RtlQueryRegistryValues), 프로세스, 쓰레드 관련 API(PsSetCreateProcessNotifyRoutine, PsGetCurrentProcessId, PsGetCurrentThreadId,

ZwOpenProcess) 등 다양한 API 들이 호출되었으며, 해당 API 들의 의존성과 호출 순서 등을 통해 각 부분의 전반적인 흐름들을 파악 가능하다.

표 1 중요API 정리

API 이름	적용 대상
ObRegisterCallbacks	thread, process
CmRegisterCallback(Ex)	Registry
FItRegisterFilter	File, etc(info)

2.2 자가보호 기법 동적 분석

자가 보호 기법에 대하여 분석하고 검증하기 위한 두 번째 방법은 동적 분석이다. 동적 분석 방법은 파일을 실행하면서 분석을 진행하는 방법을 말한다. 즉 프로그램 동작을 모니터링하거나 디버깅 도구를 사용하여 상태를 실시간으로 확인하면서 분석하는 방법이다. 동적 분석 방법을 통해 정적 분석 방법에서 확인한 내용과 API 호출에 대하여 실제 안티바이러스 실행 흐름에서 호출 되는지 여부와 드라이버의 실제 흐름을 확인한다.

자가 보호 기법에 대한 분석을 위해 동적 분석을 진행하기 위해서 디버깅 방법의 하나로 Kernel Debugging을 사용한다. 윈도우 XP 64bit 이후 운영체제에 대해선 KPP 정책이 적용되어 커널 레벨에 대한 후킹이 불가능 하다. 후킹을 통해 API나 흐름을 디버깅 할 수 없기 때문에 가상 환경에 OS 와 분석하고자 하는 안티바이러스 제품 환경을 등록하고 Windbg 디버깅 툴을 이용하여 커널 디버깅을 시행한다. 커널 디버깅은 커널 레벨에서 동작하는 흐름을 검출하는 작업으로, 정적 분석에서 확인한 드라이버와 함수의 호출 흐름을 확인할 수 있다. 또한 필요에 따라 메모리 값을 확인 가능하다. 커널 디버깅 환경을 구축하는 방법은 여러 가지 있지만 본 연구에서는 VirtualKD라는 오픈 도구를 통하여 환경을 구축한다. 그림 2는 커널 디버깅 환경에 대해 도식화 한 그림이다.



그림 2 커널 디버깅을 위한 환경 구축의 도식화

2.2 자가보호 기법 동적 분석

자가 보호 관련 검증을 진행할 때 위에서 정리한 방법을 통해 상세 분석을 진행하는 방법 외에 간단하게 어떠한 대상을 보호하고 있는지 확인하는 방법 또한 존재한다. 그림 3의 경우 Avast 제품의 관련 프로세스를 종료 시도를 자가 보호 기능이 차단하여서 실패한 그림이다.

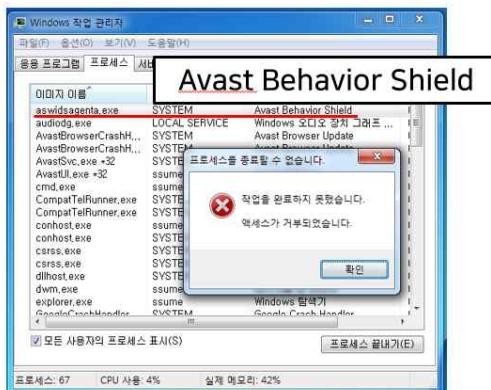


그림 3 프로세스 종료 요청 거부

해당 그림처럼 작업 관리자를 통해 간단히 확인해도 자가 보호 기능이 정상적으로 작동되고 있다면 해당 요청을 무시하게 된다. 이러한 간단한 방법을 통해서도 어떤 자가 보호가 걸려 있는지 확인할 수 있다.

이 외에도 CMD 의 Command를 통해서도 자가 보호 대상을 확인할 수 있다. Del, Move 명령 같은 간단한 명령을 통해 파일에 대한 조작을 진행하고 이에 대한 결과로 판단 할 수 있다. 레지스트리의 경우 ‘Reg delete’를 통해, 프로세스 쓰레드는 ‘Taskkill’, 서비스는 ‘Sc stop servicename’, 드라이버는 ‘fltmc unload driver_name’ 등을 통해 각각 파일, 레지스트리, 프로세스, 쓰레드, 서비스, 드라이버 등에 대한 자가 보호가 적용되어져 있는지 판단할 수 있

다.

이러한 간단한 방법을 제조사 및 프로그램에 대한 자가 보호 기능이 적용되어져 있는지 간단하게 파악이 가능하지만, 위에서 진행한 상세 분석처럼 자가 보호기능의 구현 방법에 대해서는 파악이 불가능하다. 확인한 자가 보호에 대하여 간단한 방법으로 자가 보호가 구현되었다면, 자가 보호가 적용된 대상이라 하더라도 보안적인 측면에서 안전하다고 할 수 없다.

III. 결론

안티바이러스 제품의 자가 보호 기법에 대하여 상세 분석을 진행하였으며, 분석을 위한 검증 방법을 제시하였다. 선행 작업과 정적 분석, 동적 분석을 통해 다양한 안티바이러스 제품에 대하여 분석을 진행하였고, 관련된 API 와 헤더에 대하여 파악할 수 있었다. 실제 안티바이러스 제품을 대상으로 확인 해 봤을 때 자가 보호 기능을 드라이버를 이용하여 구현하고 있는 것을 확인할 수 있었다. 드라이버를 이용한 자가 보호는 커널 레벨에서 작동하기 때문에 기존의 유저 레벨에서의 공격은 모두 방어해준다. 또한 커널 레벨에서도 높은 우선순위로 적용되어 높은 보안성을 가지기 때문에 모든 악의적인 접근을 차단한다. 결국 자가 보호 기능은 드라이버나 미니필터를 통해 구현한 것이 높은 보안성을 가지므로 해당 방법을 통해 자가 보호를 구현하는 것을 권장 한다.

[참고문헌]

- [1] F. H. Hsu, M. H. WU, C. K. Tso, C. H. Hsu and C. W. Chen, Antivirus software shield against antivirus terminators, IEEE Transactions on Information Forensics and Security, 2012
- [2] D. Weston and M. Miller, Windows10-Mitigation-Improvements, BLACKHAT, August, 2016

코드 재사용 공격 방어를 위한 제어 흐름 무결성 검증 기법 연구

여기수*, 권동현†

*부산대학교 전기컴퓨터공학부 정보컴퓨터공학 전공(학부생)

A Survey of Control Flow Integrity Techniques
against Code Reuse Attack

Gisu Yeo*, Dong-Hyun Kwon†

*Pusan National University(Undergraduate student)

요약

제어 흐름 무결성 검증 기법은 코드 재사용 공격에 대한 효과적인 방어 수단 중 하나이다. 소스 코드 레벨과 바이너리 레벨, 모두에서 구현될 수 있고, ASLR과 같은 메모리 보호 기법과는 다르게 하드웨어적 지원이 최소화된 임베디드 시스템 환경에서도 구현할 수 있다는 장점이 있다. 본 연구에서는 제어 흐름 무결성 기법의 구현에 대해 살펴보고, 실제 임베디드 시스템 환경에서 제어 흐름 무결성 검증 기법이 적용된 연구 사례에 대해 분석하였다.

I. 서론

코드 삽입 공격에 대한 방어 수단으로 $W \oplus X$ 기법이 소개된 이후 이를 우회하는 방법으로 코드 재사용 공격이 등장했다. 실행 가능한 바이너리 영역 내의 일련의 코드 조각들을 조합하여 공격을 진행하는 코드 재사용 공격은 이에 대한 여러 방어 기법이 등장했음에도 불구하고 다양한 방향으로 계속 발전하여 여전히 소프트웨어의 실행에 큰 위협이 되고 있다. 코드 재사용 공격은 원하는 코드 조각을 실행하기 위해 제어 흐름을 변조한다는 특성을 가지는데, 이러한 특성에 비추어 볼 때 제어 흐름의 무결성을 검증하고, 정해진 제어 흐름 그래프 (Control Flow Graph)를 따라서만 함수의 호출과 복귀를 할 수 있게 하는 방법은 코드 재사용 공격에 대한 효과적인 방어 수단이 된다.

본 논문에서는 이전의 논문에서 소개된 제어 흐름 무결성 검증 기법에 대해 살펴보고, 일반

적인 실행 환경이 아닌 임베디드 환경에서의 제어 흐름 무결성 기법이 적용된 사례에 대해서 살펴본다.

II. 코드 재사용 공격

코드 재사용 공격(Code Reuse Attack, CRAs)은 바이너리 내에 존재하는 일련의 명령어 셋을 연속적으로 조합하여 원래의 동작이 아닌 공격자가 원하는 임의의 코드를 실행할 수 있도록 하는 공격기법이다. 라이브러리 및 응용 프로그램 내에 존재하는 정상적인 명령어들을 활용하여 공격이 진행되므로 공격으로 인한 임의의 코드 실행이 일반적인 코드 실행과 크게 다르지 않으며, 이러한 이유로 코드 삽입 공격과는 다르게 $W \oplus X$ 기법이 적용되어 있어도 공격할 수 있다. 대표적인 코드 재사용 공격으로 ROP (Return Oriented Programming)[1] 기법이 있는데, pop-pop-ret 과 같은,

'Gadgets'이라고 불리는 작은 명령어 집합들을 연속적으로 실행되게 함으로써 Return을 통한 함수 호출이 이루어지게 한다. ARM 등 임베디드 환경에서 주로 사용되는 코드 재사용 공격 기법의 하나인 JOP (Jump Oriented Programming) [2]는 ROP와 다르게 Return으로 구성된 가젯을 이용하는 대신 간접 분기 명령을 가젯으로 이용하여 공격자가 원하는 동작을 하도록 만든다.

III. 제어 흐름 무결성 검증 기법

정상적으로 동작하는 소프트웨어의 실행은 항상 사전에 정해져 있는 제어 흐름 그래프를 따르게 되어 있다. 하지만 공격자가 코드 재사용 공격을 하게 되는 경우 사전에 정해진 제어 흐름을 벗어나 공격자가 원하는 임의의 코드를 따라가게 된다. 이러한 점에서, 제어 흐름의 무결성을 검증하여 정상적인 제어 흐름 밖으로 벗어나지 못하게 만드는 것은 많은 공격을 막을 수 있는 효과적인 방안이 된다 [3].

제어 흐름의 무결성을 검증하는 방법은 무결성을 검증하고자 하는 간선의 방향에 따라 크게 순방향 간선(Forward Edge)에 대한 무결성 검증과 역방향 간선(Backward Edge)에 대한 무결성 검증으로 나누어진다. 순방향 간선에 대한 제어 흐름의 변조는 주로 레지스터의 값에 따라 제어 흐름이 변할 수 있는 간접 분기 혹은 간접 호출 명령에서 일어난다. 따라서 순방향 간선에 대한 무결성 검증은 간접 분기 명령과 간접 호출 명령이 실행될 시 정상적인 제어 흐름 그래프를 그리는 방향으로 진행되는지 검

사하는 방법으로 진행된다. 이는 간접 분기와 간접 호출 명령을 이용하여 제어 흐름을 변화시키는 공격기법인 JOP 공격을 방어하는 데 효과적이다.

역방향 간선에 대한 제어 흐름 변조는 함수 호출로 인해 한 블록에서 다른 기본 블록으로 흐름이 변화되었다가 다시 원래의 블록으로 돌아가는 과정에서 일어난다. 하나의 기본 블록 내에서 다른 함수를 호출하는 경우 스택을 이용하여 복귀주소를 저장해두는데, Stack Buffer Overflow 등의 취약점을 이용해 해당 주소를 변조하여 제어 흐름을 바꾸는 방식으로 공격이 진행된다. 역방향 간선에 대한 무결성 검증은 메모리에 저장된 복귀주소가 변조되지 않았음을 검증하여, 호출된 함수의 동작이 끝난 후 올바른 위치로 되돌아갈 수 있도록 한다.

3.1 분기 통제

분기 통제(Branch Regulation) 기술은 제어 흐름에 변조를 유발할 수 있는 명령들에 대한 일련의 규칙을 만들어 해당 규칙들 내에서만 각 명령이 동작할 수 있도록 강제한다 [4]. 분기 통제의 대상이 되는 명령에는 Return과 간접 분기, 그리고 간접 호출 명령이다. 간접 분기 명령이 사용되는 경우는 주로 함수 내부에서 switch-case 등의 구문을 처리하기 위해 존재하기에, 간접 분기 명령이 존재하는 함수의 범위를 넘어가지 않도록 한다. 간접 호출 명령은 해당 명령의 목적지가 반드시 함수의 시작 주소를 가리키고 있는 경우에만 정상적인 동작으

Type	Example Instructions		Branch Regulation
	x86_64	ARM	
간접 호출	call rax	blx r0	<ul style="list-style-type: none"> Target = <Entry of Functions>
간접 분기	jmp rax	ldr pc, [r0, r1 #4]	<ul style="list-style-type: none"> Function Base < Target < Bound IF Target > Bound THEN Target = <Entry of Functions>
복귀	ret	pop {pc}	<ul style="list-style-type: none"> Target = <Restored Address>

[표 1] 분기 통제로 강제되는 규칙 및 영향을 받는 명령어 타입

로 간주한다. Return 명령은 대응되는 함수 호출에서 미리 복귀주소를 저장해두었다가 복귀 명령이 호출되기 직전 시점에서 스택에 있는 값과 별도로 저장된 값을 비교하여 무결성을 검증한다.

3.2 함수 서명 값 검증

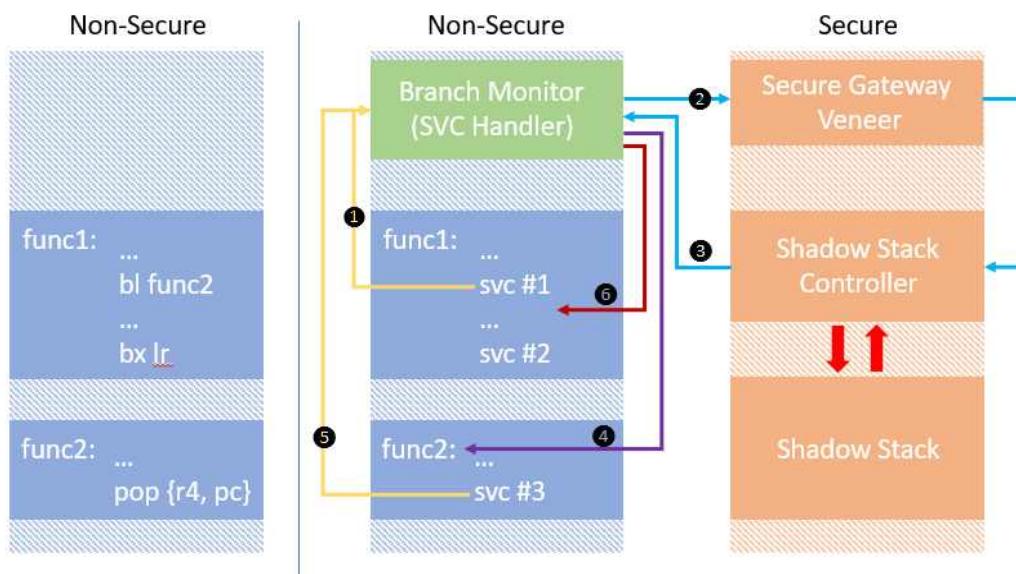
V Van Der Veen 외 9인은 그들의 논문[5]에서 소스 코드 없이 바이너리 레벨에서 구현 가능한 새로운 제어 흐름 무결성 검증 기법을 소개하였다. 일반적인 소스 코드 레벨의 제어 흐름 무결성 검사와 다르게, 바이너리 레벨에서는 함수 호출 시 전달되는 인자의 유형과 개수를 정확히 알 수 없다는 단점이 있다. 이를 극복하기 위하여 해당 논문에서는 TypeArmor라는 기능을 구현하였다. 정적 분석을 통해 각 함수에서 사용하는 최소 인자 수를 파악하고, 각 간접 호출이 발생하는 시점에서 인자로 전달될 수 있는 레지스터 수의 최댓값을 구한다. 호출하고자 하는 함수에서 필요로 하는 최소 인자 수가 호출 시 전달되는 인자 수의 최댓값보다 큰지 비교하는 방식으로 간접 분기가 가능한 대상의 개수를 줄인다. 정적 분석으로 함수에서 사용되는 인자 수를 파악하기 위해, 각 레지스터를 *read-before-write* (R), *write-before-read*

(W), *clear/unouched* (C) flag로 구분하고, 이를 종합하여 각 함수에 4byte 라벨을 붙여 실행 시점에서 검사하도록 한다.

IV. 임베디드 환경에서의 제어 흐름 무결성 검증

임베디드 환경에서 사용되는 마이크로컨트롤러(MCUs)는 일반 컴퓨터에 들어가는 프로세서에 비해 저조한 수준의 연산 능력을 갖추고 있으며, 인터럽트에 의해 동작이 시행되는, 일반적인 환경과 구분되는 독특한 특성이 있다. 이러한 특성은 전통적인 방식의 제어 흐름 무결성 검증을 적용하기 어렵게 만든다 [6]. 하지만, 다행히도 일부 아키텍처에서는 신뢰 실행 환경(Trusted Execution Environment, TEE)을 제공하기 위해 하드웨어적으로 메모리 공간을 분리하는 등 최신의 보안 기술들이 제공되고 있으며, 최근에는 이러한 기능들을 이용하여 임베디드 환경에서 제어 흐름 무결성 검증을 하고자 하는 연구가 이루어지고 있다.

Nyman 외 3인이 구현한 CFI CaRE[6]는 ARM TrustZone-M 기술을 제어 흐름 무결성 검증에 활용하였다. 인터럽트에 기반하여 동작이 시행되는 임베디드 환경의 특성에 맞게, 기존의 제어 흐름 무결성 검증 기법에 더해 비동



[그림 1] CFI CaRE Control Flow

기 인터럽트로 인한 제어 흐름 변경에 대한 무결성도 검증하고 있다. 해당 논문에서는 ARM TrustZone-M의 Secure Memory 내에 Shadow Stack과 Branch Monitor를 구현하고, Non-secure 영역 내의 안전하지 않은 명령을 Supervisor Call (SVC) 명령과 원본 명령에 대응되는 Comment로 패치하여 메모리 레이아웃의 변화를 최소화한다.

TZmCFI[7]는 CFI CaRE와 마찬가지로 Shadow Stack을 이용한 역방향 간선에 대한 제어 흐름 검증을 시행하고, Exception Return에 대한 무결성 검증을 수행한다. CFI CaRE에서 다루지 않았던 Nested exception에 관한 처리를 Shadow Exception Stack이라는 구조를 만들어 처리하고 있으며, Multi-tasking을 지원하기 위해 Thread 별로 각각의 Shadow Stack을 구현하였다.

V. 결론

제어 흐름 무결성 검증 기술은 코드 재사용 공격에 대한 효과적인 방어 수단이다. 이는 Address Space Layout Randomization (ASLR)과는 다르게 Memory Management Unit (MMU)와 가상 메모리를 지원하지 않는 임베디드 환경에서도 구현 가능하며, 바이너리와 심볼 정도밖에 없는 제한적인 상황에서도 구현할 수 있다. 하지만 바이너리 내의 모든 간접 분기와 함수 호출, 그리고 반환 명령마다 무결성을 검증하는 루틴이 들어가게 되므로 실행 시간 측면에서 큰 오버헤드를 발생시킨다. 따라서 실시간성이 중요한 환경에서는 제어 흐름 검증 기술을 실제로 적용하기 힘들 수 있다. 실제 임베디드 환경에서 제어 흐름 무결성 검증 기술이 실용적인 코드 재사용 공격에 대한 방어 수단이 되기 위해서는 보다 경량화되고, 오버헤드가 크지 않게 구현하는 기술이 연구되어야 할 것이다.

