Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.28, NO.6, Dec. 2018 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2018.28.6.1523

AI 스피커의 보안성 평가 및 대응방안 연구*

이 지 섭,† 강 수 영, 김 승 주‡
고려대학교 정보보호대학원

Study on the AI Speaker Security Evaluations and Countermeasure*

Ji-seop Lee,† Soo-young Kang, Seung-joo Kim‡
Center for Information Security Technologies(CIST), Korea University

요 약

AI 스피커는 간단한 동작으로 음악재생, 온라인 검색 등 사용자에게 유용한 기능을 제공하고 있으며, 이에 따라 AI 스피커 시장은 현재 매우 빠른 속도로 성장하고 있다. 그러나 AI 스피커는 항시 사용자의 음성을 대기하고 있어 보안 위협에 노출될 경우 도청, 개인정보 노출 등 심각한 문제가 발생할 수 있다. 이에 모든 AI 스피커의 전반적으로 향상된 보안을 제공하기 위해 발생 가능한 보안 위협을 식별하고 체계적인 취약점 점검을 위한 방안이 필요하다.

본 논문에서는 점유율이 높은 제품 4개를 선정하여 보안위협모델링을 수행하였다. Data Flow Diagram, STRIDE, LINDDUN 위협모델링을 통해 체계적이고 객관적인 취약점 점검을 위한 체크리스트를 도출하였으며, 이후 체크리스트를 이용하여 실제 기기에 대한 취약점 점검을 진행하였다. 마지막으로 취약점 점검 결과 및 각 제품에 대한 취약점 비교·분석을 통해 AI 스피커의 보안성을 향상시킬 수 있는 방안을 제안하였다.

ABSTRACT

The AI speaker is a simple operation that provides users with useful functions such as music playback, online search, and so the AI speaker market is growing at a very fast pace. However, AI speakers always wait for the user's voice, which can cause serious problems such as eavesdropping and personal information exposure if exposed to security threats. Therefore, in order to provide overall improved security of all AI speakers, it is necessary to identify potential security threats and analyze them systematically.

In this paper, security threat modeling is performed by selecting four products with high market share. Data Flow Diagram, STRIDE and LINDDUN Threat modeling was used to derive a systematic and objective checklist for vulnerability checks. Finally, we proposed a method to improve the security of AI speaker by comparing the vulnerability analysis results and the vulnerability of each product.

Keywords: AI Speaker, Threat Modeling, STRIDE, LINDDUN

I. 서 론

AI 스피커는 음성인식 AI 플랫폼으로, 사용자가 음성 명령을 전달하면 인공지능이 그 음성을 이해하여 음악 재생, 알람 설정, 인터넷 검색 등 특정 기능을 수행한다. 이렇듯 간단한 동작으로 사용자에게 유용한 기능을 제공하고 있으며, 이에 따라 AI 스피커 시장은 현재 매우 빠른 속도로 성장하고 있다.[1]

Received(11. 07. 2018), Modified(11. 22. 2018), Accepted(11. 22. 2018)
* 본 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (IITP-2017-0-00184,자기학습형 사이버 면역 기술 개발)
† 주저자, gsleegs4@naver.com
‡ 교신저자, skim71@korea.ac.kr(Corresponding author)

그러나 AI 스피커는 “Alexa”, “Ok Google” 등 Wake word의 입력 대기상태로 사용자의 음성을 듣고 있어, 보안 위협에 노출될 경우 도청, 개인정보노출과 같은 심각한 문제가 발생할 수 있다. 이와 관련하여 최근 5월에는 AI 스피커의 오작동으로 인해 사용자의 음성이 타인에게 전송되어 개인정보가 유출되는 피해가 발생하였다.[2]

이러한 보안위협에 대한 완화방안을 마련하기 위해 다양한 연구가 진행되었다. 2016년에는 사람이 인식할 수 없는 음성을 이용한 공격[3]을 통해 스피커 제어가 가능함을 보였고, 2017년에는 악성 애플리케이션 제작을 통한 도청[4], 2018년에는 플래시 메모리 덤프를 통한 원격 제어 공격[5] 등이 연구되었다. 이 외에도 여러 연구가 진행되었으나, 현재까지의 연구 실태를 보았을 때, AI 스피커에 대한 보안 연구는 분석가의 경험적인 노하우에 의존하고 있는 상황이다. 따라서 AI 스피커의 전반적인 보안성을 제공하기 위해, 발생 가능한 모든 보안 위협을 식별하고 체계적인 취약점 점검을 위한 방안이 필요하다.

이에 본 논문에서는 AI 스피커에 적합한 보안위협모델을 선정하고, 이를 통해 체계적이고 객관적인 취약점 점검을 위한 체크리스트[6]를 도출한다. 이후 체크리스트를 이용하여 실제 기기에 대한 취약점 점검 및 사례 연구를 통해 AI 스피커의 보안성을 향상시킬 수 있는 방안을 제안한다. 분석 대상은 타 모델에 비해 상대적으로 기기 보안에 대한 기대치가 높고 향후에도 다양한 사람들이 선택할 것으로 판단되는 국내·외 점유율이 높은 모델로 선정하였다.

본 논문은 서론에 이어, 2장에서는 보안위협모델링 및 AI 스피커 보안 관련 연구를, 3장에서는 AI 스피커의 보안성 평가 기준을 도출한다. 4장에서는 평가 기준을 통해 AI 스피커 보안성 평가를 수행하고, 각 제품에 대한 비교·분석을 통해 현재 AI 스피커에 대한 문제점을 지적한다. 마지막으로 5장에서는 AI스피커의 보안 문제에 대한 대응방안을 서술하고, 결론을 맺는다.

II. 관련 연구

2.1 보안위협모델링

보안위협모델링은 체계화된 구조에 따라 제품의 전반적인 보안위협을 식별하는 방법이다. 이는 보안개발수명주기(SDL, Security Development Lifecycle)의 설계 단계에서 발생 가능한 취약점을 미리 식별하여 제거하고 보안성을 강화하기 위해 사용한다. 또한 위협 발생 가능성 및 잠재적 피해 및 영향을 포함한 보안 위협의 완화 전략을 구상하는 과정이다.[7]

이 장에서는 가장 널리 사용되고 있는 보안위협모델에 대해 간략하게 설명한다. 또한 보안위협모델의 비교를 통해 본 연구에 적합한 모델을 선정하도록 하겠다.