Acknowledgement

이 성과는 정부(과학기술정보통신부)의 지원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1G1A1102193)

[참고문헌]

- [1] Hovav Shacham. 2007. The geometry of innocent flesh on the bone: return-into-libc without function calls (on the x86). In Proceedings of the 14th ACM conference on Computer and communications security (CCS '07). Association for Computing Machinery, New York, NY, USA, 552 - 561. DOI:10.1145/1315245.1315313
- [2] Stephen Checkoway, Lucas Davi, Alexandra Dmitrienko, Ahmad-Reza Sadeghi, Hovav Shacham, and Marcel Winandy. 2010. Return-oriented programming without returns. In Proceedings of the 17th ACM conference on Computer and communications security (CCS '10). Association for Computing Machinery, New York, NY, USA, 559 - 572. DOI:10.1145/1866307.1866370
- [3] Martín Abadi, Mihai Budiu, Úlfar Erlingsson, and Jay Ligatti. 2009. Control-flow integrity principles, implementations, and applications. ACM Trans. Inf. Syst. Secur. 13, 1, Article 4 (October 2009), 40 pages. DOI:10.1145/1609956.1609960
- [4] M. Kayaalp, M. Ozsoy, N. Abu-Ghazaleh and D. Ponomarev, "Branch regulation: Low-overhead protection from code reuse attacks," 2012 39th Annual International

Symposium on Computer Architecture (ISCA), 2012, pp. 94–105, DOI: 10.1109/ISCA.2012.6237009.

- [5] V. van der Veen et al., "A Tough Call: Mitigating Advanced Code-Reuse Attacks at the Binary Level," 2016 IEEE Symposium on Security and Privacy (SP), 2016, pp. 934–953, doi: 10.1109/SP.2016.60.
- [6] Nyman T., Ekberg JE., Davi L., Asokan N. (2017) CFI CaRE: Hardware-Supported Call and Return Enforcement for Commercial Microcontrollers. In: Dacier M., Bailey M., Polychronakis M., Antonakakis M. (eds) Research in Attacks, Intrusions, and Defenses. RAID 2017. Lecture Notes in Computer Science, vol 10453. Springer, Cham. DOI:10.1007/978-3-319-66332-6_12
- [7] Kawada, T., Honda, S., Matsubara, Y. et al. TZmCFI: RTOS-Aware Control-Flow Integrity Using TrustZone for Armv8-M. Int J Parallel Prog 49, 216 - 236 (2021). DOI:10.1007/s10766-020-00673-z

접근제어를 지원하는 Tagged Memory Extension 동향

이진재*, 데리 프라타마*, 권동현**, 김호원**

*부산대학교 (대학원생)

**부산대학교 (교수)

Trend of Tagged Memory Extension supporting access control

Jin-Jae Lee*, Derry Pratama*, Dong-Hyun Kwon**, Ho-Won Kim**

*Pusan National University(Graduate student)

**Pusan National University(Professor)

요약

상용 프로세서 제조사는 메모리 취약성을 효율적으로 보완하기 위한 다양한 보안 구조를 제공하고 있다. 그 중에서도 ARM의 MTE, SPARC의 ADI와 같은 TME(Tagged Memory Extension)는 Tagged Memory를 활용하여 메모리 접근을 제어하는 확장 명령어 세트이다. 본 연구에서는 이와 같은 Tagged Memory 구조를 활용하는 명령어 Extension 개발 현황 및 최근 등장한 오픈소스 ISA(Instruction Set Architecture)인 RISC-V ISA를 활용한 Tagged Memory Extension에 대해서 살펴보고 이러한 구조들의 한계점을 알아보도록 한다.

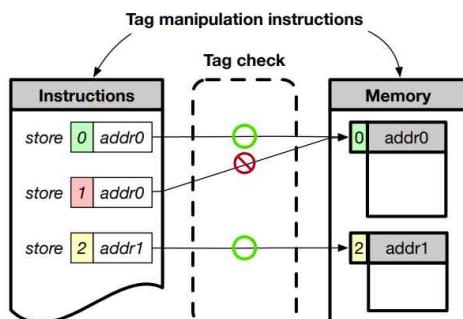
I. 서론

TMA(Tagged Memory Architecture)는 모든 메모리 블록이 현재 메모리의 상태를 나타내는 특수 목적의 메모리 태그 값을 가지는 컴퓨터 아키텍처이다. TMA는 메모리 태그 값이 어떻게 사용되는지에 따라 다양한 보안 솔루션 [1]~[4]을 지원할 수 있다고 여겨지고 있다. 예를 들면, 개발자들은 use-after-free 취약점을 탐지하기 위해 TMA를 사용할 수 있다. 구체적으로는, 데이터가 할당된 메모리 블록과 데이터 할당이 해제된 메모리 블록이 서로 다른 메모리 태그 값을 가지도록 강제할 수 있다면 할당이 해제된 메모리 블록에 대한 접근을 해당 메모리의 태그 값을 검사하는 것으로 검증할 수 있다. 이러한 TMA의 범용성으로 인해, 상용 프로세서 제조사들은 TMA를 위한 확장 명령어 세트을 발표하였다(ARM의 MTE(Memory Tagging Extension), SPARC의 ADI(Application Data Integrity) 등, 편의를 위해 이후로는 이러

한 상업용 확장 명령어 세트들을 TME(Tagged Memory Extension)라고 명칭하겠다). 그리고 이에 따라 소프트웨어 어플리케이션의 보안 수준을 강화하기 위해 TME를 사용한 몇 가지 시도가[7] 있었다.

II. 본론

2.1 TME(Tagged Memory Extension)



[그림 1] TME(Tagged Memory Extension) 동작 구조

TME의 세부적인 구현은 프로세서의 아키텍처에 따라 다르게 나타나지만, 기본적인 동작은 다음과 같다. TME에서는 물리적인 메모리뿐만 아니라 메모리에 접근하기 위한 포인터에도 태그 값이 사용된다. 본 논문에서는 물리적 메모리상에 저장되는 태그를 memory tag, 포인터에 저장되는 태그 값을 address tag라고 부르도록 한다. memory tag는 각각의 물리적 메모리 블록에 할당이 되며 실제 memory tag값은 메모리상의 별도의 공간에 저장이 된다. address tag의 경우 메모리상의 별도의 공간에 저장되지 않고 메모리에 접근하는데 사용되는 포인터 값의 상위 비트에 저장이 되어 사용된다. 이는 64비트 프로세서 구조의 address translation 과정에서 64비트 주소 값의 최상위에 위치하는 일부 비트가 사용되지 않는다는 점을 활용한 것이다. [그림1]과 같이 TME는 포인터상의 address tag와 물리 메모리상의 memory tag의 값이 일치할 때에만 메모리에 대한 접근을 허가한다. 이러한 tag 비교 연산은 특별한 명령어의 추가 없이 기본적인 메모리 접근 연산이 수행될 때 함께 수행된다.

2.2 ARM MTE

ARM MTE는 ARMv8.5-A 아키텍처에서 소개된 명령어 확장이다. ARM MTE에서는 lock과 key라는 개념을 사용하여 물리 메모리에 대한 세밀한 접근 제어 메커니즘을 제공한다. 다르게 말해서, 포인터 변수의 key 값이 목표 물리 메모리 영역의 lock 값과 일치할 때만 메모리 접근이 허가되고 key 값과 lock 값이 일치하지 않을 경우에는 접근 에러가 발생한다. ARM MTE에서는 포인터 변수상의 key 값을 pointer tag라 부르고, 물리 메모리상의 lock 값을 memory tag라고 부른다. memory tag는 4비트의 길이를 가지며 물리 메모리상에 정렬된 각 16바이트 영역에 대해 할당되며, memory tag는 메모리상에 사용자가 주소를 통해 접근할 수 없는 영역에 저장된다. 이러한 특징 때문에 일반적인 메모리 접근 명령어로는 memory tag에 접근할 수가 없는 대신에 ARM MTE는 memory tag에 접근하고 업데이트할 수 있는

특수 명령어를 제공한다(e.g., LDG, STG). pointer tag는 ARMv8-A 64비트 아키텍처의 top byte ignore(TBI) 기능을 사용하여 구현된다. TBI가 enable되면, 포인터상의 최상위 1바이트는 address translation 과정에서 메모리 주소로 해석되지 않으며 ARM MTE에서는 최상위 1바이트에서 4비트가 pointer tag를 위해 사용된다(64비트 포인터에서 [59:56]비트).

ARM MTE는 pointer tag에 랜덤 tag 값을 할당하기 위한 별도의 명령어를(i.e., IRG) 제공하지만, pointer tag가 포인터의 상위 비트에 존재하기 때문에, 산술 연산 명령어나 메모리 load 명령어를 통해 pointer tag를 조작하는 것도 가능하다.

2.3 SPARC ADI

ARM MTE와 유사하게, SPARC ADI도 tagged memory 아키텍처를 통해 lock과 key 메커니즘을 구현한다. 세부적으로는, SPARC ADI는 물리 메모리상에 저장되는 tag와 포인터상의 tag 모두를 version tag라고 부르며, version tag는 ARM MTE와 동일하게 4비트를 사용하지만 포인터상에서는 최상위 1바이트의 [63:60] 위치에 저장된다.

SPARC ADI는 물리 메모리상의 각 64바이트마다 version tag를 할당하며 메모리 영역에 접근하기 위해서는 포인터를 통해 동일한 version tag가 제시되어야만 한다.

구체적으로는, SPARC ADI에서도 address tag가 포인터의 최상위 바이트에 위치하기 때문에 ARM MTE와 유사하게 산술 연산 명령어나 메모리 load 명령어를 통해 address tag를 조작하는 것이 가능하다.

2.4 RISC-V MTE

공개되어 있는 오픈소스 ISA에서도 RISC-V ISA는 산업과 학계에서 대중적이고 컴퓨터 아키텍처 연구의 프로토타입으로 가장 많이 사용되는 ISA 표준이다. 활발한 개발자 커뮤니티의 지원으로 현재는 Linux 시스템을 동작시킬 수 있을 정도의 RISC-V ISA를 사용

한 프로세서도 개발되어 있을 정도로 많은 발전이 있었으며 현재도 연구 및 개발이 활발히 이루어지고 있다.

최근에는 이러한 RISC-V ISA를 활용한 오픈소스 tagged memory architecture도 공개되고 있다. RISC-V memory tagging extension (RISC-V MTE)[11]는 오픈소스 RISC-V 프로세서상에 tagged memory 아키텍처를 구현하기 위한 프로젝트로 물리 메모리에 할당되는 tag와 포인터로 제공되는 tag에 8비트 길이의 tag를 사용한다. 포인터상의 tag는 64비트 주소에서 최상위 1바이트 전체를 사용하며, ARM MTE와 동일하게 물리 메모리상의 각 16바이트 영역에 tag가 할당된다. 또한 물리 메모리상에 저장되는 tag 값에 접근하기 위한 별도의 명령어가 제공된다(i.e., ST(Store Tag), LT(Load Tag)).

2.5 한계점

TME를 보안 솔루션에 적용할 때에는 몇 가지 한계점도 존재한다. 첫 번째로, address tag에 대한 공격으로 TME를 사용한 보안 기법은 무력화될 수 있다. memory tag에 대한 접근은 몇 개의 특수한 명령어에 의해서만 가능하기 때문에 이미 존재하는 기법들을[8][9] 통해 이러한 명령어들에 대한 접근만 차단하면 memory tag에 대한 기밀성은 보장이 될 수 있다. 하지만, address tag는 포인터 값의 상위 비트에 존재하기 때문에 address tag는 소프트웨어 취약점(uninitialized read, integer overflow)을 통한 일반적인 메모리 접근 명령어나 산술연산 명령어를 통해 변경될 수 있다. 두 번째로, memory tag는 물리 메모리의 각 블록마다 할당이 되어 있기 때문에 멀티 코어 시스템상에서 각 코어 별로 다른 접근 권한을 부여하는 것이 불가능하다. 마지막으로, 개발자들은 TME를 이용하여 세부적인 접근 권한을 설정하는 것이 불가능하다. 예를 들어, ARM MTE에서는 address tag와 memory tag가 같은 값만 가지고 있다면 모든 종류의 메모리 접근이 허가되기 때문에 개발자가 접근하고자 하는 메모리 영역에 대해 read-only 접근 권한을 부여하는 것이 불가능하

다.

따라서, TME가 완벽한 접근제어 기능을 제공한다고는 단언할 수 없으므로 소프트웨어를 이용한 보완이나 한계점을 보완한 새로운 TME에 대한 연구가 필요하다.

III. 결론

본 논문에서는 상용 프로세서의 tagged memory extension 및 오픈소스 ISA를 활용한 tagged memory extension의 동향과 이러한 tagged memory extension이 가지는 장점과 한계점에 대해 간략히 소개하였다.

ACKNOWLEDGMENT

이 논문은 국토교통부의 스마트시티 혁신인재 육성사업으로 지원되었습니다.

【참고문헌】

- [1] N. Zeldovich, H. Kannan, M. Dalton, and C. Kozyrakis, “Hardware enforcement of application security policies using tagged memory.” in OSDI, vol. 8, 2008, pp. 225 - 240.
- [2] C. Song, H. Moon, M. Alam, I. Yun, B. Lee, T. Kim, W. Lee, and Y. Paek, “Hdfi: Hardware-assisted data-flow isolation,” in 2016 IEEE Symposium on Security and Privacy (SP). IEEE, 2016, pp. 1 - 17.
- [3] J. Woodruff, R. N. Watson, D. Chisnall, S. W. Moore, J. Anderson, B. Davis, B. Laurie, P. G. Neumann, R. Norton, and M. Roe, “The cheri capability model: Revisiting risc in an age of risk,” in 2014 ACM/IEEE 41st International Symposium on Computer Architecture (ISCA). IEEE, 2014, pp. 457 - 468.
- [4] S. Weiser, M. Werner, F. Brasseler, M. Malenko, S. Mangard, and A.-R. Sadeghi, “Timber-v: Tag-isolated memory bringing

- fine-grained enclaves to risc-v.” in NDSS, 2019.
- [5] D. Seal, ARM architecture reference manual. Pearson Education, 2001.
 - [6] K. Aingaran, S. Jairath, G. Konstadinidis, S. Leung, P. Loewenstein, C. McAllister, S. Phillips, Z. Radovic, R. Sivaramakrishnan, D. Smentek, and T. Wicki, “M7: Oracle’s next-generation sparc processor,” IEEE Micro, vol. 35, no. 2, pp. 36 - 45, 2015.
 - [7] K. Serebryany, “Arm memory tagging extension and how it improves c/c++ memory safety,” login: the USENIX Magazine, vol. 44, p. 5, 2019.
 - [8] S. Park, S. Lee, W. Xu, H. Moon, and T. Kim, “libmpk: Software abstraction for intel memory protection keys (intel {MPK}),” in 2019 {USENIX} Annual Technical Conference ({USENIX}{ATC} 19), 2019, pp. 241 - 254.
 - [9] A. Vahldiek-Oberwagner, E. Elnikety, N. O. Duarte, M. Sammler, P. Druschel, and D. Garg, “{ERIM}: Secure, efficient in-process isolation with protection keys ({MPK}),” in 28th {USENIX} Security Symposium ({USENIX} Security 19), 2019, pp. 1221 - 1238.

제로 트러스트 관점의 보안체계

고민혁*, 이대성*

*부산가톨릭대학교

Security Systems from a Zero Trust Perspective

Min-Hyuck Ko*, Daesung Lee*

*Catholic University of Pusan

요약

최근 네트워크 확장 및 클라우드 인프라 확장, 자택근무로 인한 원격 접속의 증가로 외부의 접근뿐만이 아니라 내부에서의 접근을 경계해야 할 필요성이 증가하고 있다. 이로 인해 제로 트러스트라는 새로운 네트워크 보안 모델이 주목받고 있다. 이 논문에서는 제로 트러스트의 개념을 간단히 소개하고 이를 이용한 새로운 보안 전략과 이 모델을 기반으로 하는 보안체계의 구축을 소개하려 한다.

I. 서론

제로 트러스트란 Forrester 리서치 부사장 겸 수석 애널리스트인 John Kindervag 이 처음 제시한 개념으로 엄격한 ID 확인 프로세스 기반 네트워크 보안 모델이다. 제로 트러스트는 신뢰할 수 있는 네트워크는 존재하지 않는다는 핵심원칙을 가지고 있으며 모든 네트워크 트랜잭션이 이루어지려면 먼저 인증을 받아야 하며 인증되고 권한이 부여된 사용자와 장치만이 애플리케이션 및 데이터에 접속을 허용한다[1].

II. 결론

2.1 제로 트러스트 적용 전략

클라우드 인프라가 확대되면서 애플리케이션의 위치 상관없이 내부 및 외부 관계자들은 언제든지 접속해 필요한 작업을 진행할 수 있다. 또한, 최근 코로나로 인해 재택근무 등의 원격 접속이 새로운 근무 형태로 자리 잡고 있다. 제

로 트러스트는 이러한 변화에 반영할 수 있는 솔루션으로 부상했으며 여러 기업이 현재 고려하고 있는 보안 모델이다. 제로 트러스트 모델은 모든 작업이 인터넷에 연결되어 있다는 전제하에 어떤 사용자와 장비도 신뢰하지 않고 기업의 모든 IT 자원은 접속마다 인증을 거쳐야 하며, 모든 활동과 트래픽은 모니터링의 대상이 된다. 이중 인증으로 신뢰할 수 있는 계정 정보인지 확인하고, 안전한 보안 네트워크를 위해서 작업에 직접 접속하는 게 아닌 보안 게이트웨이를 통해 접속할 수 있게 한다[2, 3].

EAA(Enterprise Application Access)는 제로 트러스트 모델을 구체화하여 원격근무에 최적화시킨 솔루션이다. 내부로의 기본 접근을 모두 막고, 권한을 부여한 장비나 사용자만 접속을 허용한다는 제로 트러스트의 기본 원칙을 이용하고 있으며, 추가로 인바운드 방화벽 포트를 열지 않고 데이터센터 안에 엔터프라이즈 커넥터라는 일종의 가상머신을 두는 방법을 제

시한다. 여기서 엔터프라이즈 커넥터는 사용자 확인과 접속만을 관리 및 허용하는 역할을 담당한다. 일단 연결만 되면 모든 내부 네트워크에 접속할 수 있는 기존 VPN과 달리 EAA는 외부 요청을 엔터프라이즈 커넥터로만 연결하며, 사전에 정의된 사용자나 장비가 접속할 때 권한에 맞는 작업 및 애플리케이션만 열어준다. 또한, 내부 네트워크가 외부 요청과 직접 연결되는 것이 아닌, 엔터프라이즈 커넥터라는 새로운 매개체에 연결되기 때문에 방화벽을 열지 않아도 된다. 별도의 전용 장비를 요구하거나 보안정책을 추가로 설정할 필요가 없다. 접근을 관리해 모든 내부 네트워크를 탐색할 수 없으며, 허용된 애플리케이션에만 접속할 수 있게 하여 보안 위협을 최소화한다[2].

2.2 제로 트러스트 기반 보안체계 구축 프로세스

일하는 장소와 장비의 다양화, 클라우드 시스템의 활용, 보안 위협의 분산화 등이 보안 대책의 과제로 제시되고 있다. 보안 강화를 위해선 통신 중에 생겨나는 로그를 확인하는 방법이 있지만, 최근에는 단말 로그, 네트워크의 로그, 인증 로그, IaaS (Infrastructure as a Service)나 SaaS (Software as a Service) 로그 등 확인해야만 하는 로그가 광범위해지고 있고 또한 사내 네트워크나 외부의 단말, 클라우드 시스템 등 로그를 확인해야만 하는 범위도 넓어지고 있다[3]. 제로 트러스트는 신뢰할 수 있는 네트워크는 존재하지 않는다는 핵심원칙에 의해 모든 행동이나 사용자, 장치, 데이터의 속성을 로그로 확인할 수 있어야 한다. 하지만 앞서 말했듯이 로그는 광범위하게 걸쳐있기 때문에 얼마나 효율적으로 보안 및 감시하는가를 중요하게 여겨야 한다. 효율적으로 감시를 하기 위해서는 사용자나 객체의 행동을 감시하는 것이 중요하다. 광범위한 로그로부터 모든 행동을 세밀하게 확인하는 것은 현실적이지 못하기 때문에 조직 및 기업에서 중요한 리소스가 무엇인지 조사하여 어떠한 사용자와 객체로부터 어떠한 행동이나 접근을 탐지 및 방어해야 하는지에 대한 보안정책을 정의하는 것이 바람직하다[4]. 이렇게

행동 기반에서 탐지를 시행할 경우 네트워크 로그 등을 모아 해석해야 하며 행동 기반을 제외한 다른 로그들도 조사하려면 번잡해질 가능성이 있다. 이러한 문제점의 솔루션 중 하나로 SIEM(Security Information & Event Management) 가 있다. SIEM 은 크게 4가지 기능으로 데이터 통합, 탐지, 조사, 대응으로 나타낼 수 있다. 이러한 기능을 이용하여 각종 로그를 집적하여 로그를 해독하는데 특화돼 있으므로 제로 트러스트를 기반으로 한 보안 운용 및 감시를 보다 효율적으로 실시할 수 있게 해준다. 광범위한 로그 문제 말고도 해결해야 하는 보안 과제로 인력이 부족하다는 문제가 있다. 이러한 문제는 제로 트러스트 보안 시스템 운용을 어떻게 자동화해야 하는가로 볼 수 있다. 현실적으로 모든 상황을 자동화하여 검사할 수 없기에 초기 대응만을 자동화하여 시나리오 및 가상 테스트를 통해 접근 및 순서를 파악하고 다양한 시나리오를 축적하여 자동화 절차를 명확하게 만드는 것을 목표로 해야 한다.

III. 결론

본 논문에서는 제로 트러스트라는 다른 관점의 네트워크 보안 모델에 대한 개념과 이를 활용한 전략 및 보안체계에 대해서 알아보았다. 현재 대부분 기업은 웹 서비스, 클라우드 시스템, 데이터센터용으로 방화벽이나 WAF(Web Application Firewall) 등의 보안 솔루션을 갖추고 있다. 하지만 현 상황으로 인해 원격 접속의 필요가 높아지면서 원격 접속 제어의 보안 취약점을 극복해야 하는 과제가 중요시되고 있다. 제로 트러스트는 모든 접근을 의심하는 핵심원칙하에서 엄격한 보안이 요구되기 때문에 앞으로의 네트워크 보안 측면에서 큰 역할을 담당할 것으로 생각한다.

[참고문헌]

- [1] 제로 트러스트란? [Internet]. Available : <https://www.vmware.com/kr/topics/glossar>

y/content/zero-trust.html

- [2] 천지용, 제로 트러스트 기반의 네트워크 보안 전략, IDG Summary AKAMAI MEGAZONE
- [3] Seo-Young Kim, Kyung-Hwa Jeong, Yuna Hwang, Dae-Hun Nyang, Abnormal Behavior Detection for Zero Trust Security Model Using Deep Learning, 한국정보처리학회 학술대회논문집, 28(1): 132-135
- [4] Zero Trust 도입가속-마이크로 소프트의 보고서 [Internet]. Available : <https://blog.naver.com/cspark14/222453433114>
- [5] 황민주, Microsoft Zero Trust Network 전략 및 구현방안, Cyber Security Solutions Group

딥러닝 기반 랜섬웨어 분류를 위한 시각화 기법 연구

이수경* 최은정**

*서울여자대학교 (대학원생) **서울여자대학교(교수)

A study of visualization techniques for Deep Learning-based Ransomware Family Classification

Sy-Gyeong Lee*, Eun-jung Choi**

*Seoul Women's University(Graduate student)

**Seoul Women's University(Professional)

요약

일상생활이 편리해지는 만큼 그에 따른 보안 문제 또한 증가하고 있다. McAfee Labs Threate Report에 따르면 랜섬웨어의 수는 빠르게 증가하고 공격의 유형이 다양화되고 지능화되고 있다. 특히 소규모 랜섬웨어는 감소하고 있는 반면 서비스형 랜섬웨어는 대규모 조직과 회사를 표적으로 공격대상이 증가하고 있다. 악성코드 분류에 대한 연구는 꾸준히 증가하고 있지만 랜섬웨어 분류에 대한 연구는 현저히 적다. 이에 따라 본 논문에서는 랜섬웨어 바이너리 파일을 시각화하여 육안으로도 구별 가능함을 보여주고 추후 CNN 모델에 적용하여 학습 모델의 정확도를 높이는 연구를 진행하고자 한다.

I. 서론

McAfee Labs Threate Report에 따르면 새로운 랜섬웨어는 계속하여 발견되고 있으며, 최근 각 분기별 발견된 최신 랜섬웨어가 수천만 개에 달한다. [1] 특히 McAfee ATR에서 관찰한 위협의 양은 분당 688개의 위협으로, 2021년 1분기에는 분당 40개의 위협이 증가하였다. 2020년 4분기부터 2021년 1분기까지 주목할만한 부문 증가는 기술, 교육, 금융 및 보험 분야임을 보아 환경의 변화 뿐만 아니라 우리의 일상생활이 편리해지는 만큼 그에 따른 보안 문제 또한 증가함을 알 수 있다. 랜섬웨어의 수는 빠르게 증가할 뿐만 아니라 공격의 유형이 다양화되고 지능화되고 있다. 특히 Babuk, Conti, Ryuk, DarkSide 등 다양한 랜섬웨어가 2021년 공격 트렌드로 등장하였으며 소규모 랜섬웨어는 감소하고 있는 반면 서비스형 랜섬웨어

(RaaS)는 대규모 조직과 회사를 표적으로 공격이 증가하였다. 이를 보아 랜섬웨어 공격대상이 더 적지만 수익성이 높은 대상으로 이동한 것으로 보인다. [2]

이러한 랜섬웨어에 대응하기 위해서는 빠른 분석과 함께 적절한 방어기법이 제공되어야 한다. 랜섬웨어를 효과적으로 분석하기 위해서는 랜섬웨어를 적절히 분류하는 것이 필요하다.

현재까지 악성코드 분류 시스템에 대한 연구는 꾸준히 진행되고 있지만 랜섬웨어 분류 시스템에 대한 연구는 적으며 새로운 유형과 버전의 증가로 인해 랜섬웨어 분석에는 많은 어려움이 따른다.

본 논문에서는 기존의 악성코드 및 랜섬웨어 시각화 기법을 분석한다. 랜섬웨어 이진 파일의 전 영역을 기반으로 하는 전처리 방식을 제안하고 추후 더 다양한 방식의 시각화 방식을 연

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과로 수행되었음(2016-0-00022)

구하고 딥러닝에 적용하여 기존의 방식보다 랜섬웨어 탐지율이 높아지는 것을 실험하고자 한다.

II. 관련연구

2.1 머신러닝 기반 악성코드 탐지 연구

머신러닝 기반 악성코드 탐지란 정상파일 및 악성코드로 모델을 학습시킨 후 학습된 모델로 악성코드를 탐지하는 것이다. 더욱 효과적으로 탐지하기 위해서는 전처리를 통해 적절한 입력 데이터를 만들어야 한다. 대표적인 방법으로 Natarai et al의 연구[3]에서와 같이 악성코드 바이너리 파일을 [0,255] 범위에 있는 8비트 벡터 이미지로 변환하여 그레이스케일 이미지로 시각화하고 평균 해시값을 사용하여 동일한 aHash 값을 가진 이미지는 동일한 유형으로 라벨을 지정하는 방식과 같이 이미지로 표현하는 방법과 kasperskey Lab의 연구[4]에서와 같이 유사성 해싱 함수를 학습 시킨 후 분류하는 방식인 특정 정보를 이용하는 방법이 있다.

악성코드는 실행 방식에 따라서 패밀리가 구분되며, 같은 패밀리의 악성코드는 유사한 특징을 가지고 있어 이미지로 변환했을 때 유사한 이미지가 형성된다. 이를 기반으로 이미지를 생성하는 방법을 설명 후 랜섬웨어에도 적용하여 패밀리별로 이미지를 생성하여 비교하려고 한다.

2.2 악성코드 시각화 관련 연구

기존의 악성코드 데이터의 시각화 기법으로는 [5] 연구에서는 악성코드 opcode 시퀀스에서 특징을 추출한 후, SimHash를 적용하여 비트 값이 0이면 픽셀 값은 1로, 1이면 255로 변환하여 그레이스케일 이미지를 생성한다.

[6]에서는 악성코드를 RGB 색상 이미지로 시각화하고 이미지에서 전역 기능을 추출하는 새로운 접근 방식을 제안하였다. 또한 일련의 특수 바이트 시퀀스를 멀웨어의 코드 섹션과 데이터 섹션에서 추출하고 이는 Simhash에 의해 기능 벡터로 처리된다.

제안된 관련 연구들은 육안으로 식별하기 어렵고 가변적인 크기의 특정 정보를 이용하여

번거롭기 때문에 본 연구에서는 전 영역을 추출하여 그레이스케일 이미지를 생성하여 랜섬웨어를 분류할 수 있는 방법을 제안한다.

III. 악성코드 및 랜섬웨어 파일 구조

악성코드 분류는 악성코드의 유사성에 기반한다. 동일한 패밀리의 악성코드는 사용하는 라이브러리, 함수 등의 특징이 유사하기 때문에 이를 기반으로 분류한다. Windows 환경에서 동작하는 PE 실행파일은 헤더(header)와 섹션(section)들로 구성된다. PE 파일의 실행을 위해서는 운영체제에 실행파일의 정보를 제공할 필요가 있다. 예를들면 파일이 실행 가능한 플랫폼의 종류, 시작 코드의 위치 등 많은 정보를 제공하며 PE 헤더의 IAT(Import Address Table)로부터 악성코드가 사용하는 API 정보를 획득할 수 있다.[7] 위 정보를 바탕으로 악성코드 및 랜섬웨어를 분석하면 PE헤더 구조체에서 유용한 정보들이 있음을 알 수 있다. 특히 랜섬웨어 제작자들은 암호화 알고리즘을 구현할 때 윈도우에서 제공하는 WinCrypt API를 사용하는 등 악성코드와의 차이점이 존재한다. 이를 기반으로 랜섬웨어의 전 영역을 추출하여 시각화하고 추후 특정 영역을 추출하여 시각화하는 방법을 연구하고자 한다.

IV. 랜섬웨어 시각화 기법 구현 및 실험

본 논문에서는 이미지 인식 분야에서 좋은 성능을 보이는 CNN을 이용하여 랜섬웨어의 패밀리를 분류할 수 있도록 랜섬웨어 바이너리에서 각 바이트를 gray-scale 이미지로 변환하여 특징 정보를 생성한다. 또한 CNN 기반의 랜섬웨어 패밀리 분류를 위해 이미지를 일정한 크기(256x256)로 변환하는 과정을 포함한다. 랜섬웨어 패밀리로는 Gandcrab, Maze, NetWalker, Ryuk, Sodinokibi를 포함한다.

다음의 표1을 통해 볼 수 있듯이 랜섬웨어 패밀리별 시각화 한 결과 육안으로도 구별이 가능할 정도의 결과가 나온 것을 볼 수 있다.

[표 1] 랜섬웨어 패밀리별 시각화 결과

Family	Output		
Gandcrab			
Maze			
NetWalker			
Ryuk			
Sodinokibi			

V. 결론

본 논문에서는 꾸준히 증가하고 다양화되는 랜섬웨어를 효과적으로 분류하기 위하여 딥러닝에 효율적으로 적용할 수 있는 랜섬웨어 시각화 전처리 기법을 분석하였다. 본 논문에서는 랜섬웨어 분류에 대한 연구가 부족함과 기존의 관련 연구들은 육안으로 식별하기 어려움을 개선하여 랜섬웨어 시각화 방법으로 랜섬웨어 바이너리 파일의 전 영역을 추출하여 각각의 바이트를 gray-scale 이미지로 변환한다.

랜섬웨어 전 영역을 추출하여 시각화 한 결과 육안으로도 구별이 가능했다. 이러한 결과를 토대로 주요 영역을 추출하여 서로 다른 색상 코드로 나타내거나 3D로 추출하여 패밀리를 분류하고 이를 CNN 모델에 적용하여 학습 모델의 정확도를 높이는 연구를 진행하고자 한다.

[참고문헌]

- [1] D.S.Milojicic, V.Kalogeraki, R.Lukose, K.Nagaraja, J.Pruyne, B.Richard, S.Rollins and Z.Xu, Peer to Peer Computing, HP Laboratories Palo Alto HPL-2002-57, March, 2002.
- [1] McAfee Labs Threats Report, <https://www.mcafee.com/enterprise/en-us/assets/reports/rp-threats-jun-2021.pdf>. (accessed August 22, 2021)
- [2] (2021) MacAfee website [online]. Available at: <https://www.mcafee.com/enterprise/ko-kr/lp/insights-preview.html>
- [3] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, "Malware images: visualization and automatic classification," Proc. of the 8th international symposium on visualization for cyber security, pp. 1-7, 2011.
- [4] Machine Learning for Malware Detection, Kaspersky Lab, Edward Raff, Malware Detection by Eating a Whole EXE, NVIDIA, 2017
- [5] S. Ni, Q. Qian, and R. Zhang, "Malware Identification Using Visualization Images and Deep Learning," Computers and Security, Vol. 77, pp. 871-885, 2018.
- [6] J. Fu, J. Xue, Y. Wang, Z. Liu, C Shan, "Malware visualization for fine-grained classification", IEEE Access, 6 (2018), pp. 14510-14523
- [7] Deok-Jo Jeon, Dong-Gue Park, "Real-time Malware Detection Method Using Machine Learning", The Journal of Korean Institute of Information Technology - Vol. 16, No. 3, pp.101-113.

망 분리 환경에서의 안전한 원격근무 지원을 위한 인프라 구성 방안

박종현*, 김창훈*, 권상오**, 박성수**

*대구대학교 대학원 컴퓨터공학

**포위즈시스템 부설연구소

Infrastructure design strategy for securing remote work in network separation environment

Jong-hyun Park*, Chang Hoon Kim*, Sang-oh Kwon**, Seongsu Park**

*Department of Computer Engineering, Daegu University

**Annex Research Institute, Forwiz System Co.LTD

요약

경계방어기반의 네트워크보안 모델은 COVID-19로 인한 재택근무가 이루어지면서 더 이상 안전성을 제공하지 못한다. 이러한 문제점을 해결하기 위해 네트워크 망분리를 구축한 국내의 많은 기관은 재택근무의 보안성을 향상시키기 위해 VDI서버 기반의 환경을 구축 및 사용하고 있다. 그러나 VDI서버는 망접점 생성, VM의 악성코드 감염 및 전파, VM 우회 공격 등 치명적인 취약점을 갖고 있어 이에 대한 대책이 절실히 요구된다. 본 논문에서는 VDI서버의 안전을 위해 Zero-Trust기반의 구조 및 서비스 절차를 제안하고 VDI서버에서 발생 가능한 취약점 및 위협에 대한 방어가 가능함을 실험을 통해 증명한다.

I. 서론

2020년 2월부터 진행된 COVID-19의 확산으로 기업 및 기관의 경우 원격접속을 통한 재택근무가 일반화되고 있다. 재택근무는 포스트 팬데믹 시대에도 계속 진행될 것으로 예상되며, 향후 일상적인 근무형태로 이루어질 것으로 전망된다.

각 기관에서는 안전한 재택근무지원을 위한 다양한 형태의 환경을 구성하여 지원하고 있다. 지원하고 있는 인프라는 보안성에 대한 기술적 확인뿐만 아니라 많은 경험과 운영노하우를 통하여 그 실효성과 안전성을 확보하고 있다. 널리 사용되고 있는 재택근무지원 인프라 환경은 가상화 데스크톱 기반(VDI: Virtual Desktop Infrastructure)을 이용하는 방식과 외부 단말기에서 업무용 단말기로 원격접속을 이용하는

방식, 외부단말기에서 내부 업무용 단말기를 경유하지 않고 내부서버 등에 직접 접속하는 방식을 사용한다[1]. 여기서, 외부 단말기에서 업무용 단말기로 원격 접속하는 방식과 내부서버 등에 직접 접속하는 방식은 망 분리 환경을 무력화할 수 있는 문제가 있어 본 논문에서는 검토 대상에서는 제외하기로 한다.

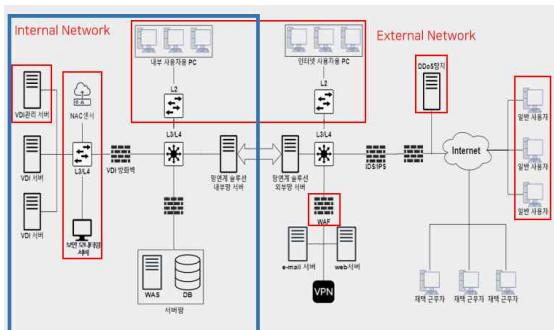
본 논문은 재택근무를 위한 기본 구성요소인 단말기, SSL VPN, 망연계 스트림 서버, VDI서버에서 가질 수 있는 대표적인 취약점을 발굴하고 이에 대한 대안을 제안한다. 제안하는 시스템은 Zero-Trust[2] 개념에 기반하며, 망 접점생성을 통한 망우회 공격, 가상단말기 악성코드 전파, 비인가 가상단말기를 통한 권한 우회의 공격에 대한 방어가 가능함을 제시한다.

II. 관련 연구

재택근무지원을 위한 기술적 환경은 사용자에 따른 다양한 컴퓨팅 환경과 업무형태에도 불구하고 기관의 중요한 정보를 보호하는데 목적이 있다. 하지만 현실에서는 동일한 범위의 네트워크에 위치한 다양한 보안 수준을 가진 컴퓨팅 기기로 인해 기관의 중앙 집중화된 보안통제 방식의 적용은 매우 어렵다. 따라서 재택근무 환경을 구성하는 구성 요소별로 발생 가능한 보안취약점을 도출하고 이에 대한 대응책을 마련해야 한다.

2.1 재택근무 구성요소별 보안 위협 요인

재택근무 지원을 위한 구성요소와 서버를 위한 연결구조는 <그림 1>과 같이 구성할 수 있다.



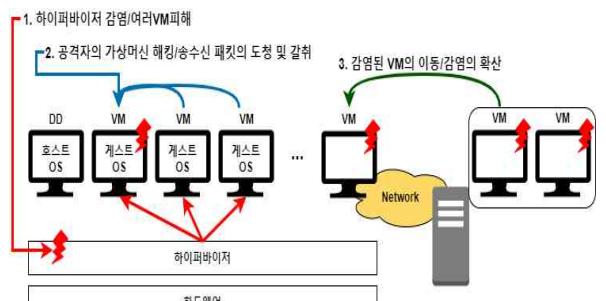
<그림1> 재택근무지원을 위한 환경구성 요소

재택 근무자들은 노트북, 태블릿, 스마트폰 등 다양한 형태의 모바일 단말기를 이용하여 노출된 물리적 공간에서 내부망의 VDI 서버로의 접근이 가능하다. 이런 업무 환경적 특징은 사용자 부주의로 인하여 단말기 도난·분실 또는 보안 관리의 부주의에 따른 악성코드 감염, 중요정보 유출 등 다양한 위협 상황을 만들 수 있다. 또한 우리가 중점적으로 고려해야 할 내용은 사용자 인증과 통신구간의 암호화를 위한 SSL VPN 장치이다. SSL VPN의 경우 특정 벤더사 제품[3]과 같이 취약점을 이용한 관리자 권한 탈취 문제가 있으며, SSL VPN 프로토콜의 구조적인 문제로 접속 완료 후 기존 인터넷 연결 세션을 차단하지 못하는 치명적인 취약점을 가지고 있다. 뿐만 아니라 망 분리 환경에서 사용 중인 망 연계 스트리밍의 경우 웹크롤링과 같은 정상적인 트랜잭션 패턴은 보안정책만으

로는 통제가 불가능하여 대용량 데이터 처리를 요구하는 DDoS공격에 무력화 될 수 있다. 그외 VDI서버를 이용한 추가적인 공격은 아래와 같다.