2.1.1 STRIDE

STRIDE는 위장(Spoofing), 변조(Tampering), 부인(Repudiation), 정보 노출(Information Disclosure), 서비스 거부(Denial of service), 권한 상승(Elevation of Privilege) 등 소프트웨어에서 발생할 수 있는 6가지 보안 위협을 다룬다.[8] 이 프레임워크는 소프트웨어에 대한 공격 유형 정의 및 위협 식별을 위해 설계되었으며, STRIDE의 각 속성에 대한 설명은 다음과 같다.

a) 위장 : MAC 주소, IP 주소, 포트, 이메일 등 네트워크 통신과 관련된 정보를 속이고, 이를 사용자가 신뢰하도록 유도하는 행위
b) 변조 : 프로세스, 파일, 네트워크 전송 값 등 대상의 구성요소를 변조시키는 행위
c) 부인 : 주체가 대상에 대해 읽기, 쓰기, 접근과 같은 행위를 한 뒤, 이를 부인하는 행위
d) 정보 노출 : 데이터 주체의 민감한 정보가 허가되지 않은 대상 또는 사람에게 노출됨
e) 서비스 거부 : 정보 시스템의 데이터나 자원을 적절한 대기 시간 내에 사용하는 것을 방해하는 행위
f) 권한 상승 : 권한 없는 사용자가 특정 권한을 획득하여 정보 시스템을 손상시키는 행위

2.1.2 Trike

Trike는 데이터 흐름(Data Flow) 및 사용 흐름(Use Flow) 내의 사용자, 자산을 식별하고 자산의 4가지 요소(Create, Read, Update, Delete)에 대한 사용자의 수행 빈도를 분석하여 각 자산에 대한 위험도를 산출하는 위협모델이다.[9] 이 프레임워크는 어택트리 및 어택 라이브러리를 활용하여 자산의 취약점을 식별하며, 위험관리(Risk Management) 관점으로 자산을 관리하는 것이 특징이다.

2.1.3 LINDDUN

LINDDUN은 연결(Linkability), 식별(Identifiability), 부인 방지(Non-repudiation), 검출(Detectability), 정보 노출(information Disclosure), 내용 몰인식(content Unawareness), 정책 및 동의 불이행(policy and consent Non-compliance) 등 7가지 개인정보 관련 위협을 다룬다.[10] 이 위협모델은 DFD(Data Flow Diagram)으로 표현된 외부 객체, 프로세스, 데이터 저장소, 데이터 흐름에 대해 위협을 식별하고, 각각의 위협을 위협트리로 작성하여 개인정보위협을 체계화한다. 이후 위협트리에 대한 상세 설명을 제공하기 위해, 오용사례 시나리오를 작성하여 식별된 위협을 문서화한다. 마지막으로 도출된 위협을 완화하기 위한 전략을 결정한다. LINDDUN의 각 속성에 대한 설명은 다음과 같다.

a) 연결 : 획득한 데이터들을 통해 데이터 주체를 연결지어(Link) 유추 가능
b) 식별 : 획득한 데이터를 통해 데이터 주체 식별 가능
c) 부인 방지 : 데이터 주체가 데이터의 소유권의 부인 불가능
d) 검출 : 데이터를 통해 데이터 주체의 관심항목(데이터 주체의 생활패턴 등) 구별 가능
e) 정보 노출 : 데이터 주체의 민감한 정보가 허가되지 않은 대상 또는 사람에게 노출됨
f) 내용 몰인식 : 데이터 주체는 자신의 데이터 수집, 처리, 저장 또는 공유 활동을 인식하지 못함
g) 정책 및 동의 불이행 : 개인정보의 처리, 저장 또는 취급이 법률을 준수하지 않음

2.1.4 PASTA

PASTA(The Process for Attack Simulation and Threat Analysis)는 7단계로 구성된 위험 위주의 위협모델이다. 이 모델은 공격자 관점으로 위협을 식별 및 분류하고 각 위협에 대한 점수를 산출하여, 보안실무자가 자산 중심의 완화전략을 개발할 수 있도록 프로세스를 제공한다.[11] PASTA 위협모델이 다른 모델과의 가장 큰 차이점은 위험 및 비즈니스 영향 분석 프로세스에 조직의 주요 의사 결정권자를 참여시키는 것이다. 즉, 자산에 대한 위협을 소프트웨어 개발자 및 보안실무자 관점으로만 보지 않고 사업에 미칠 영향까지 포함하여 고려한다.

2.1.5 보안위협모델 비교·분석

2.1.1~2.1.4에서 소개한 위협모델링 방법은 모두 분석 대상에 대한 위협을 분류하고 완화방안을 제공하지만, 각각 다른 관점으로 접근한다. Table. 1은 위협 초점, 분류 기법 및 문서화 지원 등 각각의 위협모델에 대한 차이점을 보여준다.

4개의 보안위협모델의 특성을 비교한 결과, AI 스피커 기기 자체의 보안 위협 및 개인정보 유출 위협을 동시에 고려하는 모델은 존재하지 않는다. 그러나 위협모델의 분석 초점 및 관점, 문서화 지원 정도를 보았을 때, 보안위협모델에 대한 문서가 잘 정리되어 있

Table 1. Comparison of threat modeling

	STRIDE	Trike	LINDDUN	PASTA
Analysis focus	Design	Requirements	Design	Requirements
Analysis viewpoint	Software Vulnerability	Risk management for Asset	Privacy	Risk management for Business
Identification of system element	Data Flow Diagrams	Data Flow Diagrams & Use Flows	Data Flow Diagrams	Data Flow Diagrams
Threat determination methodology	STRIDE	Actor, Asset, Action matrix	LINDDUN	Threat-attack scenarios
Complexity	Medium	High	High	High
Documentation and support	Very good	Bad	Good	Good
Last update	2018	2012	2014	N/A

고 설계 단계에서 보안 위협을 식별하며 S/W 취약점, 개인정보보호 관점으로 보안 위협을 분류하는 위협 모델인 STRIDE와 LINDDUN을 결합하는 것이 AI 스피커에 가장 적합한 접근 방법으로 판단된다.

2.2 AI 스피커 보안 연구 사례

기존의 AI 스피커 보안 관련 연구는 전문적인 경험을 바탕으로 특정 부분에 한해 위협 식별 및 분석을 수행하는 중이다.

Nicholas Carlini, Pratyush Mishra 외 6명은 사람이 해석할 수 없는 음성 명령을 통한 음성 인식 시스템의 공격 방법을 설명하였다.[3] 블랙박스, 화이트박스 모델을 이용하여 음성 인식 시스템에 대한 공격자의 접근법을 보여주었으며, 이러한 음성 공격에 대한 완화책을 제안하였다. 하지만 해당 논문은 음성 공격에 대한 위험성을 제시만 할뿐 실제 공격을 시도하지 않았다. 즉 실제 공격에 활용될 가능성에 대한 분석은 진행하지 않았으며 이론적인 수준에서만 공격 가능성을 보여주는 한계가 있다.

Hyunji Chung, Michaela lorga 외 2명은 Amazon Echo에서 발생 가능한 위협, 개인정보보호 위험 등을 4가지로 구분하여 설명하였다. 이들은 무선 업데이트 중인 펌웨어 노출, 클라우드 서버로 전송되는 음성파일 노출, AI 스피커-서버 간 전송 데이터 스니핑, 임의의 음성 명령 전송을 통한 스마트 기기 제어 위협 등을 설명하였다.[12] 그러나 본 논문은 일반적으로 예측 가능한 수준의 단순 위협을 소개하고 있으며, AI 스피커의 보안 위협에 대한 연구가 심도 깊게 진행되지는 않았다.

Ike Clinton, Lance Cook, Shankar Banik은 Amazon Echo를 대상으로 eMMC 물리적 접근 경로를 통해 실제 취약점 분석을 수행하여 스마트 기기가 개인 정보에 미칠 수 있는 영향을 설명하였다.[13] 또한 UART, JTAG와 같이 다양한 접근 경로를 통한 공격 가능성도 설명하였다. 본 논문은 물리적 접근 경로를 통한 다양한 하드웨어 리버싱 방법을 서술한 것이 특징이다. 하지만 본 연구에서 제안한 공격 방법은 연구·개발을 위한 테스트 제품이 아닌 소비자에게 판매되는 완제품에 대해서 UART, JTAG 등 물리적 접근 경로를 차단한다면 쉽게 대응할 수 있다. 또한 공격을 위해 사용자와 관련한 사전 조사 및 사용자에게 접근하는 과정이 포함되어 있으므로 실제 공격 발생 가능성이 극히 제한적이다.