2.2 VDI서버 보안위협

VDI서버는 외부사용자가 내부 정보시스템이나 데이터에 직접 접속을 차단하고 중계해주는 기능을 제공한다. 가상화 기술로 사용자 단말과 RDP(Remote Desktop Protocol) 프로토콜을 이용하여 이미지 형태의 데스크톱 서비스를 이용한다. 그러나 가상화 서버에서 제공하는 하이퍼바이저의 경우 해킹으로 인한 통제권 상실, 가상화 취약점 상속에 따라 호스트 운영체제의 감염, 그리고 게스트 운영체제로의 추가적인 악성코드 감염을 야기시킨다. 이러한 현실은 <그림 2>와 같이 가상화 환경이 가지는 자원 통합, 재분배하는 공유 시스템의 환경에서 발생하는 구조적인 문제이다[4].



<그림2> 가상머신간 악성코드 전파

하이퍼바이저의 취약점을 이용한 또 하나의 위협요소로 취약한 하이퍼바이저를 경유하여 가상게스트간의 운영체제 통제권을 획득할 수 있다[4]. 정보시스템에 접속권한을 획득하지 못한 가상게스트가 하이퍼바이저를 통해 권한을 가진 게스트운영체제의 권한을 획득함으로써 권한을 우회하는 취약점이 존재한다[4]. 이것은 기존 정보시스템 환경에서 접근제어 시스템을 무력화하는 것이며, 중앙 집중형 접근 통제 정책 차제가 무의미하게 된다. 따라서 가상의 운영체제를 컨트롤하고 실제 데이터 통신경로를 제어하는 특권을 가진 하이퍼바이저 레벨을 관리하는 보안기술과 각 게스트 운영체제간 접근 통제 차제는 반드시 필요하다.

2.3 VDI서버를 통한 망 우회공격

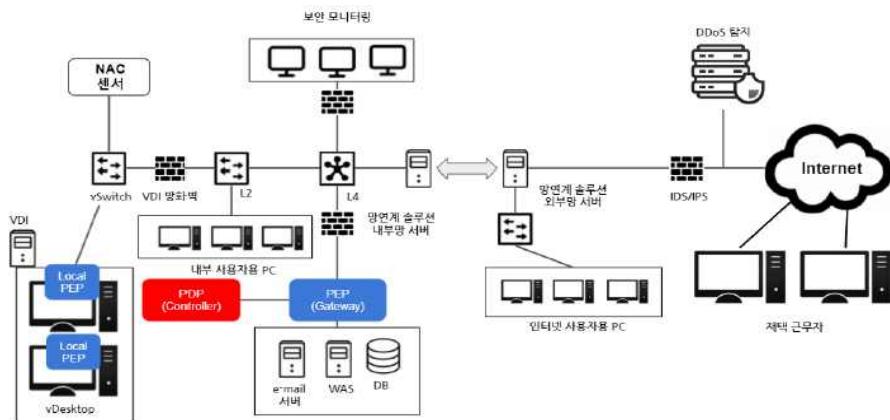
공공·금융기관은 많은 노력과 비용을 투자하여 외부의 사이버공격을 효과적으로 차단하고 내부 중요 정보자산을 보호하기 위하여 내부망과 외부망을 분리하여 운영하고 있다. 그러나 VDI서버 기반의 가상 데스크톱 환경에서 외장 형태의 무선기기를 무단으로 연결하거나 그 외 다양한 방법으로 망 분리 우회 공격이 가능하다[5]. 이러한 공격은 비정상적으로 외부와의 접점을 제공하여 사실상 망 분리를 무력화 하고 이를 통하여 악성코드를 직접적으로 감염 시킬 수 있다.

III. Zero-Trust기반 안전한 VDI 시스템 구성

본 절에서는 Zero-Trust 개념에 기반하여 2 절에서 기술한 3가지 공격에 대해 대응을 가지는 VDI서버를 제안한다.

3.1 Zero-Trust기반 VDI 시스템 구성

제안하는 시스템의 논리적인 구조는 <그림 3>과 같다. 그림 3에 제안된 개념의 실험적인 증명을 위해 윈도우 10기반의 VDI서버에 가상 데스크톱 2대(vDesktop1, vDesktop2)를 구축하고 vDesktop2에 Agent(PEP¹)을 설치한다. PEP에서 수집된 정보를 분석하여 정책을 통합적으로 관리하기 위한 Controller(PDP²)를 추가한다. 외부 사용자 단말은 설치된 VDI 클라이언트를 통해 가상데스크톱(VDI)에 접속하고 설치된 PEP는 [표 1]에 따라 테스크톱의 정보를 수집하고 사전에 정의된 목적지IP, 서비스 및 어플리케이션 기반의 화이트리스트 보안정책을 통해 접근 제어를 실시한다.



<그림3> Zero-Trust VDI 실험 구성도

[표 1] PDP 화이트리스트 정책 샘플

정책명	대상IP	포트	소프트웨어	사용자
WAS1	172.16.2.11	80	chrome.exe	test1
WAS2	172.16.2.12	80	chrome.exe	test2
VDI내부	172.16.0.23 1	1~ 65535	all	test1 test2
외부	forwiz.com	80,443	chrome.exe	test1 test2

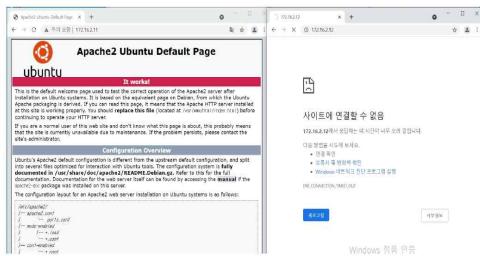
<그림 3>에서 제안한 구조를 <그림 4>와 같이 실제로 구현한 환경을 이용하여 다음의 3가지 공격 방어가 가능성을 검증한다. 1) 권한이 없는 가상데스크톱 단말기가 권한을 우회하기 위하여 다른 가상데스크톱 접근을 시도할 경우 이의 탐지 및 차단 가능 여부, 2) VDI서버의 가상데스크톱 악성코드 감염이 다른 가상 데스크톱으로의 전파 시도에 대한 차단 여부, 3) VDI 단말기에서 외부 망 접속(스마트 폰 테더링을 통한 무선 연결)을 통한 망 우회 탐지 및 차단 여부.



<그림4> <그림 3>의 구현 및 실험환경

3.2 실험 절차 및 검증 절차

화이트리스트 기반 보안 정책에 따라 Local PEP가 설치된 vDesktop2의 test1은 크롬 브라우저를 통해 WAS1 접속이 가능하고 WAS2 접속이 Local PEP에서 차단됨을 <그림 5>에서 확인 할 수 있다. 또한 test2 계정으로는 크롬 브라우저를 통해 WAS2 접속이 가능하고, WAS1 접속이 Local PEP에 의해 차단된다. Local PEP가 설치되지 않은 vDesktop1의 경우 WAS1, WAS2 접속이 모두 차단됨을 실험을 통하여 검증 하였다.



<그림5> vDesktop2의 WAS1접속허용(왼쪽), WAS2접속차단(오른쪽) 결과

vDesktop2에서는 test1 혹은 test2 계정으로 크롬 브라우저를 통해 forwiz.com 접속은 가능하였으나, naver.com 등 보안정책에 설정되지 않은 URL은 정상적으로 차단됨을 확인하였다. 가상데스크톱 간의 통신제어를 확인하기 위하여 nmap의 GUI 프로그램인 zenmap을 통해 vDesktop에서 포트스캐닝을 수행한 결과 vDesktop1, 2 모두 Local PEP를 미사용할 경우 정상적으로 상호간 포트 스캔ning이 가능함을 보이며, vDesktop2에 Local PEP를 사용할 경우 vDesktop2에서 vDesktop1으로의 포트스캐닝이 Local PEP에 의해 차단됨에 따라 vDesktop1의 자세한 정보를 확인할 수 없다. 이는 가상머신 간 악성코드 전파 및 다른 가상머신을 경유하여 권한을 획득할 수 없음을 보여준다.

또한 Local PEP가 설치된 vDesktop2에 외부 무선기기를 연결하여, USB등 매체제어 기능과 외부 네트워크 정보 생성여부 탐지 및 차단을

1) PEP(Policy Enforcement Point) Zero Trust 정책을 적용하여 네트워크 플로우 차단 등의 기능을 수행

2) PDP(Policy Decision Point) Zero Trust 정책을 각 PEP에 배포하는 Controller의 역할을 수행

통해 망간 접점을 이용한 우회 공격을 시도하였지만 차단됨을 확인하였다.

3.3 VDI서버의 위치에 따른 장·단점 비교

[1]의 권고에 따라 VDI서버는 내부망에 설치가 가능하고 이의 경우 망접점이 발생하면 치명적인 단점으로 작용한다. 반대로 VDI서버를 사내 외부망(DMZ 구간 내)에 설치 한다면 각 VM들의 접근 정책을 개별적으로 적용해야하기 때문에 관리의 어려움이 따르고, 특히 VDI서버와 내부망 정보시스템간에 암호화 통신을 구현하더라도 트래픽이 노출되어 보안에 취약할 수 있다. 따라서 본 논문에서는 VDI 서버를 내부망에 위치시키고 실험을 진행 하였다. [표 2]는 VDI서버 구성에 따른 장·단점을 보여준다.

[표 2] VDI서버 구성 방식 비교(망 분리 환경)

구분	외부망설치	내부망설치
내용	VDI서버를 내부망에 설치	VDI서버를 외부망에 설치
구성		
Data 측면	외부사용자 내부망 직접 접속 불필요	VDI서버-정보시스템간 Data 외부망 노출 불필요
접근 허용	VDI서버-정보시스템간 Data 외부망 노출 발생 가능	외부사용자 내부망(VDI서버) 접속 필요
관리 측면	가상머신별 정보시스템 허용 정책으로 관리의 어려움	VDI서버에 대한 접근 정책으로 관리가 쉬움

IV. 결론

본 논문에서는 Zero-Trust기반 안전한 VDI 시스템을 제안하였고 재택근무에 따른 다양한 보안 위협요소에 대응이 가능함을 실험을 통하여 검증하였다. Local PEP와 PDP Controller를 이용하여 가상머신간의 접근 제어를 실시함으로써 악성코드 전파 및 가상머신 경유를 통한

권한 우회 차단을 수행할 수 있었고, 망 접점을 생성하는 취약점에 대한 방어가 가능하였다. 따라서 본 논문에서 제안하는 구조는 보다 안전한 재택근무 환경 제공을 위한 하나의 대안이 될 수 있음을 보여준다.

[참고문헌]

- [1] 금융보안원, “금융회사 재택근무 보안 안내서”, 금융보안원, 2020.12
- [2] 스콧 더블유 로즈, 올리버 보처트, 스튜어트 미첼, 존 코넬리 “제로 트러스트 아키텍처”, NIST, 2020.8
- [3] 보안뉴스, “시큐워즈 VPN 이어 지니언스 NAC까지... 관리자 권한 탈취 가능 취약점 비상!”, 2021.08.19
- [4] 최도현, 유한나, 박태성, 도경화, 전문석, “클라우드 서비스 가상화 내부 환경을 위한 BareMetal Hypervisor 기반 보안 구조 설계”, 한국통신학회논문지, 2013, vol.38, no.7
- [5] 박종현, 김창훈, “NAC솔루션을 활용한 효과적인 망혼용 탐지 기법”, 한국정보보호학회 학술논문, 2021.02

블록체인을 활용한 후불식 톨게이트 지불 시스템 구현

박재훈*, 권혁동*, 서화정*†

*한성대학교 IT융합공학부 (대학원생)

*† 한성대학교 IT융합공학부 (교수)

Implementation of Tollgate Deferred Payment System Using Blockchain

Jae-Hoon Park*, Hyeok-Dong Kwon*, Hwa-Jeong Seo*†

*Division of IT convergence engineering, Hansung University.
(Graduate student)

*† Division of IT Convergence Engineering, Hansung University
(Professor)

요약

고속도로를 비롯한 유료 도로에 설치된 톨게이트는 통행료를 걷기 위해 필수적으로 존재해야 하며, 톨게이트 진입 시 속도를 낮춰야 하고 도로의 폭이 좁아지는 등의 문제로 인해 톨게이트 앞은 상습적으로 정체 구간이 된다. 본 논문에서는 이러한 문제를 해결하기 위해 차량 인식을 통한 후불식 톨게이트 지불 시스템을 제안한다. IBM의 프라이빗 블록체인 프레임워크인 하이퍼래저 패브릭을 사용하여 구현하였으며, 블록체인을 이용하였기에 설치 비용을 절약하고 데이터의 신뢰성을 보장할 수 있다. 본 시스템을 이용하여 톨게이트 구간에서의 정체를 완화하는 것을 기대한다.

I. 서론

톨게이트는 고속도로를 비롯한 유료 도로에서 통행료를 걷기 위한 시스템이다. 과거에는 현금만을 지불하였으나, 하이패스가 이용되기 시작한 이후로 정부에서는 하이패스의 이용을 좀 더 권장하고 있다. 실제로 그림 1과 같이 2007년 하이패스 도입 이후 10년 만인 2017년에 하이패스 이용률이 넘었을 정도로 하이패스의 이용률은 증가하고 있다.



(그림 1) 하이패스 이용률 변화[1]

하지만 하이패스를 이용함에도 불구하고 톨게이트 앞에서의 정체는 여전히 유발된다. 이는 지불 방법이 현금, 하이패스인 것에 관계 없이 톨게이트 앞에서는 결국 속도를 줄여야 하기 때문이다. 본 논문에서는 이러한 문제를 개선할 수 있도록 블록체인을 활용한 후불식 톨게이트 지불 시스템을 구현하였다.

II. 관련 연구

2.1. 블록체인

블록체인은 사토시 나카모토가 비트코인을 개발할 당시 도입한 방식으로, P2P (Peer-to-Peer) 네트워크에서 데이터의 무결성을 보장하기 위해 데이터를 블록형태로 만든 뒤 체인 형태로 묶은 것이다[2]. 이것을 통해 특정 데이터가 위/변조 되더라도 타 블록과의 해시값 비교를 통해 무결성을 보장할 수 있다.

2.2. 하이퍼래저 패브릭

하이퍼레저 패브릭은 리눅스 재단에서 시작된 프라이빗 블록체인 프로젝트인 하이퍼레저 프로젝트로부터 시작된, IBM에서 주도하고 있는 오픈소스 블록체인 프레임워크이다. 블록체인의 신뢰성을 이용해 체인코드[3]라는 비즈니스 로직을 호스팅 하며, 이것은 Go, Node.js, Java와 같은 상용 언어들을 통해 작성될 수 있다.

본 논문에서는 하이퍼레저 패브릭 2.2 LTS 버전을 이용하여 후불식 톤게이트 지불 시스템을 구현하였다.

III. 후불식 톤게이트 지불 시스템

본 논문에서는 차량 번호 인식을 통해 차후 운전자에게 요금을 부과하는 완전 후불 방식을 제안한다. 블록체인을 이용하여 중앙 서버 없이 각 톤게이트에서 장부를 관리한다. 그렇기에 설치비용이 절약되며, 향후 DID 등을 이용하여 추가적인 확장이 가능하다. 또한 결정적으로 이것을 통해 고속도로 흐름을 유지하여 정체를 완화시킬 수 있다.

시스템을 구현하기 위해 하이퍼레저 패브릭 2.2 LTS 버전을 이용하였으며, 체인코드는 TypeScript를 통해 작성되었다. 네트워크 상에는 1개의 채널이 존재하며, 예시를 위해 톤게이트 조직을 1개만 이용하였다. 본 서비스가 실제로 이용될 경우 각 톤게이트가 하나의 조직으로써 네트워크에 참여할 수 있다.

체인코드에는 Record, Car, Driver라는 3개의 클래스가 존재하며, 각각 표 1, 2, 3과 같다. Record 클래스는 각 부과 기록에 관한 클래스이며, Car 클래스는 차량 등록에 관한 정보, Driver 클래스는 운전자에 관한 정보이다.

Field	Type	Description
key	string	key of the record
car_num	string	car number
timestamp	number	time when the car entered

tollgate	string	tollgate name
money	number	money to charge

(표 1) class Record

Field	Type	Description
key	string	key of the car
car_num	string	car number
car_type	string	size type
car_name	string	car name
owner	string	owner's ID

(표 2) class Car

Field	Type	Description
key	string	key of the car
phone	string	phone number (ID)
charged	number	entire charged money

(표 3) class Driver

체인코드는 톤게이트를 지나는 운전자에게 요금을 부과하고, DB에 새 통과 기록을 저장하는 charge 함수와 요금을 납부할 때 사용하는 pay 함수가 있으며 각 표 4, 5와 같다.

function charge(car_num, tollgate_name)
let car := getCar(car_num)
let driver := getDriver(car.owner)
driver.charged = driver.charged +
getChargedMoney(car.car_type)
Update driver to the database
Create new record using car_num,
timestamp, tollgate_name, money
Insert the record to the database

(표 4) function charge

function pay(phone, payed)
let driver := getDriver(car.owner)
driver.charged = driver.charged - payed
Update driver to the database

(표 5) function pay

IV. 결론

본 논문에서의 시스템은 블록체인을 이용하여 중앙 서버가 필요한 통상의 시스템에 비해 설치 효과가 절감되며, 신뢰성 있는 요금 관리가 가능하다. 이 시스템이 실제로 적용되어 고속도로 등 톨게이트가 설치된 곳에서의 도로 흐름 유지 및 정체 완화를 기대한다.

V. Acknowledgment

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.2018-0-00264, IoT 융합 형 블록체인 플랫폼 보안 원천 기술 연구, 100%).

[참고문헌]

- [1] “하이패스 이용률 80% 돌파…2007년 도입 10년 만에 달성”, 매일신문 [Internet] available: <https://bit.ly/3j5eg9t>
- [2] Nakamoto, Satoshi. Bitcoin: A peer-to-peer electronic cash system.
- [3] Beckert, Bernhard, et al. "Formal specification and verification of Hyperledger fabric chaincode." 3rd Symposium on Distributed Ledger Technology (SDLT-2018) co-located with ICFEM. 2018.

RE100 실현을 위한 블록체인 기반 REC 거래 플랫폼

김재석* 황보규민* 최윤호**

*부산대학교(대학원생) **부산대학교(교수)

A blockchain based REC transaction platform for realization RE100

Jae-Seok Kim* Hwangbo Gyumin* Yoon-Ho Choi**

*Pusan National University(Graduate student) **Pusan National University(Professor)

요약

RE100은 기업이 사용하는 모든 에너지를 신재생에너지로 대체하겠다는 글로벌 신재생에너지 캠페인으로, 국내에서는 이를 실현하기 위해 에너지 거래에 있어서 에너지 거래량을 증명 가능한 REC(Renewable Energy Certificate)를 함께 거래하고 있다. 하지만, 기존 REC 계약은 한국 전력공사를 통한 독점적 계약 형태로 이루어져 있어 거래 내역의 무결성 침해를 일으킬 수 있으며, 거래를 검토하고 승인하는 과정이 수동적으로 이루어지고 있다. 본 논문에서는 이러한 REC 계약 구조를 블록체인으로 재구성하여 자동화하고, 데이터의 무결성을 유지하는 블록체인 기반 REC 거래 플랫폼을 제안한다. 제안 블록체인은 트랜잭션 이중서명을 통한 쌍방계약 증명, 친환경적인 합의과정이라는 특징을 지니며, 이로 인해 거래의 신뢰성과 접근성을 확보하여 국내 전력시장 개방과 국내 기업의 RE100 참여도 향상이 가능할 것으로 예상된다.

I. 서론

RE100은 기업이 사용하는 모든 에너지를 신재생에너지로 대체하겠다는 글로벌 신재생에너지 캠페인으로 국내에서 이를 실현하기 위해 에너지 거래에 있어서 REC(Renewable Energy Certificate)를 함께 거래하고 있다. REC는 에너지 종류, 에너지량, 공급자와 소비자 정보 등을 담고 있으며, RE100에서는 기업의 에너지 보유량을 REC 보유 여부로 증명 가능하다. 하지만, 한국전력공사를 통한 독점적 계약 형태로 이루어져 있는 기존 REC 계약은 REC 거래 내역의 무결성이 침해될 수 있으며, 거래를 검토하고 승인하는 과정이 수동적으로 이루어지고 있다.