Tencent Blade Team은 DEFCON 26에서 MIIO, Amazon Echo를 대상으로 인증 토큰 노출 및 펌웨어 획득으로 인해 발생 가능한 공격 위협을 설명하였다. 이 연구는 AI 스피커의 특정한 부분에 대해 공격을 진행한 것으로, 원격 제어 공격에 대한 가능성을 보여주었다.[5] 하지만 AI 스피커의 전반적인 보안성 향상을 고려하지 않고 있어, 기기의 특정 기능에 대한 완화책만 제시해주는 한계를 가지고 있다.

이렇듯 현재까지 진행된 AI 스피커 보안 연구는 분석가의 경험을 기반으로 기기의 특정한 부분에 대한 취약점 식별하거나 이론적 접근을 통한 공격 가능성을 제안하는 등 기기 취약점에 따른 위험성이 많은 연구를 통해 증명되고 있다. 하지만 분석가의 지식 범위 내에서만 연구되고 있다는 한계점이 존재한다. 이에 본 논문에서는 한계점 보완을 위해 위협모델링을 활용하여 AI 스피커의 보안 위협을 체계적으로 식별·분석하고, 이에 대한 대응방안을 제시하였다.

III. AI 스피커에 대한 보안성 평가 기준 도출

3.1 Data Flow Diagram 도출

본 논문에서는 취약점, 개인정보보호 관점에서 위협을 식별하기 위해 STRIDE와 LINDDUN을 적용하였다. 우선 AI 스피커의 전반적인 구조를 파악하기 위해 DFD(Data Flow Diagram)를 활용하였다. DFD는 분석 대상의 외부 객체, 프로세스, 데이터 저장소, 데이터 흐름에 대한 시각화를 통해 데이터 흐름을 보여주기 때문에, 정확하게 작성되었을 경우 분석 대상의 공격 지점 및 방법의 식별이 용이하다.

AI 스피커의 DFD 작성 결과는 Fig. 1과 같으며, 이에 대한 전체 설명은 다음과 같다.

a) 사용자가 AI 스피커를 사용하기 위해, 먼저 앱 스토어에 공개된 애플리케이션(P10)을 사용자의 스마트폰에 설치한다.(P12)
b) 애플리케이션 로그인(P14)를 통해 사용자 인증을 진행한다. 이후 AI 스피커의 네트워크 연결을 위해 애플리케이션 내에서 Wi-Fi 인증을 진행한다.(P2)
c) 사용자가 AI 스피커를 이용하기 위해 음성으로 지시한다.(P4) 사용자의 음성에 “Alexa”, 혹은 “Ok Google”과 같이 AI 스피커를 깨우는 단어가 감지될 경우(P5), 사용자의 음성을 음성 파일로 변환 후, 서버에 전송한다.(P3)

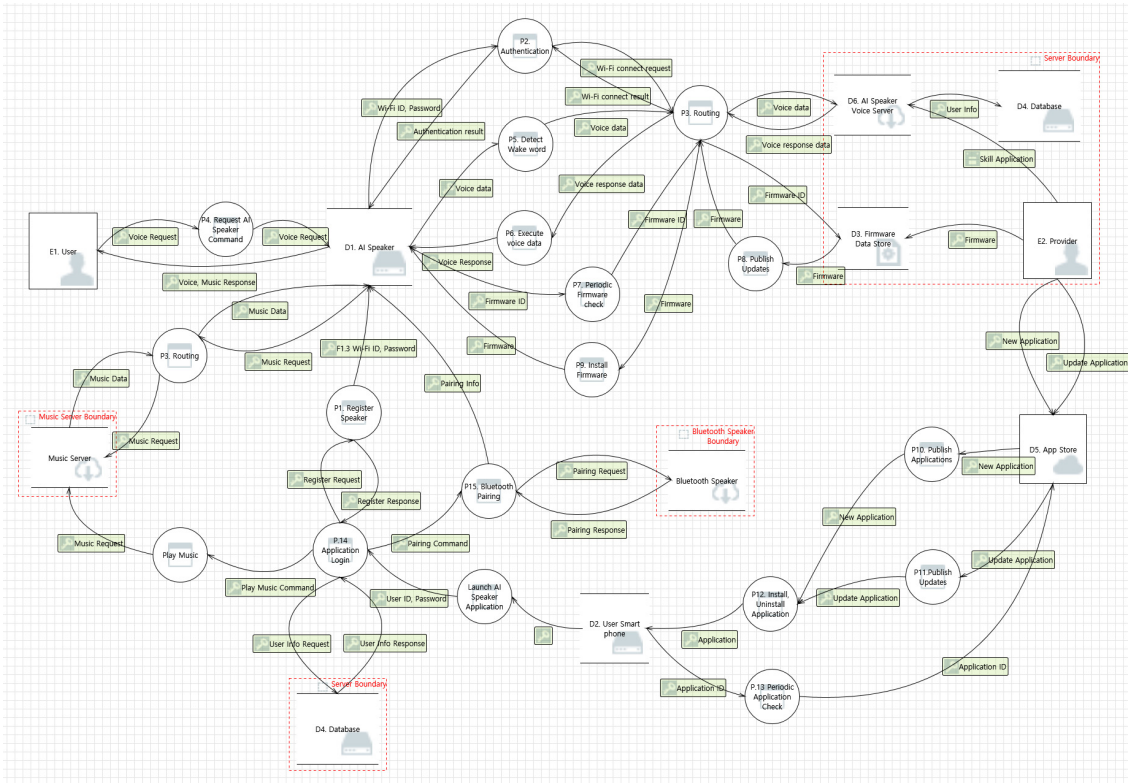


Fig. 1. AI Speaker Data Flow Diagram

d) 서버는 자동 음성 인식(Automatic Speech Recognition, ASR), 자연어 이해(Natural Language Understanding, NLU) 등을 통해 전송된 음성 명령을 이해하고 처리한다. 또한 처리 결과를 음성 파일로 변환 후, AI 스피커에 전송하여 실행한다.(P6)

e) 주기적인 펌웨어, 애플리케이션 버전 체크를 통해(P7, P13) 업데이트가 필요한 경우, 관련 요청을 서버에 전송하고 무선으로 업데이트를 진행한다.(P8, P11)

3.2 STRIDE 위협 모델링

3.2.1 어택 라이브러리

어택 라이브러리는 DFD의 각 요소에서 발생 가능한 보안위협을 정확하게 식별했는지 판단할 수 있는 근거를 제공한다. 이는 논문, 컨퍼런스, 기술문서 및 취약점 데이터베이스, 표준 등 다양한 자료를 수집하여 구축할 수 있다. Table 2는 본 연구를 위해 수집한 위협 목록 중 일부를 보여준다.

이러한 어택 라이브러리를 작성할수록 분석 대상의 보안위협을 구체화할 뿐만 아니라 현재까지 알려진 모든 공격 지점 및 방법, 유형을 보안성 평가 지표에 세밀하게 반영할 수 있는 장점이 있다.

3.2.2 STRIDE

3.1에서 작성한 DFD의 구성 요소는 각자의 잠재적인 위협을 가지고 있으므로, STRIDE를 통해 각 요소가 가지고 있는 위협을 식별할 수 있다. Table. 3은 Fig. 1의 DFD에 STRIDE 모델을 적용하여 식별한 91개의 보안위협 중 일부를 보여준다.

STRIDE는 보안위협을 단순 열거한 방법으로, 식별된 위협이 실제 활용되는지 확인하기 위해 3.2.1에서 구축한 어택 라이브러리와의 상관관계를 Table. 3에 포함하였다.

Table. 2. AI Speaker Attack Library

Type	Year	Title	Author	Ref
Conference	2013	Embedded Devices Security and Firmware Reverse Engineering	Zaddach	[14]
	2017	Exploiting BlueBorne in Linux-based IoT devices	Ben Seri	[15]
	2018	Breaking Smart Speaker, We are Listening to you	Tencent Blade Team	[5]
Vulnerability	2018	CVE-2018-9070	MITRE	[16]
Paper	2016	Hidden Voice Commands	Nicholas Carlini	[3]
	2017	DolphinAttack: Inaudible Voice Commands	Xiaoyu Ji	[17]
	2017	Security Analysis of the Amazon Echo	William Haach	[18]
Technical Report	2016	Security Vulnerabilities in Speech Recognition Systems	Oxford University	[19]
	2016	A Survey of Various Methods for Analyzing the Amazon Echo	Vanderpot	[13]
	2017	Exploiting the Amazon Echo Dot, Part 1: Intercepting firmware updates	Vanderpot	[20]

Table 3. STRIDE for AI Speaker DFD

Type	No	Name	Threat Description		Attack Library	Threat No
Entity	E1	User	S	The attacker masquerades as a User	18, 24, 25, 27	T1
			R	The attacker denies the control of the speaker	18, 24, 25, 27	T2
Entity	E2	Provider	S	The attacker masquerades as a Provider	36, 45	T3
			R	The attacker denies the provided of the speaker	36, 45	T4
Process	P1	Register Speaker	T	Threats to manipulate authentication values in transit	4, 10, 13, 16, 31, 57	T5
			I	Threats that expose User's ID, Password	4, 13, 16, 22, 31, 43	T6
			D	Threats that prevent you from performing User authentication by passing invalid argument values	4, 10, 31, 43, 57	T7
Process	P2	Authentication	T	Threats to manipulate authentication values in transit	4, 13, 20, 31, 43, 57	T8
			I	Threats that expose User's ID, Password	4, 13, 20, 22, 31, 43	T9
			D	Threats that disable data transmission of authentication values	4, 20, 31, 43	T10
			E	Threats to gain the privileges of a router through a specific attack	20, 29, 30	T11
Process	P3	Routing	S	After an attacker obtains an administrator account using a random assignment attack, Attacker masquerades as a administrator	29, 30	T12
			T	Threats to tamper with existing file system files	29, 30, 45	T13
			R	Threats denying actions such as accessing, running, or tampering with the file system	29, 30, 43, 45	T14
			I	Threats that expose users' data flowing through the router	4, 13, 20, 22, 31, 43	T15

omit

Data Store	D6	AI Speaker Voice Server	T	The threat that an attacker sends a malformed voice file to the server	18, 24, 25	T90
			R	A threat that denies attackers from sending malformed voice files to the server.	18, 24, 25	T91

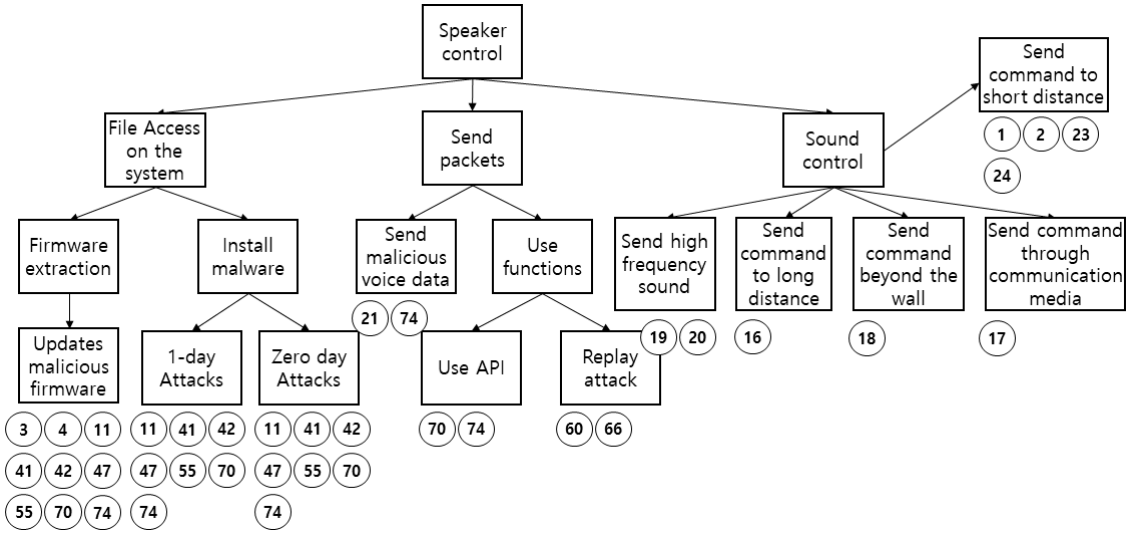


Fig. 2. AI Speaker Attack Tree

3.2.3 어택트리

분석 대상에 대한 모든 보안위협을 체계화하기 위해 어택트리를 작성한다. 이를 위해 STRIDE를 적용하여 식별한 보안위협 및 분석 대상에 대한 세부 위협 목록인 어택 라이브러리를 활용하였다.

어택트리를 작성한 결과, AI 스피커에 대한 공격은 데이터 획득, 데이터 변조, 서비스 거부 공격, 스피커 제어 등 크게 4가지 유형으로 분류할 수 있다. Fig. 2는 작성한 결과 중 일부를 보여준다.

지금까지 STRIDE 위협모델링을 통해 취약점 관점에서 AI 스피커의 보안위협을 식별하였다. 그러나 STRIDE로는 개인정보에 대한 위협을 도출하기 어려우므로, 3.3에서는 LINDDUN을 이용하여 개인정보 보호 관점으로 AI 스피커의 보안위협을 도출한다.

3.3 LINDDUN 위협 모델링

3.3.1 LINDDUN

LINDDUN은 개인정보보호 관점에서 작성하는 위협모델로, DFD를 기반으로 분석을 진행하는 부분에 있어 STRIDE와 유사하다.