블록체인은 수동적으로 이루어지는 이러한 REC 계약을 합리적인 합의 알고리즘을 통해 자동화할 수 있으며, 거래 내역이라는 트랜잭션을 개인키로 서명함으로써 데이터 무결성을 유지할 수 있다는 장점을 가지고 있다. 하지만 이러한 REC 계약 구조를 기존 블록체인에서 활

용하기에는 여러 한계점이 존재한다.

첫 번째, 기존 블록체인에서는 암호화폐의 단방향 전송이라는 특징으로 인해 트랜잭션은 전송자의 개인키로만 서명되고 수신자의 서명은 필요하지 않다. 반면, REC 계약은 에너지 가격과 양을 소비자와 공급자 양측의 동의가 모두 필요한 쌍방계약으로 기존 블록체인의 트랜잭션 구조를 그대로 활용할 수 없다.

두 번째, 블록체인의 대표적인 합의 알고리즘으로 하드웨어 자원을 소모하여 블록을 생산하고 검증하는 작업증명(PoW), 노드가 보유한 지분에 따라 블록 생성권이 결정되는 지분증명(PoS)이 존재한다. 하지만, 작업 증명은 블록 생성 과정에서 많은 에너지 소비를 일으키므로 RE100의 목적에 반하게 되며, 지분 증명은 지속적으로 거래가 이루어지는 에너지를 지분으로 활용할 수 없다.

본 논문에서는 이러한 한계점을 해결하기 위해 트랜잭션 이중서명 구조를 통한 REC 쌍

FROM(ID)	ENERGY	MONEY	TO(ID)	signature1	signature2	Fee
Amy	50	41\$	M1	Sig(Amy, Tx1)	Sig(M1, Sig(Amy, Tx1))	5%
Bob	30	22\$	M3	Sig(Bob, Tx2)	Sig(M2, Sig(Bob, Tx2))	7%
...

그림 1 제안 이중서명 기반 트랜잭션 구조

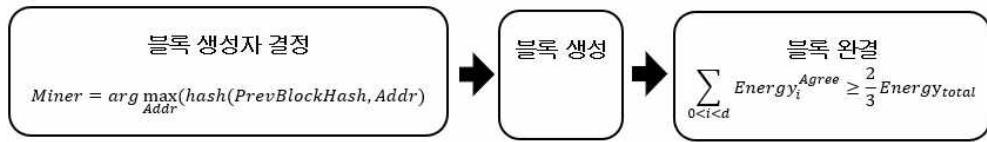


그림 2 제안하는 합의 알고리즘 동작 순서

방계약, 친환경적인 합의 알고리즘을 제안하며, 제안된 블록체인을 기반으로 동작하는 REC 거래 플랫폼을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서 제안 블록체인의 특징인 트랜잭션 이중서명, 친환경 합의 알고리즘을 설명한 뒤, 3장에서 프로토 타입 구현 결과를 서술한다. 이후 4장에서 논문의 결론을 서술한다.

II. 블록체인 기반 REC 거래 플랫폼

본 장에서는 REC 계약이 블록체인 기반으로 동작하기 위해 필요한 트랜잭션 이중서명 및 친환경 합의 알고리즘에 대해 기술한다.

2.1 트랜잭션 이중 서명

기존 블록체인의 트랜잭션은 송신자, 수신자, 암호화폐량, 송신자의 서명, 수수료로 구성되어 있다. 이러한 트랜잭션의 구조는 수신자의 동의가 필요 없는 단방향 거래를 지원하며, 실행될 때 채굴자에게는 수수료가, 수신자의 지갑에는 기입된 암호화폐량이 전송된다.

이와 달리, 제안 블록체인에서 REC 계약을 위한 트랜잭션은 에너지 구매자, 에너지 판매자, 암호화폐량, REC(에너지 거래량), 에너지 구매자의 서명, 에너지 판매자의 서명, 수수료로 구성된다. 이러한 트랜잭션은 에너지 구매자가 에너지 판매자에게 일정 암호화폐를 지불하고 에너지를 받는 쌍방계약을 지원할 수 있다. 트랜잭션이 실행 시에는 기존 블록체인과 동일하게 수수료와 암호화폐가 전송되며, 추가적으로 판매자로부터 구매자에게 트랜잭션에 기입된 에너지 거래량이 전송된다. 이러한 트랜잭션은 블록에 담기고 블록체인에 연결됨으로

인해 거래 내역이라는 의미를 지니게 되므로, 판매자는 구매자에게 실제 REC와 에너지를 공급할 의무를 가진다. 그림 1은 제안하는 이중서명 기반 트랜잭션 구조를 나타낸다. 트랜잭션을 생성하는 에너지 구매자는 서명을 제외한 트랜잭션 각 항목에 원하는 거래를 기입한다. 이후 signature1에 트랜잭션을 자신의 비밀키로 서명한 값을 기입하여 네트워크로 전파한다. 에너지 판매자는 자신의 주소와 전파된 트랜잭션의 수신자가 일치하는지 확인한다. 이후 에너지량과 암호화폐량의 합리성을 검토하고 signature2에 자신의 비밀키로 서명하여 네트워크로 다시 전파한다. 이는 구매자와 판매자의 쌍방계약이 완료되었음을 의미하며, 블록 생성자는 이중 서명이 완료된 트랜잭션만을 블록에 담아 채굴한다.

2.2 친환경 합의 알고리즘

본 논문에서 제안하는 친환경 합의 알고리즘은 2단계를 거쳐 이루어진다. 먼저 블록 생성자를 결정하고, 블록이 생성되면 Tendermint[1]의 2/3 Majority Condition을 기반으로 이를 검증하고 완결한다. 에너지 구매자와 판매자의 수를 고려하여 합의 과정에 참여하는 노드는 판매자로만 구성한다. 이는 소수의 노드로만 이루어진 합의로 인해, 네트워크의 블록 생산 속도를 향상시킨다.

합의에 참여하는 모든 노드는 블록 생성자가 되기 위한 후보군이다. 이러한 후보군 중에서 블록 생성자를 결정하는 수식은 다음과 같다.

$$\text{Miner} = \arg \max_{\text{Addr}} \text{hash}(\text{PrevBlockHash}, \text{Addr})$$

먼저, 각 노드는 이전 블록의 해시와 각 후보의 주소를 합쳐 새로운 해시를 구한다. 이 해시 중 가장 큰 해시를 띠는 후보는 블록 생성

```
=====
모든 Node가 생성되었습니다.
제네시스 블록을 생성합니다.
현재 메인 네트워크 참가자 크기: 10

Peer List 출력 -----
Node : fe75a5b0e56d21c0, Money : 10, Energy : 1000, State : 1
Node : b402eb076b652d2, Money : 20, Energy : 2000, State : 1
Node : eb944df2bd01e5b, Money : 30, Energy : 3000, State : 1
Node : afc23a5c7ef20c4a, Money : 40, Energy : 4000, State : 1
Node : c0b622cd3a18ced7, Money : 50, Energy : 5000, State : 1
Node : 7a773b68764157b3, Money : 100, Energy : 10, State : 0
Node : a95d0fad6933e2c4, Money : 200, Energy : 20, State : 0
Node : e8431ad34886f174, Money : 300, Energy : 30, State : 0
Node : 4c30c5db63586bc3, Money : 400, Energy : 40, State : 0
Node : 9509138a041bfff8f, Money : 500, Energy : 50, State : 0
```

그림 3 프로토타입 네트워크의 노드 리스트
자가 된다. 이는 이전 블록 해시를 블록 생성 전까지 알 수 없으므로 블록 생성자의 예측이 불가능하도록 하며, 불필요한 Nonce 계산을 없애 하드웨어 자원 소모를 줄인다.

블록 생성자가 결정되고 트랜잭션을 담은 블록을 네트워크에 전파하면, 모든 후보군은 블록에 대해 검증한다. 검증은 블록 생성자 검증, 트랜잭션 검증(이중 서명 여부, 정상 실행 여부)으로 이루어진다. 각 후보는 검증이 정상적으로 완료되면 투표를 실행한다. 투표는 Tendermint의 2/3 Majority Condition을 따르며, 이를 적용한 수식은 다음과 같다.

$$\sum_{0 < i < d} Energy_i^{Agree} \geq \frac{2}{3} Energy_{total}$$

만약 전체 후보군의 에너지 보유량 3분의 2 이상을 만족하는 투표권이 모이면, 해당 블록을 완결하고 합의를 종료한다. 이는 블록 생성자가 악의적인 블록 생성을 하는 경우, 검증자 역할을 하는 후보군에 의해 생성된 블록을 차단하는 역할을 하여 블록에 신뢰성을 부여한다.

III. 프로토타입 구현

프로토 타입은 노드 생성, 블록 생성자 및 후보군 결정, 트랜잭션 생성 및 이중 서명, 블록 생성, 블록 검증 및 완결 순서로 동작한다.

노드 생성은 RSA 키 및 지갑 생성으로 이루어지며, 각 노드는 State에 따라 에너지 구매자와 판매자로 구분된다. 프로토타입 네트워크에서 생성된 전체 노드는 10개로, 그 중 5개의 노드를 블록 생성자 후보군인 에너지 판매자로 설정하였으며, 이는 그림 3에서 볼 수 있다. 노

```
===== TxID: ad1837c8050b95f450c5f87c946eeebdae25c48c =====
From: 7a773b68764157b3
To: fe75a5b0e56d21c0
Energy: 10
Money: 10
GasPrice: 0.02
sig1: 008F532CCB1C42CA07829AD0128CEB30
sig2: 32ECB9C3713CA44C5B52EBF1A18B35C8
=====
```

그림 4 이중 서명이 완료된 트랜잭션

드 생성이 끝난 후, 후보군 내에서 블록 생성자를 결정한다. 그림 4는 이중 서명이 완료된 트랜잭션으로 From, To의 각 서명이 sig1, sig2에 저장된 모습을 보이고 있다. 블록 생성자는 이렇게 이중 서명 여부가 모두 확인된 트랜잭션만을 블록에 담아 패키징한다. 블록 검증 단계에서는 트랜잭션의 실행 결과를 검증하고, 투표를 진행한다. 프로토타입은 투표 과정에서 임의의 노드 네트워크에 장애를 일으키면서 2/3 Majority Condition 만족 여부를 검증한다. 블록 검증과 투표가 정상적으로 완료되면, 블록을 체인에 연결하고 다음 블록의 생성자를 결정하며 이전 동작을 반복한다.

IV. 결 론

본 논문에서는 기존 REC 계약 구조를 블록체인으로 재구성하여 신뢰성과 접근성을 높인 REC 거래 플랫폼을 제안하였다. 제안 블록체인은 REC 거래 내역의 무결성이 제공됨에 따라 국내 전력시장 개방을 유도할 수 있다. 또한, 국내 기업의 RE100 참여도 향상과 신재생에너지 거래 시장 활성화에 기여할 수 있을 것으로 기대된다.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업(IITP-2020-0-01797) 및 융합보안핵심인재양성사업(No. 2019-0-01343)의 연구결과로 수행되었음

[참고문헌]

- [1] Ethan Buchman, Jae Kwon, Zarko Milosevic, The latest gossip on BFT consensus, November, 2019.

단순 전력 분석 공격에 강인한 타원 곡선 point multiplication 방법

최필주*

*부경대 IT융합응용공학과

Elliptic Curve Point Multiplication Method Strong Against Simple Power Analysis Attack

Piljoo Choi*

*Dept. of IT Convergence and Application Engineering, Pukyong National University

요약

타원 곡선 암호의 주요 연산인 point multiplication을 수행할 때 소모 전력을 관찰하면 점에 곱하는 비밀 값을 분석할 수 있다. 이를 방지하기 위해 비밀 값에 의존적인 연산을 비밀 값과 상관없이 일정하게 수행되도록 하는 point multiplication 방법이 있으나 실행 시간이 증가한다는 단점이 있다. 본 논문에서는 이러한 overhead를 줄인 새로운 point multiplication 방법을 제안한다. 실행 시간을 예측한 결과는 제안 방법이 affine 좌표계를 사용하는 경량 응용에 대해 point multiplication에 필요한 modular multiplication과 modular addition의 수를 약 4/5 수준으로 줄일 수 있음을 보여준다.

I. 서론

대표적인 공개키 암호 알고리즘 중 하나인 타원 곡선 암호 [1], [2]의 주요 연산은 타원 곡선 위의 점에 대한 곱셈 연산, 즉 point multiplication (PntMult)이며 타원 곡선 암호에 기반한 ECDH, ECDSA 등의 암호 프로토콜에서 PntMult 수행 시 점에 곱하는 값은 개인 키 또는 임시로 사용되는 비밀 난수 값 등의 비밀 정보이다. 이러한 비밀정보에 따라 PntMult 수행 시 반복 수행되는 point doubling (PntDbl)과 point addition (PntAdd)의 실행 여부가 달라지며 이는 전력 파형을 통해 관찰할 수 있다. 이러한 공격을 단순 전력 분석(Simple Power Analysis, SPA) [3]이라고 하며 이를 통해 효과적으로 비밀정보를 추출할 수 있다.

SPA를 방지하기 위해 비밀정보에 상관없이 항상 PntDbl과 PntAdd를 수행하는 double-and-add-always (DAA) [4], Montgomery ladder (ML) [5] 등의 PntMult 방

법이 사용되나 이러한 방법들을 적용할 경우 실행 시간이 많이 늘어난다는 단점이 있다. 본 논문에서는 기존의 DAA와 ML의 overhead를 줄인 SPA에 강인한 새로운 PntMult 방법을 제안한다.

II. 배경 지식

이 장에서는 배경 지식으로서 기존의 PntMult 방법에 대해 살펴보고 PntDbl과 PntAdd 실행에 필요한 연산을 살펴본다.

1.1 기존의 PntMult 방법

타원 곡선 위의 점 P 에 대해 k -bit scalar 값 s 를 곱하는 PntMult, 즉 sP 를 수행하는 가장 기본적인 방법은 binary method이다. 이 방법은 s 의 비트를 왼쪽에서 오른쪽으로 관찰하면서 PntDbl과 PntAdd를 반복하면서 PntMult를 수행하는 것으로 자세한 알고리즘은 표 1에 나타나 있다.

	Algorithm
Binary method [4]	$Q \leftarrow P$ $\text{for } (i \leftarrow k-2 ; i \geq 0 ; i \leftarrow i-1)$ $Q \leftarrow 2Q$ $\text{if } (s[i] = 1) Q \leftarrow Q + P$ return Q
DAA [4]	$Q_0 \leftarrow P$ $\text{for } (i \leftarrow k-2 ; i \geq 0 ; i \leftarrow i-1)$ $Q_0 \leftarrow 2Q_0$ $Q_1 \leftarrow Q_0 + P$ $Q_0 \leftarrow Q_{s[i]}$ return Q_0
ML [5]	$(Q_0, Q_1) \leftarrow (P, 2P)$ $\text{for } (i \leftarrow k-2 ; i \geq 0 ; i \leftarrow i-1)$ $Q_{s[i]} \leftarrow 2Q_{s[i]}$ $Q_{1-s[i]} \leftarrow Q_0 + Q_1$ return Q_0

표 1. 기존의 PntMult 방법

표 1에서 $s[i]$ 는 s 의 i 번째 bit을 나타내며 k -bit s 의 최상위 비트는 1인 것으로 가정되었다. 즉 $s[k-1] = 1$ 이다. 표 1에서 볼 수 있듯이 binary method에서 PntAdd는 $s[i]=1$ 일 때에만 수행되므로 전력 과형을 관찰하면 PntAdd의 수행 여부를 알아낼 수 있고 이로부터 s 의 값도 분석할 수 있다.

표 1은 SPA 방지 PntMult 방법인 DAA와 ML도 보여주고 있다. 두 가지 방법 모두 $s[i]$ 에 상관없이 항상 PntDbl과 PntAdd를 모두 수행하고 있다. 따라서 단순 전력 분석 공격을 방지 할 수 있으나 binary method보다 더 많은 PntAdd를 수행하므로 그만큼 실행 시간이 늘어 난다는 단점이 있다.

1.2 PntDbl과 PntAdd의 연산

타원 곡선 위의 점은 x, y 좌표를 가지며 두 점 (x_0, y_0) 과 (x_1, y_1) 간의 덧셈 결과 (x_2, y_2) 는 다음과 같다.

$$x_2 = \lambda^2 - x_0 - 2x_0, \quad y_2 = \lambda(x_0 - x_2) - y_0$$

여기서 더하는 두 점이 같을 때, 즉 $(x_0, y_0) = (x_1, y_1)$ 이면 PntDbl로 $\lambda = \frac{3x_0^2 - a}{2y_0}$ 이다. 두 점이 다

를 때, 즉 PntAdd에서는 $\lambda = \frac{y_0 - y_1}{x_0 - x_1}$ 이다. 이

수식들은 나눗셈을 포함하고 있고 field 상에서의 나눗셈은 연산이 매우 복잡하다.

복잡한 나눗셈을 회피하기 위해 projective 좌표계가 많이 사용되기도 한다. 이는 Z 좌표를 더 추가하여 나눗셈을 수행하지 않고 나눌 값을 Z 에 누적되는 것으로 볼 수 있다. 이러한 방법은 나눗셈은 회피되나 다른 연산, 즉 곱셈이나 덧셈 등의 횟수가 더 늘어나며 최종 결과를 얻을 때는 나눗셈 또는 곱셈의 역 계산을 한번 이상 반드시 수행해야 한다는 단점이 있다.

기존의 x, y 좌표만 있는 좌표계를 affine 좌표계라고 부르며 affine 좌표계 상의 점 (x, y) 는 대표적인 projective 좌표계인 Jacobian 좌표계 상 점 $(X:Y:Z)$ 와의 관계식은 다음과 같다.

$$(x, y) = (X/Z^2, Y/Z^3)$$

Jacobian 좌표계 위의 점 $(X_0:Y_0:Z_0)$ 에 대한 PntDbl한 결과 $(X_2:Y_2:Z_2)$ 는 다음과 같이 계산된다.

$$X_2 = A^2 - 2B, \quad Y_2 = A(B - X_2) - 8Y_0^4, \quad Z_2 = 2Y_0Z_0$$

여기서 $A = 3X_0^2 + aZ_0^4$, $B = 4X_0Y_0^2$ 이다. Jacobian 좌표계 위의 점 $(X_0:Y_0:Z_0)$ 와 affine 좌표계 위의 점 (x_1, y_1) 과의 PntAdd한 결과 $(X_2:Y_2:Z_2)$ 는 다음과 같이 계산된다.

$$X_2 = A^2 - B^2C, \quad Y_2 = A(X_0B^2 - X_2) - Y_0B^3, \quad Z_2 = BZ_0$$

여기서 $A = y_1Z_0^3 - Y_0$, $B = x_1Z_0^2 - X_0$, $C = x_1Z_0^2 + X_0$ 이다.

	Affine 좌표계	Jacobian 좌표계
PntDbl	3M+D+9A	10M+10A
PntAdd	2M+D+6A	11M+8A

표 2. PntDbl, PntAdd에 필요한 연산 수

표 2는 Affine 좌표계와 Jacobian 좌표계 상에서의 PntDbl과 PntAdd에 필요한 연산 횟수를 나타낸다. 여기서 M, D, A는 각각 prime field 상에서의 곱셈, 나눗셈, 덧셈/뺄셈, 즉 modular multiplication (ModMult), modular division (ModDiv), modular addition/subtraction (ModAdd)을 나타낸다. Affine 좌표계와 Jacobian 좌표계 사용 시 필요한 연산 수를 비교해보면 상대적으로 계산이 간단하여 전

체 실행 시간에서 차지하는 비중에 매우 적은 ModAdd을 제외하였을 때 ModMult가 ModDiv에 비해 적어도 7~8배 이상 빨라야 Jacobian 좌표계를 사용하는 것이 속도 면에서 이득을 얻을 수 있다. 반대로 ModMult가 아주 빠르지 않으면 affine 좌표계 사용이 실행 속도 면에서 더 유리하며 수식도 더 간단해 구현 복잡도도 더 낮다는 장점이 있다.

III. 제안하는 PntMult 방법

표 2를 보면 affine 좌표계 사용 시 PntAdd가 PntDbl보다 하나 더 적은 ModMult을 요구한다는 것을 알 수 있다. 따라서 PntDbl을 PntAdd로 대체한다면 PntMult에 소모되는 총 실행 시간을 줄일 수 있다. 제안하는 PntMult 방법은 Algorithm 1에 나타나 있다.

Input: 타원 곡선 위의 점 P , k -bit scalar s
Output: sP
1: $(Q_0, Q_1) \leftarrow (2P, P)$
2: for ($i \leftarrow k-2$; $i \geq 0$; $i \leftarrow i-1$)
3: $Q_1 \leftarrow Q_0 + Q_1$
4: if ($s[i] = 1$) $Q_0 \leftarrow Q_1 + P$
5: else $Q_0 \leftarrow Q_1 - P$
6: return $Q_{s[0]}$

Algorithm 1. 제안하는 PntMult 방법

Algorithm 1을 보면 Q_0 을 초기화할 때에만 $2P$ 를 계산하기 위한 PntDbl이 계산된다. 그 이후에는 3번째 줄에서 PntAdd를, 4~5번째 줄에서는 $s[i]$ 값에 따라 PntAdd 또는 point subtraction (PntSub)이 수행된다.

	PntDbl	PntAdd	필요 modular 연산		
			M	D	A
DAA	$k-1$	$k-1$	$5k-5$	$2k-2$	$15k-15$
ML	k	$k-1$	$5k-2$	$2k-1$	$15k-6$
제안 방법	1	$2(k-1)$	$4k-1$	$2k-1$	$12k-3$

표 3은 각 PntMult 방법을 수행할 때 필요한 연산들을 나타낸다. PntSub은 빨 점의 y 좌표를 $-y$ 로 바꿔 PntAdd로 계산할 수 있으므로 PntAdd에 포함되었다. 표 3을 보면 필요한

ModDiv 수는 세 방법 모두 거의 비슷하나 ModMult, ModAdd의 경우 제안 방법이 DAA와 ML에 비해 약 4/5만 필요함을 알 수 있다.

IV. 결론

제안하는 PntMult 방법은 PntDbl을 PntAdd로 대체하여 기존의 SPA 방지를 위한 PntMult 방법보다 더 적은 실행 시간을 갖는다. 제안 방법의 효과는 affine 좌표계 사용할 때로 한정되기는 하나 저면적 ECC 하드웨어의 경우 ModMult의 속도를 빠르게 하기 어려워 affine 좌표계 사용이 유리하다. 따라서 affine 좌표계를 사용해야 하는 IoT 등의 경량 응용에서 더 적은 실행 시간으로 SPA 방지 효과를 얻을 수 있을 것으로 기대된다.

[참고문헌]

- [1] V.S.Miller, “Use of elliptic curves in cryptography,” in *Proc. Adv. Cryptology (Crypto)*, pp. 417–426, 1985.
- [2] N.Koblitz, “Elliptic curve cryptosystems,” *Math. of Comput.*, vol. 48, pp. 203–209, 1987.
- [3] P.Kocher, J.Jaffe, and B.Jun, “Differential power analysis,” in *Proc. CRYPTO*, Santa Barbara, CA, USA, 1999, pp. 388 – 397.
- [4] D.Hankerson, A.J.Menezes, and S.Vanstone, *Guide to elliptic curve cryptography*, New Your, NY, USA: Springer-Verlag, 2004.
- [5] M.Joye and S.-M.Yen, “The Montgomery powering ladder,” in *Proc. 4th Int. Workshop Cryptograph Hardw. Embedded Syst. (CHES)*, London, U.K.: Springer-Verlag, 2003, pp. 291 - 302.

부채널 내성 딥러닝 네트워크 동향

권혁동* 박재훈** 심민주** 서화정***

*한성대학교 정보컴퓨터공학과 (대학원생)

**한성대학교 IT융합공학부 (대학원생)

***한성대학교 IT융합공학부 (교수)

Technology trends of Side-Channel attacks resistance Deep Learning Network

Hyeok-Dong Kwon* Jae-Hoon Park** Min-Joo Sim** Hwa-Jeong Seo***

*Dept. of Information computer engineering, Hansung University
(Graduate student)

**Dept. of IT convergence engineering, Hansung University
(Graduate student)

***Dept. of IT convergence engineering, Hansung University
(Professor)

요약

부채널 공격은 암호 알고리즘을 자체적으로 공격하는 대신, 암호 알고리즘의 연산 과정에서 발생하는 전력 소모, 열과 같은 부가적인 정보를 사용한다. 이를 통해 암호화에 사용된 비밀키와 같은 중요 정보를 탈취할 수 있다. 이와 같이 부채널 공격은 알고리즘 자체에 대한 공격이 아닌, 연산을 진행하는 하드웨어에서 발생하는 부채널 정보를 사용하기 때문에 굉장히 위협적인 공격으로 알려져 있다. 부채널 공격은 딥러닝 네트워크에도 유효하며, 딥러닝 네트워크의 구성을 파악할 수 있다. 이를 통해 이미 존재하는 딥러닝 모델을 동일하게 복원하는 방법에 대해서도 활발하게 연구 중에 있다. 훈련이 완료된 딥러닝 네트워크는 일종의 자산으로도 취급될 수 있으므로, 부채널 공격을 방지하는 기법의 중요성이 대두되고 있다. 본 논문에서는 부채널 공격에 내성을 지니는 딥러닝 네트워크에 대한 최신 연구 동향에 대해서 확인해보며, 이후 딥러닝 네트워크가 지녀야 할 부채널 공격 내성에 대해서 알아본다.

I. 서론

인공지능의 발전은 다계층 뉴런을 활용하는 딥러닝 네트워크의 등장 이후로 빠른 속도로 발전하고 있다. 딥러닝 네트워크는 여러 가지 학습 방법에 따라서 자신만의 모델을 생성하게 되며, 이때 내부적으로 여러 매개변수가 확립된다. 일반적으로 이러한 매개변수는 정확히 알 수 없지만, 부채널 공격을 통해 매개변수를 확보하고 딥러닝 네트워크를 복원해내는 기법들이 제안되었다. 본 논문에서는 딥러닝 네트워크에 대한 부채널 공격 및 이를 방지하기 위한 부채널 공격에 내성을 지니는 딥러닝 네트워크 동향에 대해서 알아본다.

논문의 구성은 다음과 같다. 2장에서 부채널 공격 및 딥러닝 네트워크에 대한 부채널 공격에 대해 확인한다. 3장에서 부채널 공격에 내성을 지니는 딥러닝 네트워크 기법에 대한 동향을 확인한다. 4장에서 본 논문의 결론을 맺는다.

II. 부채널 공격과 딥러닝 네트워크

2.1 부채널 공격

부채널 공격(Side channel attacks)은 Kocher가 최초로 제안한 새로운 유형의 암호 알고리즘 공격 기법이다[1]. 부채널 공격은 암호

알고리즘의 취약점을 분석 및 공격하는 것이 아닌, 암호 알고리즘 연산 중 하드웨어 상에서 발생하는 각종 부채널 정보를 통해서 공격을 시도한다. 부채널 정보에는 전력 소모 정보, 연산 타이밍, 별열과 같은 정보들이 포함된다.

부채널 공격은 크게 그림 1과 같은 모습으로 표현할 수 있다. 하드웨어에서 암호 알고리즘 연산을 위해 비밀키와 평문을 입력하여 암호문을 생성할 때 발생하는 각종 부채널 정보를 수집하여 비밀 정보를 복구한다.

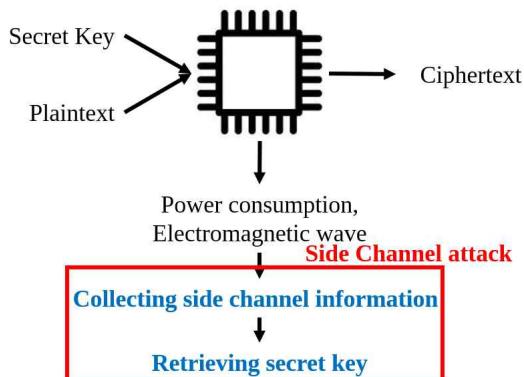


Fig. 1. Overview of side channel attacks.

대표적인 부채널 공격으로는 전력 소모량을 측정하여, 이를 대수적으로 분석하는 전력 분석 공격(Power analysis attack)[2]이 존재한다.

2.2 딥러닝 네트워크에 대한 부채널 공격

부채널 공격은 암호 알고리즘을 대상으로 공격을 시도하지만, 공격을 통해 비밀 정보를 복구해낸다는 특성을 활용하여 딥러닝 네트워크를 공격하는 방법이 제안되었다.

딥러닝 네트워크의 매개변수는 공개되지 않는 정보이지만, 부채널 공격을 통해서 복구될 수 있었다. [3]은 FPGA 기반의 CNN(Convolutional Neural Network)를 공격하여 최대 89%의 인식 정확도를 보였다. [4]는 8-bit AVR, 32-bit ARM 프로세서 상에서의 부채널 공격을 시도하였고, 최대 98.15%의 정확도를 보였다. [5]는 Arduino Uno 보드 상에서 활성화 함수를 복원하는데 성공하였다.

III. 부채널 내성 딥러닝 네트워크

부채널 공격을 통해서 딥러닝 네트워크 분석을 방지하기 위해서는, 각각의 공격에 효과적인 내성 기법이 필요하다. 딥러닝 네트워크에 대해 부채널 공격을 제안한 연구자들은, 기법에 대한 방어 방법도 제안하였고 이는 다음과 같은 기법들이 있다.

우선은 각각의 연산 순서를 바꾸는 셔플링 기법이 있다[6]. MLP(Multi-Layer Perceptron)의 hidden layer 실행은 순서대로(Sequential)하게 실행되어야 하나, 각각의 레이어 내부의 연산은 독립적으로 실행될 수 있다. 즉, 레이어 내부의 곱셈 연산을 무작위 순서로 수정하는 것으로 DPA(Differential Power Analysis) 공격을 약화시킬 수 있다.

또 다른 방법으로 마스킹 기법이 있다[7]. 마스킹 기법은 부채널 공격에서 자주 사용되는 방어 기법으로, 연산 값에 임의의 마스킹 값을 포함시켜서 실제 값과 부채널 정보 사이의 종속성을 제거한다. 마스킹 기법은 다른 기법에 비해 구현 난이도가 낮은 편이며 단순하지만, 알고리즘이 효과적으로 부채널 내성을 지니게 해준다.

[3]에서 가중치(Weight)는 CPA(Correlation Power Analysis)를 통해서 복구해냈다. 마스킹을 사용하게 되면, 부채널 정보와 실제 비밀 정보의 종속성이 사라지므로 제대로된 상관 관계 연산이 어려워진다. 따라서 CPA에 대해 효과적으로 방어할 수 있다.

타이밍 공격을 방어하기 위해서는 constant timing 구현을 통해 방어할 수 있다. [3]과 [5]에서 활성화 함수를 복원하는데 있어서 타이밍 공격을 시도하였다. constant timing 구현은 연산 시간을 일정하게 만드는 것으로, 부채널 정보 중 시간 정보를 무력화할 수 있다. 활성화 함수에 따라 구현 방법은 조금씩 차이가 있다. 지수 연산을 사용하는 활성화 함수는 [8]과 같은 공개키 알고리즘의 constant timing 구현 기법을 사용할 수 있다. 입력 값에 따라 연산 시간이 일정한 함수도 constant timing 구현이 용이하다. ReLU 함수 같은 경우, 입력 값이 0 이전일 때와 0 이후일 때만 연산 시간이 다르기 때문에 구현이 편리하다[9].

IV. 결론

본 논문에서는 부채널 내성을 지니는 딥러닝 네트워크 구현을 위한 기법에 대해서 확인하였다. 주된 기법으로는 셔플링, 마스킹, 그리고 constant timing 구현과 같은 기법을 확인할 수 있었다. 이러한 부채널 방지 기법은 일반적으로 암호 알고리즘 구현에 사용되지만, 딥러닝 네트워크의 복원을 방지하는데도 적용할 수 있음을 알 수 있었다.

딥러닝 네트워크의 모델은 개인, 기업의 자산으로 활용될 수 있으므로, 부채널 공격을 통한 네트워크의 복원은 금전적인 위협으로도 다가올 수 있다. 따라서 지속적인 부채널 위협에서 네트워크를 보호할 수 있도록 예방 방법에 대한 연구가 계속적으로 이루어져야 한다.

V. Acknowledgment

이 성과는 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1F1A1048478, 100%).

[참고문헌]

- [1] P.C.Kocher, “Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems,” In *Annual International Cryptology Conference*, Springer, Berlin, Heidelberg, pp. 104–113, 1996.
- [2] P.Kocher, J.Jaffe, and B.Jun, “Differential power analysis,” In *Annual international cryptology conference*, Springer, Berlin, Heidelberg, pp. 388–397, August, 1999.
- [3] L.Wei, B.Luo, Y.Liu, and Q.Xu, “Wei, L., Luo, B., Li, Y., Liu, Y., & Xu, Q. (2018, December). I know what you see: Power side-channel attack on convolutional neural network accelerators,” In *Proceedings of the 34th Annual Computer Security Applications Conference*, pp 393–406, December, 2018.
- [4] L.Batina, S.Bhasin, D.Jap, and S.Picek, “CSI NN: Reverse Engineering of Neural Network Architectures Through Electromagnetic Side Channel,” In *28th USENIX Security Symposium*, pp. 515–532, 2019.
- [5] G.Takatoi, T.Sugawara, K.Sakiyama, and Y.Li, “Simple Electromagnetic Analysis Against Activation Functions of Deep Neural Networks,” In *International Conference on Applied Cryptography and Network Security*, pp. 181–197, Springer, Cham, 2020.
- [6] N.Veyrat-Charvillon, M.Medwed, S.Kerckhof, and F.X.Standaert, “Shuffling against side-channel attacks: A comprehensive study with cautionary note,” In *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 740–757, Springer, Berlin, Heidelberg, December, 2012.
- [7] J.S.Coron, and L.Goubin, “On boolean and arithmetic masking against differential power analysis,” In *International Workshop on Cryptographic Hardware and Embedded Systems(CHEs)*, pp. 231–237, Springer, Berlin, Heidelberg, August, 2000.
- [8] G.Hachez, and J.J.Quisquater, “Montgomery exponentiation with no final subtractions: Improved results,” In *International Workshop on Cryptographic Hardware and Embedded Systems(CHEs)*, pp. 293–301, Springer, Berlin, Heidelberg, August, 2000.
- [9] T.Nakai, D.Suzuki, F.Omatsu, and T.Fujino, “Evaluation of timing attacks against deep learning on a microcontroller and countermeasures,” In *2020 Symposium on Cryptography and Information Security(SCIS)*, pp. 28–31, Kochi, Japan, January, 2020.

Quantum Information Set Decoding 연구 동향

장경배* 송경주* 오유진** 서화정***

*한성대학교 (대학원생)

**한성대학교 (학생)

***한성대학교 (교수)

Quantum Information Set Decoding Research Trends

Kyung-Bae Jang* Gyeong-Ju Song* Yu-Jin Oh** Hwa-Jeong Seo*

*Korea University(Graduate student)

**Korea University(Undergraduate student)

***Korea University(Professor)

요약

IBM, Google, Microsoft 등의 국제 대기업의 양자 컴퓨터에 대한 전폭적인 투자는 양자 컴퓨팅 시대를 앞당기고 있다. Shor 알고리즘을 구동시킬 수 있는 고수준의 양자 컴퓨터가 개발된다면, 현재 널리 사용되고 있는 공개키 암호 시스템 RSA와 ECC는 더 이상 사용할 수 없다. 이에 NIST(National Institute of Standards and Technology)에서는 RSA와 ECC(Elliptic Curve Cryptography)를 대체할 수 있는 양자내성암호 표준화 공모전을 주최하였다. 현재 Finalists가 선정된 상태이며 격자기반, 코드기반, 다변수기반 암호로 구성되어 있다. 본 논문에서는 코드기반 암호에 가장 효율적인 공격법으로 알려진 Information Set Decoding에 대한 연구 동향을 살펴본다.

I. 서론

양자 알고리즘을 사용하는 양자 컴퓨터는 암호 시스템들이 기반하고 있는 난제를 빠르게 해결할 수 있다. 대표적으로 Grover 알고리즘[1]과 Shor 알고리즘[2]이 암호학계의 안전성을 위협하고 있다. Grover 알고리즘은 전수조사를 가속화하여 n -bit 보안 레벨의 대칭키 암호를 $n/2$ -bit 보안 레벨로 낮출 수 있다. 따라서 기존의 보안 레벨을 만족시키기 위해서는 키 길이를 2 배로 늘려야 한다. 대칭키 암호와는 달리 공개키 암호의 경우 양자 컴퓨터의 파급력이 더 강력하다. Shor 알고리즘은 공개키 암호 RSA와 ECC에서 기반하고 있는 소인수 분해 문제와 이산대수 문제를 다행시간 내에 해결할 수 있다. 대칭키 암호와는 달리 안전성이 완전히 무너지기 때문에 이를 대체할 새로운 양자내성암호가 필요한 상황이다. 이에 NIST는 RSA와

ECC를 대체할 수 있는 암호 알고리즘을 선정하기 위한 양자내성암호 공모전을 주최하였다. 양자 컴퓨터에서도 풀기 어려울 것으로 여겨지는 문제들을 기반으로 한 암호 알고리즘들이 공모전에 참가하였으며 현재 Round 3의 Finalists에는 격자기반, 코드기반, 다변수기반의 암호 알고리즘들이 남아있는 상태이다. 본 논문에서는 코드기반암호에 대한 연구 동향 및 코드기반암호에 대해 가장 효율적인 공격 알고리즘인 Information Set Decoding에 대해 살펴보자 한다.

II. 관련연구

2.1 코드기반암호

코드기반암호는 공개키의 크기가 너무 크다는 단점을 가지고 있어 주목받지 못하였지만 양자 컴퓨터 연구가 활발히 진행되면서 다시 주목을 받고 있다. 격자기반, 다변수기반 암호와 함께

양자 컴퓨터의 등장 이후에도 안전성을 보장받을 수 있을 것으로 신뢰되기 때문이다.

2.2. ClassicMcEliece

ClassicMcEliece는 송수신자간 동일한 세션 키를 설립하는 KEM 구조이다. Classic McEliece가 기반하고 있는 안전성은 아래의 신드롬 디코딩 문제이다.

$$He^T = C \quad (1)$$

ClassicMcEliece는 Goppa 코드로부터 생성한 패리티체크행렬 H 를 공개키로 사용한다. 송신자는 특정 무게 조건을 가지고 있는 벡터 e 를 생성한 뒤, 공개키 H 와 행렬 곱을 수행한다. 이렇게 생성된 신드롬 값 C 는 암호문 역할을 한다. 행렬 H 와 암호문 C 는 공개되지만 낮은 무게 조건을 특징으로 하는 벡터 e 를 복구하는 것은 NP-hard 문제로 분류된다. 올바른 수신자는 Goppa 코드로부터 공개키 H 를 생성할 때 사용 된 개인키로 암호문 C 를 디코딩하여 원본 e 를 복구하고 해시함수를 거쳐 송수신자 간 동일한 세션 키를 설립할 수 있다.

2.3. Information Set Decoding(ISD)

ISD는 개인키를 복구하는 과정 없이 암호문에서 비밀정보인 벡터 e 를 바로 찾아내는 공격 알고리즘이다. 1962년, Prange가 가장 기본적인 ISD를 제안하였다[3]. Prange의 ISD로부터 Lee-brickell, Stern등 다양한 ISD가 제시되었지만 성능의 개선은 조금씩 있었지만 획기적으로 복잡도를 줄이진 못하였다. 알고리즘에 따라 접근 방식이 다르긴 하지만 ISD는 신드롬 값을 생성함과 동시에 특정 무게조건을 만족하는 벡터 e 를 찾기 위한 전수조사를 기본적으로 수행한다.