LINDDUN은 DFD의 각 요소에 따라 적용되는 개인정보 위협의 범위가 다르며, Table. 4는 DFD의 각 요소별 적용 가능한 LINDDUN의 범위를 보여준다.[10] 위 표를 근거로 AI 스피커에 대한

Table 4. Mapping LINDDUN to DFD element

	L	I	N	D	D	U	N
Entity	X	X				X	
Process	X	X	X	X	X		X
Data Store	X	X	X	X	X		X

Table 5. LINDDUN for AI Speaker DFD

DFD element	ID	L	I	N	D	D	U	N
Entity	E1		X				X	
	E2		X				X	
Process	P1	X				X		X
	P2					X		X
	P3	X	X			X		X
	P4	X	X		X	X		X
	P5	X	X		X	X		X
	P6	X	X		X	X		X
	P7					X		X
	P8					X		X
	P9							X
	P10							X
	P11					X		X
	P12							X
	P13					X		X
	P14	X	X			X		X
	P15				X	X		X
Data Store	D1	X	X		X	X		X
	D2	X	X		X	X		X
	D3					X		X
	D4	X	X		X	X		X
	D5	X			X	X		X
	D6	X	X		X			X

Table 6. A part of Data Store of LINDDUN

ID		Description	Attack library
D1	L	The user can be inferred through voice command information of the user in the AI speaker.	6, 10, 16, 23, 25, 35, 40, 42, 53
	I	Identify users through voice files in AI speakers.	6, 10, 16, 23, 25, 35, 40, 42, 53
	D	The living pattern and environment of the user can be grasped through the voice command information of the user in the AI speaker.	6, 10, 16, 23, 25, 35, 40, 42, 53
	D	User's voice command information exposure	6, 16, 18, 19, 23, 24, 25, 33, 34, 35, 37, 39, 46

LINDDUN 위협을 식별하였으며, 그 결과는 Table. 5에 작성하였다. 그리고 식별 위협에 대한 설명은 Table. 6에 일부 작성하였다. LINDDUN 또한 STRIDE와 동일한 이유로, 3.2.1의 어택 라이브러리와의 관계를 Table. 6에 표현하였다.

3.3.2 위협트리

DFD의 구성요소에 대한 위협이 식별되면 각 위협의 상세 분류를 위한 위협트리를 작성할 필요가 있다. 위협트리는 3.3.1에서 도출한 Table. 6의 LINDDUN 위협에 대해 작성하며, 각 위협의 범주가 세분화되어 표현될 수 있다. Table. 7은 문서 형태로 작성된 AI 스피커의 위협트리 중, 일부를 보여주며, 이 중 *로 표시된 노드는 이미 작성된 위협트리의 재사용을 의미한다.

Table 7. A part of Data Store Threat Tree

Disclosure of information			
3	ID_d		
	3.1	ID_d1 : Bypass protection measures	
		3.1.1	ID_d5 : Protection measures not supported
		3.1.2	ID_d6 : Easy acquired access
		3.1.3	ID_d7 : Insufficient verification process
			3.1.3.1 ID_d17 : Supports passwordless accounts
			3.1.3.2 ID_d18 : Predictable credentials
			3.1.3.3 ID_d19 : Tampering threats to the authentication process
	3.2	ID_d2 : Acquisition of data through memory access	
		3.2.1	ID_d8 : Acquisition of data through malicious applications
		3.2.2	ID_d9 : Acquisition of data through physical access

3.3.3 오용사례

위협트리에 대한 이해도를 높이기 위해 오용사례(MUC)를 작성한다. 이 과정은 각 위협이 실제로 어떻게 발생할지 식별자, 공격자, 시나리오, 결과 등을 포함하여 작성할 수 있다. Table. 8은 데이터 저장소의 정보노출에 관한 MUC를 보여준다.

LINDDUN 위협모델의 경우 MUC 작성 후 위협의 피해를 최소화하기 위한 완화 전략을 결정하지만, 이 전략은 분석 대상에 대한 실제 분석을 진행하지 않고, 이론적으로 가능한 보안대책을 제시할 뿐이다.

이에 본 연구에서는 STRIDE, LINDDUN 위협모델로 식별한 위협을 참고하여 AI 스피커를 분석하

Table 8. A part of Misuse Case

Misuse Case	Details
MUC 11	**Tree** : ID_d **Summary** : The user's information can be acquired by accessing the user's storage, transmission data, and data not authorized to the user. **Main Attacker** : A competent inner / outer person **Scenario** : bf1. Obtain an accessible path to the repository. bf2. Obtain access to the repository. bf3. It retrieves and seizes user's data through repository access. bf4. The user's information can be obtained from the data with insufficient confidentiality.

고, 그에 대한 대응방안을 제안할 것이므로 완화 전략 과정은 생략하기로 한다.

3.4 위협 우선순위 지정

3.2.3, 3.3.3에서 식별한 위협을 체크리스트로 작성하기 위해서는 먼저 각 위협의 영향 정도를 파악하여 우선순위를 지정한다. 제품에 대한 모든 위험을 수용하기에는 많은 시간, 금전이 소모될 수 있으므로, 비용대비 효과적으로 위험을 관리할 필요가 있다.

이에 본 논문에서는 잘 알려진 마이크로소프트의 DREAD 모델을 활용하여 우선순위를 결정하였다.[21] DREAD는 피해 가능성(Damage potential), 공격의 재현성(Reproducibility), 악용 가능성(Exploitability), 영향 받는 사용자(Affected users), 취약점의 발견 가능성(Discoverability) 등 5가지 요소를 기준으로 위험 점수를 계산한다. 위험 점수는 낮음(1), 중간(2), 높음(3) 등 간단한 체계를 이용할 수 있으며, DREAD의 각 요소별 합산한 점수가 클수록 높은 우선순위를 지정한다. 본 연구에서 수행한 DREAD의 결과는 Table. 9와 같다.

Table 9. DREAD

STRIDE						
Threat	**D**	**R**	**E**	**A**	**D**	**Score**
Message sniffing	3	3	3	2	1	12
MITM	2	2	3	2	2	11
SQL injection	3	3	3	3	1	13
Brute force attack	2	3	2	3	1	11
Infinite reboot loop	1	2	1	1	2	7
Fork bomb	1	2	1	1	2	7
Update malicious firmware	3	1	1	2	1	8
Jamming	1	2	1	2	2	8
Misuse API	2	2	1	2	1	8
Flooding	1	2	1	2	1	7
Smurf	1	2	1	2	1	7
Ping of death	1	2	1	2	1	7
Replay attack	3	2	2	2	2	11
Send high frequency sound	2	1	1	1	2	7
Send command to long distance	2	3	1	1	2	9
Send command beyond the wall	2	3	1	1	2	9
Send command through communication media	3	2	1	3	2	11
Send command to short distance	2	3	1	1	2	9
Send malicious Voice data	2	1	1	2	2	8
Transmits the modulated packets	2	2	2	2	1	9
File modulation	2	1	1	1	1	6
LINDDUN						
Misuse Case	**D**	**R**	**E**	**A**	**D**	**Score**
MUC 01	2	3	3	2	1	11
MUC 02	2	2	1	2	1	8
MUC 03	2	2	1	2	1	8
MUC 04	2	2	1	2	2	9
MUC 05	2	2	1	2	2	9
MUC 06	3	2	1	2	2	10
MUC 07	2	1	1	3	1	8
MUC 08	2	2	1	3	1	9
MUC 09	2	2	1	3	1	9
MUC 10	1	2	1	3	1	8
MUC 11	3	2	1	3	1	10

3.5 취약점 점검을 위한 체크리스트

일반적으로 기기의 취약점을 점검할 때 네트워크를 통한 공격 및 시스템 상 취약점을 이용해 접근하는 등 분석가만의 노하우를 이용하거나, 알려진 체크리스트를 이용하여 점검을 진행한다. 그러나 이 방법은 점검 대상에 대한 모든 영역을 포괄하지 않으므로 전반적인 보안성 향상을 기대하기 어렵다. 따라서 기존 방식에 대한 미비점을 보완하기 위해 AI 스피커에 특화된 체크리스트를 작성하도록 한다.

STRIDE, LINDDUN 위협모델링을 적용하여 도출한 위협목록에서 실제 취약점이 발생할 수 있는 공격 벡터를 크게 애플리케이션, 네트워크, 하드웨어, 시스템 4가지 영역으로 구분하였다. 또한 체크리스트

Table 10. A part of AI Speaker Checklist

Type	Surface	No	Detail	Threat No
Application	Authentication Policy	A1	Check number of login attempts limit	T11, T60, T61
		A2	Confirm password rule settings, including password length and special symbols, alphanumeric characters	T11
		A3	Confirm of privacy exposure	T6, T9, T58
		A4	Confirm of authentication related information (cookies, session values, tokens, etc.)	T5, T8, T10, T59
	APP Permission	A5	Confirm unnecessary permission requests for APP behavior	T70, T71
	Apply Encryption	A6	Check APK source code obfuscation	T70, T74
		A7	Communication data encryption check	T58, T59
Network	Port Scanning	N1	Open port check	T68, T71
		N2	Identify unnecessary management ports	T68, T71
	Packet Sniffing	N3	Confirm whether sensitive information is encrypted	T27, T32, T33, T70, T78, T86
		N4	Acquiring sensitive information through MITM	T6, T58, T69, T70
		N5	Confirm firmware acquisition	T33
	Packet Transmission	N6	Verifies speaker control through arbitrary commands or voice file transfer	T23, T24
		N7	Confirm replay attack possible (user account access, speaker function, etc.)	T23, T62
Hardware	Check Debugging port	H1	Check UART port	T68
		H2	Check JTAG port	T68
	Firmware Acquisition	H3	Firmware acquisition via UART / JTAG port	T38, T39, T71
		H4	Firmware acquisition through flash memory dump	T36, T39
	Firmware Modification	H5	Modify the firmware through the boot loader	T38
System	System shell Check	S1	Confirm ID, password exposures for administrators in the system shell	T68, T70
		S2	Confirmation of general user's access to administrator shell	T68, T70
	Privacy Processing	S3	Confirm of non-identification of personal information	L_d3
		S4	Whether or not the user's personal information is destroyed after the retention period (2 years)	L_d2, L_d3, L_d4, L_d5

의 완전성 입증을 위해 각 위협모델에서 식별한 위협 목록을 각 체크리스트의 항목과 연관 지어 표현하였으며, 그 결과는 Table. 10과 같다.

현재까지 DFD, STRIDE, LINDDUN 등 체계적인 위협모델링 절차에 따라 AI 스피커의 모든 위협을 식별하였으며, 또한 연구 중인 모든 AI 스피커 관련 자료 수집 및 발표된 모든 AI 스피커의 보안 위협을 체크리스트에 포함하였다. 이러한 연구는 보안실무자에게 AI 스피커의 보안점검을 빠짐없이 수행할 수 있도록 접근 방법을 제공해줄 것이다.

IV. AI 스피커에 대한 보안성 평가

4.1 보안성 평가 기준을 활용한 취약점 점검

3장에서는 체계적인 절차에 따른 위협모델링을 통해 모든 AI 스피커에 적용 가능한 체크리스트를 도출하였다. 이 장에서는 3장에서 도출한 체크리스트를 이용하여 세계적으로 점유율이 높은 A사, B사와 국내에서 점유율이 높은 C사, D사의 AI 스피커를 대상으로 보안성 평가를 진행하였다. 이후 각 제조사별 취약한 부분을 비교·분석하고 모든 AI 스피커에 대한 전반적인 보안위협을 제기하였다.

보안성 평가 환경은 평가 내용이나 방법의 공정성 및 신뢰성을 위해 실제와 유사한 환경으로 구축하였다. 우선, AI 스피커는 최신 펌웨어로 업데이트 후 점검하였다. 스마트폰은 AI 스피커와 관련한 애플리케이션을 최신 버전으로 설치하였으며, 그 외 악성 파일 및 인증서를 설치하는 등 임의적인 환경은 구성하지 않았다.

4.1.1 애플리케이션

애플리케이션 분류의 보안성 평가 결과는 Table. 11과 같다. 이 중 A1, A2 항목은 애플리케이션 내 기능을 사용하여 점검하였으며, A3, A4, A7는 데이터 스니핑, MITM(Man In The Middle)[22]을 활용, A5, A6은 애플리케이션 코드의 정적 분석을 통해 점검하였다.

Table 11. Evaluation results of Application type

Type	No	A	B	C	D
Application	A1	O	O	O	O
	A2	O	O	O	O
	A3	X	O	X	O
	A4	O	O	O	O
	A5	O	O	O	O
	A6	O	O	O	O
	A7	O	O	O	O

Table 내의 O는 점검 결과, 보안성 문제가 없음을 나타내고, X는 보안성 문제가 존재함을 나타낸다. X가 표시된 A3은 개인정보 노출에 관한 항목이며, 점검 결과는 다음과 같다.

a) A3 : A, C사의 애플리케이션 내 로그인 과정에서 사용자의 ID, PW가 그대로 노출되는 취약점이 존재한다. Fig. 3은 해당 취약점에 대한 점검 결과를 보여준다.

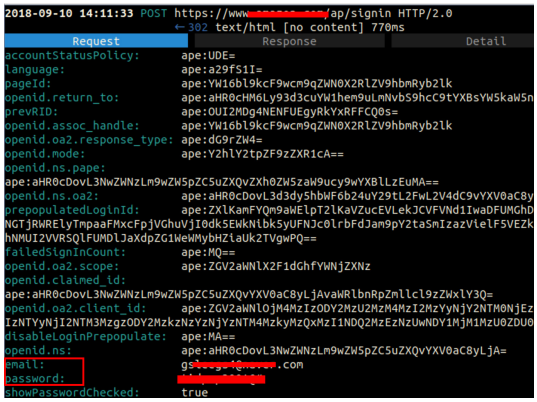


Fig. 3. Exposure of ID, Password

또한 A사의 경우, 애플리케이션 로그인 후 통화&메시지 기능을 이용할 시, 사용자가 가지고 있는 연락처 정보(이름, 전화번호 등)가 전부 노출되는 취약점이 존재하였다. Fig. 4는 탈취한 사용자 연락처 정보의 일부를 보여준다. 이러한 취약점은 사용자뿐만 아니라 사용자와 관련된 사람들에게도 피해가 확대될 수 있다.

각 제조사의 애플리케이션은 개인정보에 대한 보호 정책, 코드 난독화, SSL을 적용한 암호화 통신 및 피닝(Pinning)[23][24] 등 다양한 보호기법이 적용되어 있었다. 그러나 일부 제조사에서 MITM에 대한 취약점이 존재하였으며, 이로 인해 사용자의 계정정보, 연락처 정보 등 개인정보 노출 위협이 발생할 수 있다.