2.4 Grover 탐색 알고리즘

Grover 탐색 알고리즘은 대칭키 암호 키에 대한 전수조사 및 해시 함수의 pre-image 공격을 가속화 할 수 있는 양자 알고리즘이다. 양자 컴퓨터는 중첩 상태의 큐비트를 사용하는데, 이를 활용하여 n -qubit은 2^n 개의 상태를 동시에 확률로서 표현할 수 있다. Grover 탐색 알고리즘은 Oracle과 Diffusion operator로 구성되며, Oracle은 해답을 반환하고 Diffusion operaor는 반환된 해답의 관측 확률을 증가시킨다. 고전

컴퓨터의 n -bit 값 중, 특정 해답을 찾기 위해 $O(2^n)$ 번의 탐색이 필요하다면 양자 컴퓨터에서 n -qubit은 모든 경우를 표현할 수 있고 Oracle과 Diffusion operator를 약 $2^{n/2}$ 번 반복하여 해답의 확률을 충분히 높이고 마지막에 높은 확률로 해답을 찾아낼 수 있다.

III. Quantum Information Set Decoding

Overbeckr, Sendrier은 Information set decoding에 Grover 알고리즘을 적용했을 때 성능 개선이 크게 이루어지지 않음을 제시하였다 [4]. 대칭키 암호에 대해 키를 찾을 때 Grover 알고리즘을 적용한다면 복잡도를 절반으로 줄일 수 있지만 코드기반 암호의 신드롬 계산 문제에는 Grover 알고리즘을 적용하여도 완벽히 적용될 수 없다는 것이다. n -bit 키의 경우 2^n 개의 경우의 수가 존재한다. Grover 탐색은 2^n 의 후보군 중 해답을 찾아내야하고 전수조사의 경우 $O(2^n)$ 번의 탐색이 필요한 것을 약 $2^{n/2}$ 번의 탐색으로 복잡도를 절반으로 줄일 수 있다. 하지만 신드롬 계산의 경우, n -bit 벡터에서 특정 무게 조건을 고려하기 때문에 후보군이 좁혀지고, 2^n 의 후보군에서 Grover 알고리즘으로 복잡도를 줄이는 것은 비효율적인 것이다. Table 1.은 [4]에서 고전적인 ISD와 Grover search를 적용한 ISD의 복잡도를 비교한 결과이다.

Table 1. Attack Complexity Comparison

McEliece Parameters m, t	Classical	Quantum
11, 32	2^{91}	2^{86}
11, 40	2^{98}	2^{94}
12, 22	2^{93}	2^{87}
12, 45	2^{140}	2^{133}

하지만 Bernstein은 코드기반암호에 Grover 알고리즘이 완벽히 적용될 수 있음을 제시하였다[5]. [4]에서는 Grover 알고리즘과는 별개로 벡터의 무게를 추정하는 복잡도를 고려하였지만 Bernstein은 Grover 알고리즘의 입력 단계에서 무게 조건을 같이 고려함으로써 고전적인 ISD에 Grover 알고리즘이 완벽히 적용되어 복

잡도를 절반으로 줄일 수 있음을 보였다. Grover 알고리즘에서 무게 조건을 고려하는 중첩 상태의 입력을 준비하는 회로는 Figure 1과 같다.

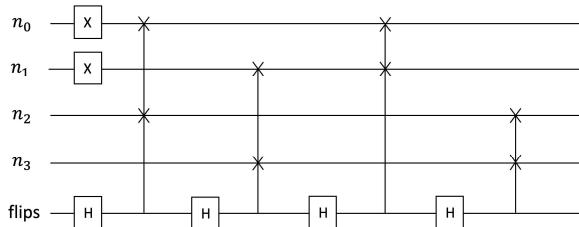


Figure 1. superposition state input (weight=2)

해당 회로는 대상 n -qubit에 Hadamard 게이트를 통해 중첩 상태를 준비하지 않는다. X 게이트로 벡터의 무게를 설정하고 flips 큐비트에 Hadamard 게이트를 활용하여 중첩 상태의 입력을 만든다. 이를 통해 후보 군을 우리가 탐색해야 하는 집합으로 줄인 뒤, Grover 탐색을 수행한다. Grover 탐색은 후보군 중에서 해답의 관측 확률을 높이기 위해 반복하는데, 이 횟수는 후보군의 크기에 따라 결정된다. 입력 상태에서 무게 조건은 이미 맞추고 시작하기 때문에 신드롬 값을 생성하는 벡터를 찾는 Grover 탐색을 수행함으로써 기존 ISD의 복잡도를 절반으로 줄일 수 있다. 공개키인 패리티 체크 행렬은 이미 알고 있으며 이에 따라 CNOT 게이트들만을 사용하여 신드롬 계산 양자 회로를 구성하는 것은 비교적 간단하다. 고전적인 ISD에서는 Prange, Lee-brickell, Stern 순으로 효율성이 조금씩 개선되었지만 양자 ISD를 고려할 때는 구현에 필요한 자원을 고려해야 한다. [3]에서도 Stern의 ISD가 고전 컴퓨터에서는 더 효율적일 수 있지만 양자 컴퓨터상에서는 Prange의 알고리즘이 더 적은 양자 자원으로 구현될 수 있음을 강조하였다.

IV. 결론

본 논문에서는 양자내성암호인 코드기반암호에 대한 공격 알고리즘 ISD 동향에 대해 조사하였다. 초기 연구에는 ISD에 양자 알고리즘인 Grover 알고리즘이 완전히 적용될 수 없었지만, Bernstein은 Grover 알고리즘이 온전히 ISD의 복잡도를 절반으로 줄일 수 있음을 보였다. 코드기반암호는 발전하는 양자 컴퓨터와 양자 알고리

즘의 시대에서 양자내성암호의 역할을 할 수 있을 것이라 여겨지고 있다. 그리고 이 안전성이 계속 보장되는지에 대해서는 ISD 뿐만 아니라 새로운 양자 알고리즘의 개발 동향을 주시할 필요가 있다.

V. Acknowledgment

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (<Q|Crypton>, No.2019-0-00033, 미래컴퓨팅 환경에 대비한 계산 복잡도 기반 암호 안전성 검증 기술개발, 100%).

[참고문헌]

- [1] Grover, L.K, “A fast quantum mechanical algorithm for database search,” in Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, pp.212 - 219, 1996.
- [2] Peter W. Shor, “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer” SIAM review, Vol. 41. No. 2. 303-332. 1999
- [3] Prange. “The use of information sets in decoding cyclic codes” IRE Transactions. IT-8. S5-S9. 1962
- [4] Raphael Overbeck, Nicolas Sendrier, “Code-based cryptography”, Post-Quantum Cryptography, pp. 95-145, 2009
- [5] Daniel J. Bernstein, “Grover vs. McEliece” in PQCrypto 2010. pp. 73 - 80. 2010.

블록체인-IoT-클라우드 기반 수산 식품 공급망 모델 제안

전미현*, 조강우*, 신상욱**

*부경대학교 대학원 정보보호학과 (대학원생)

**부경대학교 IT융합응용공학과 (교수)

Proposal Blockchain-IoT-Cloud Based Model for Fisheries Food Supply Chain

Mi Hyeon Jeon*, Kang Woo Cho*, Sang Uk Shin**

*Dept. of Information Security, Graduated School,

Pukyong National University (Graduate Student)

**Dept. of IT Convergence and Application Eng.,

Pukyong National University (Professor)

요약

수산 식품은 다른 식품들보다 식품 안전사고가 많이 발생할 수 있는 신선 식품 중 하나지만 수산 식품 공급망 분야는 발전이 매우 더딘 편이다. 식품 안전사고를 예방 및 대처하기 위해서는 공급망 엔터티 간의 정보의 공유와 추적성이 보장되어야 하기 때문에 블록체인 기반 공급망 모델에 관한 연구가 활발히 진행되고 있지만, 수산 식품 공급망에 관한 블록체인 기술 적용의 연구는 매우 미흡하다. 블록체인 기술은 불변성, 입증성, 추적성을 제공할 수 있어 식품 공급망에 좋은 선택이 될 수 있지만, 성능과 프라이버시 보호 부분에 문제가 발생할 수 있다. 본 논문에서는 IoT-클라우드 기술과 블록체인의 연동, 그리고 암호학적 기술들을 적용하여 문제점들을 극복하고, 또한 목적으로 별도의 원장을 두어 정보 신뢰성과 프라이버시 보호 및 추적성을 제공하여 안전하고 신뢰성 있는 수산 식품 공급망 모델을 제안한다.

I. 서론

현재 우리가 살고 있는 빅데이터 시대에는 전 세계적으로 데이터들이 폭발적으로 발생하고 있고, 이러한 데이터들을 적절하게 사용한다면 여러 분야에서 혜택을 볼 수 있다. 식품 공급망 분야 역시 데이터들을 필요 그룹들끼리 공유하고 활용한다면 식품 안전사고를 예방할 수 있고 추적하는 데에도 사용할 수 있다. 하지만 기존의 식품 공급망에서 공급망 엔터티 간의 데이터 공유가 원활히 진해되지 않았고 데이터의 흐름이 제대로 저장·공유 되지 않았기 때문에 말고기 스캔들[1], 파파야 살모넬라균 사건 등

이 발생하였다.

위와 같은 문제들은 농산물, 수산물, 축산물 등 모든 식품 분야에서 나타날 수 있지만, 신선 식품인 수산물, 축산물 분야에 매우 중요하다. 하지만 현재 수산 식품 공급망에 4차 산업혁명의 대표 기술을 접목시킨 사례가 거의 없다. 수산 식품에 대한 데이터 수집은 생산자(위치, 온도, 종, 보관상태 등), 가공자(공정 과정, 식품명, 공정 기계 등), 운송자(위치, 온도, 습도, 보관상태 등), 판매자(위치, 가격, 보관상태 등)로부터 가능하고, 수집된 데이터들은 이후에 발생할 수 있는 안전사고에 대한 추적에 사용될 수 있다.

데이터 저장을 위해 식품 공급망에 점차 클라우드 기술을 채택하는 공급망 기업들이 늘어나고 있다. 클라우드는 Pay On Demand, 높은 탄력성 특성을 갖고 있기 때문에 식품 공급망에 효율적이지만, 중앙 집중화된 서버를 갖고 있어 SPoF(Single Point of Failure)문제, 병목현상이 발생할 수 있고, 또한 보안 및 프라이버시 문제도 발생할 수도 있다.

블록체인은 수학적 알고리즘을 기반으로 하는 기술로써 변조 방지, 탈중앙화 특성을 갖고 있다. 블록체인과 클라우드 기술을 융합해 중앙 집중형 모델에서 발생할 수 있는 문제들을 해결하고, 변조 방지 특성을 이용해 데이터의 신뢰성을 높일 수 있다. 따라서 본 논문에서는 블록체인과 클라우드 기술을 공급망 분야에 접목시키는 모델을 제안한다.

II. 관련연구

블록체인 기술을 기반으로 하는 공급망에는 여러 사례들이 존재한다. Malik 등이 제안한 모델인 TrustChain은 블록체인에 업로드 된 데이터의 신뢰성을 높이기 위해 평판 및 신뢰 모델을 추가했다. 정직하게 거래하는 기업에게는 보상을 줌으로써 계속 정직하게 거래하도록 유인하고, 정직하지 않게 거래한 기업들에게는 특정 기간 동안 네트워크에 참여하지 못하도록 함으로써 기업에 불이익을 주는 방식으로 데이터에 신뢰성을 부여한다[2]. 또 다른 제안 모델인 TradeChain에서는 블록체인에 게재되는 정보들이 모든 사람에게 보여 지기 때문에 기업의 영업비밀이 노출될 수 있어 Malik 등은 프라이버시 보호를 위해 두 개의 별도 대장을 두어 공급망 거래자의 정체성과 거래 이벤트를 분리한다. 이로써 자격이 있는 참여자만 해당 거래 이벤트를 확인할 수 있게 해 중요 정보들을 보호 할 수 있다[3]. 마지막으로 Tao 등은 클라우드 기술과 블록체인 기술을 융합하여 기존의 클라우드 기반 공급망 모델들의 문제점을 해결하는 모델은 제안하였다[4]. 또한 블록체인의 성능적인 이유로 공급망과의 융합이 불가능했지만 클라우드 기술을 이용함으로써 이 문제도 해결하였다.

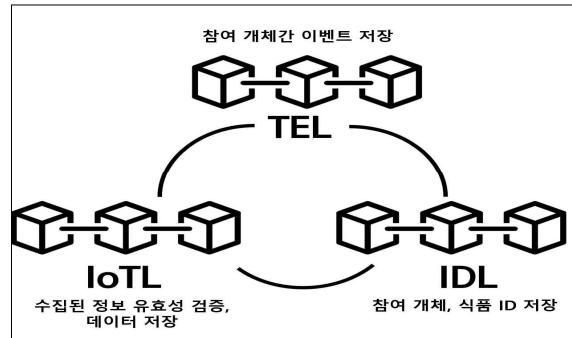


Fig 1. 목적에 따른 별도의 원장들

III. 블록체인-IoT-클라우드 기반 식품 공급망 모델

3.1 모델 제안

본 논문에서는 기존의 공급망 시스템에 IoT 센서와 클라우드, 블록체인 기술을 융합해 블록체인의 성능을 높이면서 또한 수집된 데이터의 신뢰성을 높이는 모델을 제안한다. Fig 1에서 볼 수 있듯이, 제안 모델은 수집된 데이터의 유효성을 검증하고 저장하는 IoT(Ledger)과 공급망 참여 개체 간의 이벤트 트랜잭션들이 저장되는 TEL(Transaction Event Ledger), 공급망 내의 식품들과 참여 개체들의 ID를 저장하는 IDL(Identification Ledger) 총 3가지의 블록체인이 사용되고 있다. 각각의 원장에서 블록을 생성할 수 있는 합의 노드들은 규제/감독기관에서 지정한다. IoT는 Fig 2에 표현되어 있는 엔티티들로부터 주기적으로 수집한 데이터들(위치 정보, 온도, 습도 등)이 유효한지 검증하고 블록체인에 트랜잭션으로 게시한다. 이때 실제 정보는 용량 등의 문제로 암호화 된 후 클라우드에 저장이 되며 트랜잭션에는 해시 값 등을 포함한 메타 데이터가 포함된다. TEL 역시 공급망 참여 개체 간에 발생하는 거래들이 블록체인에 게시되고, 실제 정보는 역시 IoT와 마찬가지로 클라우드에 저장된다. TEL에는 IoT센서로 수집한 데이터들의 모든 정보를 포함하지 않고 정해둔 표준에 따라 기준 값의 만족 여부와 같은 일부 정보만 포함한다. IDL은 IoT와 TEL에 참여하는 엔티티들의 ID와 공급망 참여 엔티티 ID, Fig 2에서

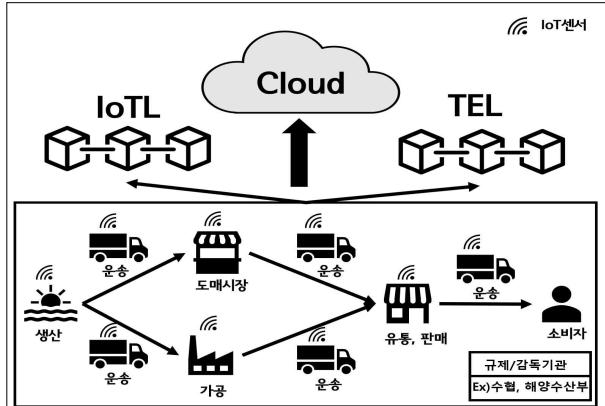


Fig. 2. 블록체인-IoT-클라우드 기반 공급망 제안 모델

보여 지는 과정으로 전달되는 식품ID가 저장되는 원장이다. 이 3개의 원장들은 각각의 목적에 따라 구분되지만 다른 원장들과는 연결되어 있다.

3.2 모델 분석

본 논문에서 제안한 모델은 기존의 중앙 집중형 공급망이 가지고 있던 단점들을 보완하는 특징들을 지닌다. (1) 데이터의 신뢰성. 중앙 집중형 공급망에서 수집된 데이터는 변조 및 탈취될 가능성이 존재했다. 하지만 제안 모델에서는 IoT센서로부터 수집된 데이터가 IoT의 참여자들로부터 검증된 후 블록체인에 게시되기 때문에 데이터의 신뢰성을 제공할 수 있고, 또한 블록체인의 특성인 불변성으로 변조의 위험을 줄일 수 있다. (2) 공급망과 결합되었을 때의 더 나아진 블록체인의 성능. [2], [3]에서 제안된 모델들은 블록체인과 공급망과의 결합이기 때문에 느린 처리속도, 용량 등과 같은 블록체인의 단점을 해결하기 어렵다. 하지만 본 논문에서의 모델은 클라우드를 사용함으로써 높은 저장 공간을 제공할 수 있어 문제를 해결할 수 있게 되었다. (3) 추적성. 블록체인을 기반으로 하는 공급망은 블록체인의 특성인 불변성을 이용해 쉽게 추적 가능해진다. 본 제안모델에서는 IDL에 저장되어있는 식품 ID를 기반으로 사고에 대해 추적할 수 있다.

IV. 결론

본 논문에서는 기존 중앙 집중형 식품 공급망에 4차 산업혁명 기술인 클라우드와 IoT, 블록체인 기술들을 접목시켜 기존 식품 공급망에 존재했던 문제점들의 해결안을 제시한다. 3개의 원장, 각 단계에 존재하는 IoT센서, 수집된 데이터의 저장을 담당하는 클라우드 기술로 제안된 모델은 데이터의 신뢰성, 추적성의 특징을 갖고 있어 안전사고에 대한 충분한 해결 방안이 되지만, 구체적인 트랜잭션 구조, 암호학적 메커니즘의 적용 등에 관한 추가 연구가 필요하며, 이를 통해 수산 식품 안전사고를 예방 및 추적하는데 활용될 수 있을 것이다.

Acknowledgement

이 논문은 2021년 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구입니다 (과제명: 미래수산식품 연구센터).

[참고문헌]

- [1] J. Premanandh, "Horse meat scandal - A wake-up call for regulatory authorities", Food control 34.2, p. 568–569, 2013.
- [2] S. Malik, V. Dedeoglu, S. Kanhere and R. JulDak, "TrustChain: Trust Management in Blockchain and IoT Supported Supply Chains.", 2019 IEEE International Conference on Blockchain(Blockchain), p. 184–193, 2019.
- [3] S. Malik, N. Gupta, V. Dedeoglu, S. Kanhere and R. JulDak, "TradeChain: Decoupling Traceability and Identity inBlockchain enabled Supply Chains.", arXiv preprint arXiv:2105.11217, 2021.
- [4] Q. Tao, Q. Chen, H. Ding, I. Adnan, X. Huang and X. Cui, "Cross-Department Secures Data Sharing in Food Industry via Blockchain-Cloud Fusion Scheme.", Security and Communication Networks 2021.

TOR 브라우저에서 비트코인을 사용한 전략물자 불법 거래 내역 추적 방안 연구*

선하라*, 윤지원**

*고려대학교 (대학원생), **고려대학교 (교수)

A Study on the Tracking Method of Illegal Transactions of Strategic Materials Using Bitcoin in the The Onion Browser

Ha-Ra Seon*, Ji-Won Yoon**

*Korea University(Graduate student), **Korea University(Professor)

요약

TOR(The Onion Router) 브라우저는 오늘날까지 익명성을 갖춘 브라우저로 대중들에게 알려져 있으며, 불법 거래가 창궐하는 다크웹이 산재하고 있다. 이에 따라 법적으로 무역이 제한되는 전략물자가 TOR 브라우저의 다크웹에서 거래되고 있다. 본 논문에서는 불법 전략물자 거래를 비트코인으로 수행했을 경우 거래 내역을 추적하는 방법을 제안한다. 본 연구 목적을 달성하기 위해 크롤링 및 CNN(Convolutional Neural Network) 방식으로 불법 전략물자 거래를 탐지한 후, Dusting 공격 기법 및 Passive Finger Printing을 활용하여 거래 발생 위치 및 거래 내역을 추적하여 전략물자 불법 거래 근절에 이바지한다.