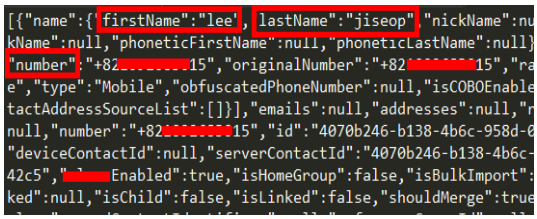


Fig. 4. User contact information

4.1.2 네트워크

네트워크 분류의 보안성 평가 결과는 Table. 12와 같다. 이 중, N1, N2는 포트 스캐닝을 이용해 점검하였으며, N3, N5는 스니핑, N4는 MITM, N6은 애플리케이션 코드 분석을 통해 발견한 API 활용, N7은 패킷을 가로챈 후 재전송하는 방법으로 점검하였다.

Table 12. Evaluation results of Network type

Type	No	A	B	C	D
Network	N1	O	O	O	O
	N2	O	O	O	O
	N3	O	O	O	O
	N4	O	O	O	X
	N5	O	O	O	X
	N6	O	O	O	O
	N7	O	O	O	O

문제점이 존재하는 항목 중 N4는 MITM 공격으로 인한 사용자의 민감한 정보 노출 항목이고 N5는 무선으로 업데이트되는 펌웨어 획득 여부 항목이다. 점검 결과는 다음과 같다.

a) N4 : D사에서는 애플리케이션 내 특정 기능을 이용해 AI 스피커에 음성을 전송할 수 있다. 이에 대한 취약점을 점검한 결과, Fig. 5와 같이 애플리케이션에서 기기로 전송한 사용자의 음성 데이터를 획득할 수 있었다. 사용자의 음성이 노출될 경우 개인정보 유포, 사생활 침해 등 2차 피해로 이어질 가능성이 존재한다.

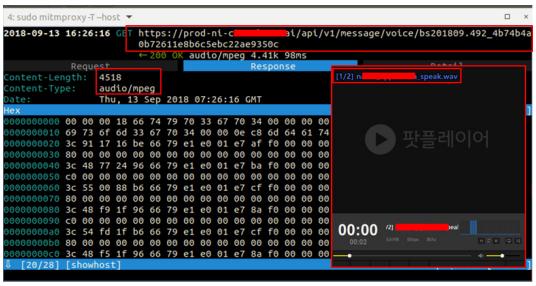


Fig. 5. Exposure of voice data

b) N5 : 스니핑을 통해 D사의 무선 업데이트 중인 펌웨어를 획득할 수 있었다. Fig. 6은 이와 관련한 그림을 나타낸다. 대부분의 AI 스피커는 기기의 주기적인 펌웨어 버전을 확인하고 자동으로 무선 업데이트를 진행한다. 그러나 만약 펌웨어가 노출될 경우, 공격자가 펌웨어 내 취약점을 이용해 도청 혹은 개인정보를 탈취하는 등 사용자의 개인정보가 노출될 위험이 존재한다.

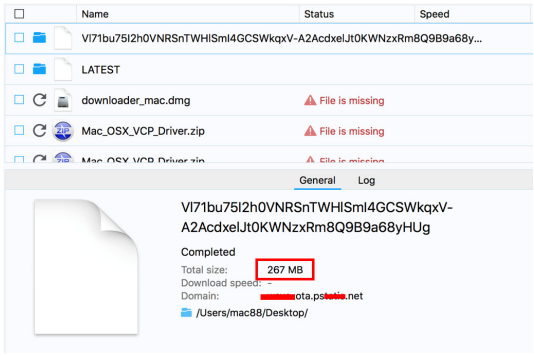


Fig. 6. Obtain firmware being wireless update

애플리케이션 - 서버 및 AI 스피커-서버 내 네트워크 통신에 대한 점검 결과, 대부분의 제조사는 불필요한 포트 미사용, SSL을 적용한 암호화 통신 및 피닝 등 보안위협에 대응하기 위한 방안이 적용되어 있었다. 그러나 D사의 경우, 일부 기능에서 통신 암호화 미적용, MITM에 대한 보호방안이 존재하지 않아 사용자의 음성 도청, 펌웨어 탈취 등의 위협이 발생할 가능성이 존재한다.

4.1.3 하드웨어와 시스템

UART, JTAG 디버깅 포트를 식별한 후 H1, H2, H3을 점검하였으며, 그 결과는 Table. 13과 같다.

Table 13. Evaluation results of Hardware type

Type	No	A	B	C	D
Hardware	H1	O	O	O	O
	H2	O	O	O	O
	H3	O	O	O	O

모든 제조사에서는 물리적 접근을 통한 AI 스피커 디버깅, 펌웨어 획득·수정을 방지하기 위해 UART, JTAG 포트의 접근 방안을 논리적으로 제거하였고, 이로 인해 H4, H5 항목 및 시스템의 모든 항목을 점검할 수 없었다. 위 결과처럼 모든 AI 스피커에서 디버깅 포트의 접근을 차단한다면 펌웨어 획득·수정, 기기 내 파일 시스템의 접근이 어려워, 이와 관련한 공격이 어려울 것으로 판단된다.

4.2 AI 스피커 취약점 비교·분석

각 제조사별 AI 스피커의 보안성 평가 결과, 취약

Table 14. Compare AI speaker vulnerabilities

Check list	A	B	C	D
Application A3	Account and Contact information	X	Account information	X
Network N4	X	X	X	Voice file exposure
Network N5	X	X	X	Firmware acquisition

점이 발견된 항목은 Table. 14와 같다. 체크리스트의 애플리케이션 분류에서는 인증정책의 개인 정보 노출(A3) 항목에서 취약점이 발견되었다. 그리고 네트워크 분류에서는 패킷 스니핑의 MITM 공격(N4), 펌웨어 획득(N5) 항목에서 취약점이 발견되었다.

AI 스피커에서 발견된 취약점을 분석한 결과, 선정한 기기 중 대부분은 애플리케이션의 개인 정보 노출 취약점을 포함하여 공통적으로 네트워크 공격 중 스니핑, MITM 공격에 취약하였다. Table. 14의 애플리케이션의 개인 정보 노출 또한 MITM 공격으로 인해 발생한 결과이다. 이러한 취약점은 선정한 기기 외 여러 AI 스피커에서도 존재할 수 있다고 판단되며, 이를 통해 사용자의 개인정보가 유출 및 악용될 수 있는 심각한 문제가 발생할 수 있다.

V. 결 론

본 논문에서는 STRIDE, LINDDUN 위협모델링을 적용하여 취약점 및 개인정보보호 관점으로 AI 스피커의 보안성 평가 기준을 도출하고, 이를 활용하여 실제 기기에 대한 보안성 평가를 수행하였다. 그 결과, 선정한 AI 스피커에서는 사용자의 개인 정보, 무선 업데이트 중인 펌웨어, 음성 파일 등이 노출되는 문제가 존재하였다. 마지막으로 보안성 평가 결과를 기반으로 AI 스피커에서 발생 가능한 위협을 비교·분석하였으며, 대부분의 기기에서 MITM 공격 위협이 발생할 수 있을 것이라는 결과가 도출되었다.

MITM 공격은 MAC, IP 등을 속이고 사용자-서버 간의 송·수신 패킷을 가로채어 엿듣는 행위이다. 이를 통해 사용자의 연락처, 음성과 같은 민감한 개인정보가 노출될 수 있는 심각한 문제가 발생할 수 있다. 또한 펌웨어 유출로 인해 기기 자체의 취약점을 이용한 공격을 시도하는 상황이 발생할 수도 있다. 따라서 AI 스피커에 SSL 암호화 통신이 기본적으로 적용되어 있다는 전제 하에 인증서 및 공개키 피닝 혹은 Expect-CT(Certificate Transparency, 인증서 투명성)[25]를 적용한다면 MITM 공격으로 인한 개인정보 유출 피해를 완화할 수 있을 것으로 기대된다.

AI 스피커는 각 제조사별로 지원하는 기능이 다르지만 음성 인식 및 일정, 알람, 메모 등 사용자의 개인정보와 관련된 공통된 기능이 포함되어 있다. 따라서 본 논문에서 제시한 보안성 평가 기준을 활용한다면 AI 스피커 제조사의 제품에 알맞은 보안 체크리스트를 구현할 수 있을 것으로 기대된다.

그러나 본 연구 결과는 실제 제품 개발과 관련한 이해관계자들의 참여가 제한되어 있으며, 알려진 취약점 및 공개된 자료를 기반으로 위협을 식별 및 분석하였다. 그러므로 AI 스피커 개발과 관련된 이해관계자들이 참여하여 잠재적인 취약점까지 고려하도록 보안성 평가 기준을 도출하는 것이 향후 과제로 남아있다.

References

[1] S&P Global Market Intelligence, https://www.spglobal.com/marketintelligence/en/news-insights/research/smart-speakers-take-off, Sep. 2018

[2] CNBC, https://www.cnbc.com/2018/ 05/24/amazon-echo-recorded-conversation-sent-to-random-person-report.html

[3] Nicholas Carlini, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner and Wenchao Zhou, "Hidden Voice Commands", 25th *USENIX Security Symposium*, 2016

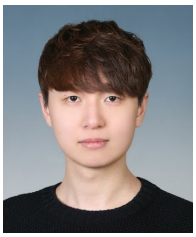
[4] The Hacker News, https://thehackernews.com/2018/04/amazon-alexa-hacking-skill.html, Aug. 2018

[5] Tencent Blade Team, "Breaking Smart Speaker - We are Listening to you", *DEFCON26*, 2018

[6] Chun Yu Cheung, "Threat Modeling Techniques", *Delft University of Technology, the Netherlands*, 2016

〔7〕 Adam Shostack, "Threat Modeling", *WILEY*, pp. 109-160, 2014
〔8〕 Microsoft, https://docs.microsoft.com/ko-kr/azure/security/azure-security-threat-modeling-tool-threats, Aug. 2018
〔9〕 Eddington, Michael, Brenda Larcom, and Eleanor Saitta, "Trike v.1 Methodology Document", *Octotrike*, Jul. 2012
〔10〕 DistriNet, https://linddun.org/linddun.php, Aug. 2018
〔11〕 "Process for Attack Simulation and Threat Analysis", *InfosecurityEurope*, 2014
〔12〕 Hyunji Chung, Michaela lorga, Jeffrey Voas and Sangjin Lee, "Alexa, Can I Trust You?", *IEEE Computer*, pp.100-104, 2017
〔13〕 Ike Clinton, Lance Cook and Dr. Shankar Banik, "A Survey of Various Methods for Analyzing the Amazon Echo", *The Citadel, The Military College of South Carolina*, 2016
〔14〕 Jonas Zaddach and Andrei Costin, "Embedded Devices Security and Firmware Reverse Engineering", *BlackHat*, 2013
〔15〕 Armis Lab, "Exploiting BlueBorne in Linux-based IoT devices", pp. 22-30, 2017
〔16〕 NIST, https://nvd.nist.gov/vuln/detail/CVE-2018-9070, Aug. 2018
〔17〕 Guoming Zhang, Chen Yan and Xiaoyu Ji, "DolphinAttack:Inaudible Voice Commands", *ACM SIGSAC Conference on Computer and Communications Security*, pp.103-117, 2017
〔18〕 William Haack, Madeleine Severance, Michael Wallace and Jeremy Wohlwend, "Security Analysis of the Amazon Echo", *MIT University*, 2017
〔19〕 Mary Bispham, "Security Vulnerabilities in Speech Recognition Systems", *OXFORD University*, 2016
〔20〕 Medium, https://medium.com/@micaksica/exploring-the-amazon-echo-dot-part-1-intercepting-firmware-updates-c7e0f9408b59, Aug. 2018
〔21〕 Microsoft, https ://docs.microsoft.com/en-us/previous-versions/msp-n-p/ff648644(v=pandp.10), Sep. 2018
〔22〕 Blackhat, https://www.blackhat.com/presentations/bh-europe-03/bh-europe-03-valleri.pdf, Jul. 2018
〔23〕 OWASP, https://www.owasp.org/index.php/Certificate_and_Public_Key_Pinning, Oct. 2018
〔24〕 Mahesh Bhor and Dr. Deepak Karia, "Certificate pinning for Android Applications", *International Conference on Inventive Systems and Control*, Oct. 2017
〔25〕 IETF Tools, https://tools.ietf.org/html/dra ft-ietf-httpbis-expect-ct-02, Oct. 2018

〈저자소개〉



이 지 섭 (Jiseop Lee) 학생회원
2016년 2월: 조선대학교 컴퓨터공학부 학사
2016년 3월~현재: 고려대학교 정보보호대학원 석사과정
〈관심분야〉 정보보호, 취약점 분석, 보안성분석평가

강 수 영 (Soo-Young Kang) 학생회원
2006년 2월: 순천향대학교 컴퓨터공학부 공학사
2008년 2월: 순천향대학교 컴퓨터공학부 공학석사
2008년 5월~2010년 10월: 한국인터넷진흥원(KISA) 연구원
2010년 10월~2014년 10월: 안랩(Ahnlab) 주임연구원
2013년 3월~현재: 고려대학교 정보보호대학원 박사과정
〈관심분야〉 보안성 평가/인증, 위협 모델링, 소프트웨어 보안

김 승 주 (Seungjoo Kim) 종신회원
1994년~1999년: 성균관대학교 정보공학과(학사, 석사, 박사)
1998년~2004년: 한국인터넷진흥원(KISA) 팀장
2004년~2011년: 성균관대학교 정보통신공학부 부교수
2011년~현재: 고려대학교 사이버국방학과/정보보호대학원 정교수
2017년~현재: 고려대학교 사이버무기시험평가연구센터(CW-TEC) 부센터장
2004년~현재: 한국정보보호학회 이사
2007년: 국가정보원장 국가사이버안전업무 유공자 표창
2010년: 방송통신위원회 정보통신망 침해사고 민관합동조사단 위원
2011년~현재: (사)화이트해커연합 HARU 및 국제해킹대회 SECUINSIDE 설립자 및 이사
2012년: 선관위 디도스 특별검사팀 자문위원
2014년~2015년: 육군사관학교 초빙교수
2014년~2016년: 다음카카오 프라이버시 정책 자문위원회 위원
2015년~현재: 방위사업청 방산기술보호 자문관
2016년~2018년: 개인정보분쟁조정위원회 위원
2016년~현재: 산업통상자원부 전략물자기술 자문위원
2016년~현재: 한국카카오뱅크 정보보호부문 자문교수
2017년~현재: 국방보안연구소 정보보호분야 자문위원
2017년~현재: 여신금융협회 신용카드 단말기 시험 인증위원회 위원
〈관심분야〉 보안공학 및 SDL, 위협 리스크 모델링, 보안성 평가/인증, 암호학, Usable Security