I. 서론

오늘날, 수많은 다크웹에서 비트코인으로 불법 거래가 진행되고 있다. 특히, 매매가 금지된 불법적 물품을 구매 후 비트코인 등 현물 화폐보다 익명성이 높은 화폐를 거래 수단으로 활용하여 법의 옮가미를 피하는 경우가 꾸준히 증가하고 있다. 법적으로 승인된 경우를 제외하고 매매를 금지하는 전략물자 또한 TOR 브라우저에서 매매되고 있다. 그러나 비트코인 믹서, 셔플링 등 자금의 위법한 출처를 숨길 수 있는 기술이 발달함에 따라 범법자들을 징벌할 만한 명확한 증거를 찾기 쉽지 않다. 코로나바이러스로 재택근무가 늘어남에 따라 기업 내부망에 침투를 위해 다크웹을 통해 민감정보를 구매하여 개인 및 기업을 공격하는 사례가 늘어나고 있다고 한다.

본 논문에서는 다크웹을 통해 전략물자를 다

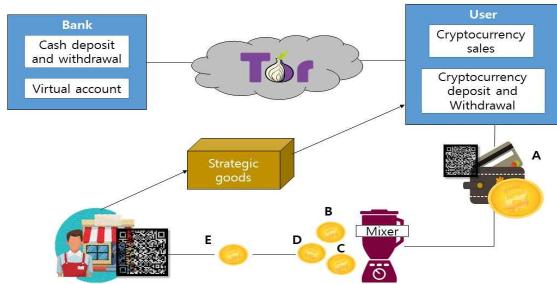
루는 기업이나 개인이 전략 물자관리원에 신고 없이 매매하고 자신의 범법 행위를 은폐하고자 암호화폐인 비트코인으로 거래하는 범법 행위를 탐지하는 방안을 연구하고자 한다.

II. 이론적 배경

TOR(Then Onion Router)는 토르 브라우저라고도 불리며, 네트워크 우회와 익명성이 특징이다. 패킷 전송 시 여러 국가에 거쳐서 릴레이 호스트를 지나가기 때문에 느린다. 하지만 패킷 간 암호화를 지원하기 때문에 민감 정보를 위한 통신 도구로 많이 사용된다. 왜냐하면 출구 릴레이 호스트에서 복호화를 실행하기 때문에 릴레이 호스트를 조사해도 사용자의 IP를 확인 할 수 없기 때문이다.

토르 브라우저에서 주로 일어나는 불법 거래들은 다음과 같은 방식으로 발생한다. 온라인상에서 사용자들은 은행에 보관한 현금을 거래소를 통해 비트코인으로 바꾸고 비트코인을 토르 브라우저에 있는 비트코인 믹싱 월렛(Mixing wallet)이나 셔플 월렛(Shuffle wallet) 사이트 같은 여러 사용자에게 비트코인을 받아 믹싱

* 전략물자관리원의 연구 용역과제 “오픈 소스 정보를 활용한 우려거래 타겟팅 방안”에 지원 받아 수행한 과제임



[그림 1] TOR 브라우저에서 전략물자 불법거래 과정

알고리즘으로 잘게 분할하여 정해진 양만큼 분할함으로써 출처를 불분명하게 만드는 서비스를 제공해주는 곳에 보낸다. 그 후 그 월렛에 있는 비트코인으로 구입하고자 하는 불법 매매품을 다크웹에서 비트코인 지갑 주소로 비트코인을 보냄으로써 구매할 수 있다.

비트코인이란 탈중앙성 분산성, P2P 기술을 활용한 해시 함수기반 암호화폐이다. 비트코인은 P2P 거래에서 거래 당사자 간 신원 증명과정 없이 해시값으로 이루어진 거래 주소를 통해 거래가 진행된다. 진행된 거래는 새로운 블록으로 생성되어 모든 참여자에게 전송되어 공유된다. 이때 블록 거래 내역은 블록체인 참여자 모두 볼 수 있지만, 해당 주소가 누구의 소유인지 알 수 없다. 왜냐하면 비트코인에서 소유주소는 공개키의 해시값이며, 제한 없이 공개 키 쌍을 생성할 수 있기 때문에 익명성을 보장한다.

비트코인 믹서란[2] 개인정보를 보호하기 위한 목적으로 사용자가 다른 사용자와 코인을 혼합하여 비트코인을 세탁하는 것이다. 중앙집중식 믹싱 서비스와 탈중앙식 믹싱 서비스로 나뉜다. 비트코인 거래소에서는 사용자의 정보가 있기에 비트코인 지갑 주소가 해시값이라고 할지라도 개인을 식별할 수 있다. 따라서 믹서를 사용하게 되면 자신의 비트코인을 다른 사람의 비트코인과 바꾸거나 작은 단위의 비트코인을 대량의 사람에게서 받아서 출처를 알 수 없게 하는 방법 등이 있다. 일부 믹싱 서비스 웹사이트에서는 토르 네트워크를 통해 직접 접근하여 사용자와 서비스 운영자 모두 IP 주소를

노출 시킬 필요 없는 경우도 있다.[4]

III. 관련 연구

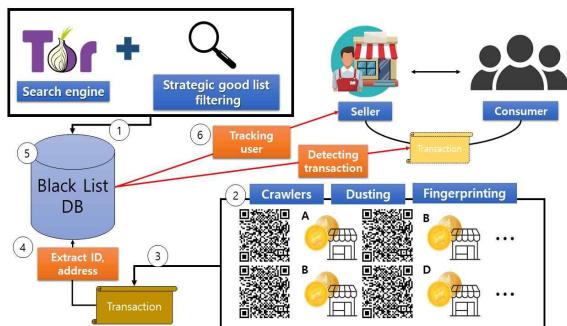
첫째로, TaoPu와 JuanWang이 제안한 CNN 기법을 이용한 평거 프린팅(Finger printing)을 활용한 익명 통신 추적 방법[1]은 범죄자가 악용하는 익명 통신 시스템을 감시하기 어려운 점을 노려 1차원 컨볼루션 신경망을 기반으로 한 TOR 익명 교통특징 추출 및 식별방법을 제안한다. 이 방법은 기능 추출 및 예측 분류를 통합하며, CNN(Convolutional Neural Network) 모델을 사용하여 TOR 익명 트래픽에 대한 지문 인식을 달성합니다. 실험 데이터 집합에서 87.5%의 정확도를 달성했다. TOR 익명 트래픽 및 익명 통신을 효과적으로 식별할 수 있다. 네트워크는 감독 및 검토를 하는 역할을 한다. 이 글은 또한 Naive Bayes, Random Forest 및 다른 방법들을 비교함으로써 1차원 신경망 방법의 발전과 다용성을 보여준다. 실험 결과에 따르면 데이터 세트 수가 클수록 인식 정확도가 높아진다.

둘째로, Perri Reynolds의 연구[3]에 의하면 자금세탁 방지 법률과 고객 식별 보안 표준은 온라인 플랫폼과 가상화폐 시스템에서 사용할 수 있도록 정립돼 있지 않다. 사용자의 공개 키가 거래 내역을 통해 추적될 수 있지만 계정과 관련된 전자 메일 주소 등록과 같은 다른 신원 데이터 확인 요건이 수반되지 않는 한 여전히 익명으로 남아 있다. 그 결과, 기술적으로 숙련된 범죄자들은 신원을 밝히지 않고도 신원 관리를 우회하고 거래할 수 있다.

IV. 전략물자 불법 거래 탐지시스템

전략물자는 나라에서 일정량 이상 비축해두어야 할 필요가 있는 물품 및 기술로서 수출입 시 제한을 받는다. 대외무역법 시행령에 의해 전략물자인지 검증을 받은 후 매매를 진행해야 한다. 하지만 전략물자가 다크웹에서 불법적으로 거래될 위험이 있다는 것은 분명하다.

이 절에서는 더스팅 공격(Dusting attack)과 CNN을 활용한 평거 프린팅을 결합하여 불법 거래 탐지시스템을 구축할 것을 제안한다.



[그림 2] 제안 시스템 구성도

먼저 전략물자관리원에서 지정한 각 항목 별 전략물자 리스트를 목록화하여 딥웹 전체를 크롤링하였을 때 나오는 결과물과 비교하여 다크웹에 있는 사이트가 전략물자를 취급하는 곳인지 필터링한다. 둘째, 전략물자를 토르 브라우저 상의 다크웹에서 검색하여 전략물자를 취급하는 사이트에 방문하고 거기서 비트코인 주소를 크롤링한다. 셋째, 더스팅 공격과 CNN을 활용한 핑거프린팅 기술을 사용하여 거래 내역 정보를 얻는다. 넷째, 크롤링한 비트코인 주소와 사용자 신원 정보를 목록화하여 블랙리스트로 등록하고 데이터베이스화한다. 이 시스템을 실시간으로 동작시켜 블랙리스트 데이터베이스가 토르 브라우저의 변동 사항에 맞추어 상시 업데이트되도록 한다.

더스팅 공격[4]이란 미량의 비트코인을 사용자 지갑으로 송금하여 암호화폐 사용자의 프라이버시를 침해하는 공격을 말한다. 사용자는 먼지(dust) 같이 거래 수수료보다도 작은 아주 작은 양의 비트코인을 송금하면 대개 많은 사람이 이러한 이벤트를 간과한다. 이를 이용하여 미량의 비트코인을 수신한 지갑 주소의 소유기관 및 소유주를 알 수 있다.

이것을 이용하면, 불법으로 전략물자를 구매 또는 판매하기 위해 코인 믹싱월렛(Mixing wallet)을 사용하여 비트코인 주소로 송금 또는 수금했다고 할지라도 미량의 코인이 섞여 있는 한 해당 믹서에 속한 사용자의 거래정보와 믹싱월렛에 대한 정보를 얻어 자금세탁을 관계자 정보까지 수집할 수 있다.

V. 결론

본 논문에서는 더스팅 공격과 CNN을 활용한 핑거 프린팅 기법을 결합하여 전략물자를 불법으로 취급하는 곳을 추적하고 더불어 블랙리스트를 실시간으로 업데이트할 수 있는 시스템을 제안했다. 그러나 사무라이 지갑 등 미량의 트랙잭션을 자동으로 보고하는 기능을 갖춘 지갑이 개발되었으며, 토르 브라우저의 익명성을 향상시키기 위한 연구가 지속적으로 진행 중이다. 이에 대응하여 본 논문의 저자는 전략물자가 거래될 시 데이터에 마크하는 방법 및 암호화폐를 사용한 자금세탁 및 거래를 추적할 수 있는 도구 개발을 위한 지속적인 연구를 진행할 것이다.

[참고문헌]

- [1] Pu, Tao, and Juan Wang. "TOR Anonymous Traffic Fingerprint Extraction and Recognition Based on Neural Network." Journal of Physics: Conference Series. Vol. 1757. No. 1. IOP Publishing, 2021.
- [2] J. Seo, M. Park, H. Oh and K. Lee, "Money Laundering in the Bitcoin Network: Perspective of Mixing Services," 2018 International Conference on Information and Communication Technology Convergence (ICTC), 2018
- [3] Reynolds, Perri; Irwin, Angela SM. Tracking digital footprints: anonymity within the bitcoin system. Journal of Money Laundering Control, 2017.
- [4] Crawford, Jesse, and Yong Guan. "Knowing your bitcoin customer: Money laundering in the bitcoin economy." 2020 13th International Conference on Systematic Approaches to Digital Forensic Engineering (SADFE). IEEE, 2020.

DID 지갑을 위한 하드웨어 키 관리 방안 연구

Zhuohao Qian, 이경현*

부경대학교 정보보호학과

*부경대학교 IT융합응용공학과

zhuohaoq@gmail.com, khrhee@pknu.ac.kr

A Research of hardware key management method for DID wallet

Zhuohao Qian and Kyung-Hyune Rhee*

Department of Information Security, Graduate School,
Pukyong National University

*Department of IT Covergence and Application Engineering,
Pukyong National University

요약

DID 자기주권 신원은 블록체인 기술의 발전과 함께 공무, 금융, 의료, 사물인터넷 여러 분야의 신원인증에 활용하고 있다. 전통적인 신원증명 메커니즘에 있는 플랫폼한 상호운용성 문제를 해결한다. DID시스템에 신원증명의 절차는 사용자가 보유한 개인키를 기반으로 이루어지며 개인키의 분실은 신분 위장 등의 보안 위협을 초래할 수 있기에 사용자 개인키에 대한 안전한 관리에 관한 연구가 반드시 필요하다. 본 논문에서는 이동 단말기 환경의 하드웨어 보안 기술(TEE 및 SE)를 이용해 키 생성, 저장, 백업 및 복원을 관한 개인키 관리방법을 제안한다.

I. 서론

자기주권신원(Self-Sovereign Identity, SSI) 사용자가 스스로 자신의 신원을 관리 및 활용 할 수 있는 체제를 의미하며 블록체인 기술의 발전과 함께 새로운 신원증명 패러다임으로 주목받고 있다. 전통적인 신원증명 메커니즘은 중앙 인증기관에 전적으로 의존하는 중앙집중화와 중앙화된 서비스 플랫폼으로 인한 플랫폼한 상호운용성 문제[1]가 존재하였다. 이러한 문제를 해결하기 위한 방법으로 탈중앙 신원증명(Decentralized Identifier, DID)가 제안되었다. DID는 블록체인과 같은 분산 원장 기술을 바탕으로 탈중앙 신원증명 체계를 구축하고 탈중앙성을 통해 여러 플랫폼 간의 상호운용성을 보장하여 앞서 언급한 문제를 해결할 수 있었다.

블록체인은 공개키 암호 기술에 기반하여 시스템에서 자신의 권한을 증명한다. DID 시스템에서 식별자는 특정한 주소의 포맷으로 블록체인 위에 기록되어 있으며 사용자 신원정보의

검증은 해당 식별자와 이에 연결된 검증가능한 크리덴셜(Verifiable Credential)을 보임으로서 이루어진다[2]. 이러한 모든 신원증명의 절차는 사용자가 보유한 개인키를 기반으로 이루어지며 개인키의 분실은 신분 위장 등의 보안 위협을 초래할 수 있기에 사용자 개인키에 대한 안전한 관리에 관한 연구가 반드시 필요하다. 본 논문에서는 이러한 개인키 관리에 대해 단순한 암호화 기술을 이용한 관리를 넘어 단말기 내의 하드웨어적으로 안전한 영역에 개인키를 보관하는 방법을 제안한다.

II. 하드웨어 기반의 개인키 관리

2.1 TEE 기반 솔루션

신뢰실행환경(Trusted Execution Environment, TEE)은 메인 프로세서 내에 별도로 독립된 보안 영역으로 내부에 업로드되는 프로그램 코드와 처리되는 데이터가 안전하게 보호되도록

보장할 수 있다. ARM의 Trustzone은 모바일 환경에 적합한 TEE 기술로 디바이스 SOC(System-on-a-Chip)의 하드웨어 및 소프트웨어 리소스를 보안과 비보안의 두 가지 실행 환경으로 나누고 보호되어야 하는 실행 함수를 TEE OS에서 TA(Trusted Application)를 사용하여 인코딩하면 Rich OS쪽의 프로그램이 API 호출로 실행환경을 전환하고 TA가 그 실행 결과를 반환한다[3]. 그림 1은 DID 지갑에서의 TEE 기반 솔루션 모델을 도식화 한 것이다.

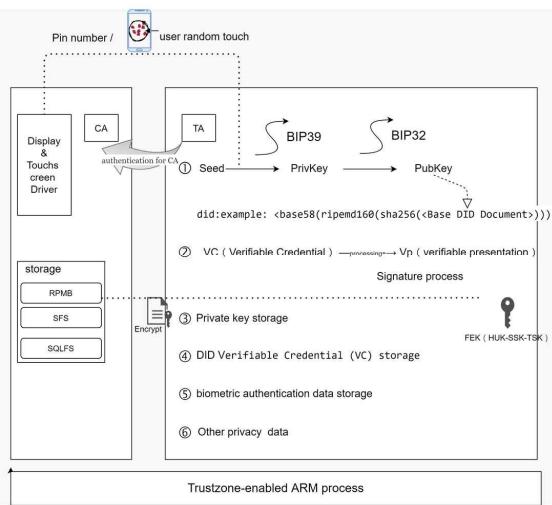


그림1. TEE기반 키 관리 모델

TEE 기반의 개인키 관리 솔루션은 터치스크린 입력 및 디스플레이 출력 보안(PIN 코드 입력 / seed 생성 프로세스에 사용자의 터치 난수 추가), TEE에서 키 생성 과정 및 서명 연산 등을 실행 하여 DID에서 개인키와 관련한 연산을 독립된 안전한 환경에서 실행함으로서 외부의 위협에서 사용자의 개인키를 안전하게 보관 할 수 있다.

2.2 SIM카드 기반의 키 관리

TEE 기반 솔루션은 개인키 보관 및 연산에 대한 안전성을 제공하지만 상호운용성 및 이동성에 대해서는 여전히 한계를 지니고 있어 개선이 필요하다[4]. 모바일 장치의 SIM 카드는 SE(Secure Element)의 일종으로 TEE보다 물리적으로 보다 독립된 실행환경을 제공할 수 있으며 이동성 또한 뛰어나다.

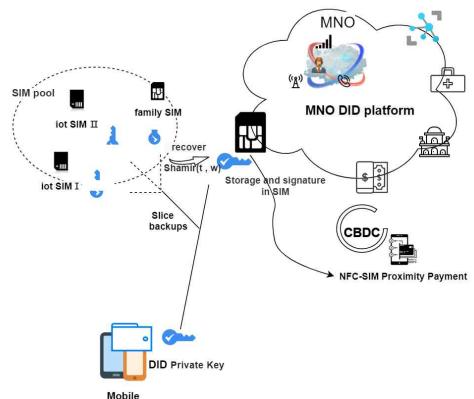


그림2. SIM카드기반 키 관리 모델

그림2와 같이 DID에 사용되는 개인키를 SIM 카드에 저장함으로서 기존 TEE보다 우수한 실행환경의 제공이 가능하며 표준화된 SIM 카드 아키텍처를 바탕으로 단말의 교체 등으로 인한 저장 데이터의 이동 필요시 별다른 제약 없이 손쉽게 데이터를 다른 단말에 이식하는 것이 가능하다.

III. 결론

본 논문에서는 탈중앙 신원증명 체계에서 핵심적인 역할을 담당하는 개인키의 관리를 위한 방법으로 TEE와 SIM카드를 사용하는 솔루션을 소개하였다. 두 접근 방식 모두 인가되지 않은 앱은 접근할 수 없는 독립된 영역에 개인키를 보관하고 연산을 수행한 뒤 그 결과만을 Rich OS의 앱에게 반환하여 해킹 등으로 인한 개인키의 노출을 해결할 수 있다. 하지만 TEE 기반의 솔루션은 플랫폼 간의 상호운용성이 보장되지 않고 단말 간 보관된 데이터의 이식에 한계가 존재하였다. 제안한 SIM카드 기반의 솔루션은 저장 독립성, 단말간 데이터 이동성, 개인키 복구 등의 장점이 있으나 기술의 타당성 면에서 아직 다각적인 연구와 논의가 필요하다.

[사사표기]

본 연구는 과학기술정보통신부 및 정보통신 기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었으며 (IITP-2021-2020-0-01797) 일부는 2021년도 정부(교육부)의 재원으

로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2021R1I1A304659011)

[참고문헌]

- [1] O. Avellaneda et al., "Decentralized Identity: Where Did It Come From and Where Is It Going?," in IEEE Communications Standards Magazine, vol. 3, no. 4, pp. 10–13, December 2019, doi: 10.1109/MCOMSTD.2019.9031542.
- [2] M. -H. Rhie, K. -H. Kim, D. Hwang and K. -H. Kim, "Vulnerability Analysis of DID Document's Updating Process in the Decentralized Identifier Systems," 2021 International Conference on Information Networking (ICOIN), 2021, pp. 517–520, doi: 10.1109/ICOIN50884.2021.9334011.
- [3] W. Dai, J. Deng, Q. Wang, C. Cui, D. Zou and H. Jin, "SBLWT: A Secure Blockchain Lightweight Wallet Based on Trustzone," in IEEE Access, vol. 6, pp. 40638–40648, 2018, doi: 10.1109/ACCESS.2018.2856864.
- [4] Z. Ahmad, L. Francis, T. Ahmed, C. Lobodzinski, D. Audsin and P. Jiang, "Enhancing the Security of Mobile Applications by Using TEE and (U)SIM," 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing, 2013, pp. 575–582, doi: 10.1109/UIC-ATC.2013.76.