군 사 과 학 연 구 지
제14권 제2호(2021년 12월)

음성인식 기반의 상용서비스에 대한 사이버 위협 분석

Cyber Threat Analysis for Voice Recognition-based Commercial Services

홍천호[1] · 조영호[2]
Cheonho Hong · Youngho Cho

ABSTRACT

ASR(Automatic Speech Recognition) technology is greatly attracting attentions in the era of the 4th industrial revolution and will be widely adopted in various fields of defense and military area. Meanwhile, concerns about cyber threats on voice recognition-based (or speech recognition-based) commercial services without ASV(Automatic Speaker Verification) are also increasing. According to existing studies on security of ASR techniques, researches on devising and advancing ASR techniques are being actively conducted, but researches on attack models and scenarios against voice recognition services are not being conducted in a variety and depth. In this study, we propose a formal attack model with three steps against a common voice recognition-based commercial service, conduct experiments to show that the proposed attack model implemented by using a synthetic voice tool (Prosody) can partially control commercial services (or systems) auch as AI speaker, air purifier and a silent vacuum cleaner connected to the AI speaker, and a commercial vehicle. We also discuss possible cyber threats based on the experimental results. Consequently, we inform the seriousness of such cyber threats on ASR-based commerical services when they do not have appropriate security countermeasures such as ASV.

Key Words: Automatic Speech Recognition, Automatic Speaker Verification, Speech Synthesis, AI Speaker, Cyber Threat Analysis

논문접수일 : 2021년 11월 11일, 심사일 : 2021년 11월 30일, 게재확정일 : 2021년 12월 20일

1)국방대학교 컴퓨터공학/사이버전협동과정 석사과정
2)국방대학교 컴퓨터공학/사이버전협동과정 주임교수 / 교신저자(Corresponding author)

1. 서 론

국방부는 주변국들의 경쟁적 군비 증강과 북한의 지속적인 위협 그리고 격변하는 안보 환경 속에서 강한 군대를 만들기 위해 4차 산업혁명 시대의 최첨단 기술을 접목한 '스마트 국방혁신'을 추진하고 있다[1]. 4차 산업혁명 시대를 맞이하여 인공지능(Artificial Intelligence: AI) 기반의 서비스가 다양한 분야에 접목되어 사용되고 있으며, 국방 분야에서는 차세대 전투기인 F-35에 음성인식(Automatic Speech Recognition: ASR) 기능을 탑재하여 사용되고 있다[2]. 또한, 고령화 속에서 노인복지 문제의 대안으로 음성인식 서비스가 제시되기도 했다[3].

과학기술정보통신부는 '2020 인터넷 이용 실태조사 결과'를 3월에 발표하였다[4]. 결과에서 우리나라의 인터넷 이용자는 약 4,600만 명으로 만 3세 이상 91.9%가 이용하고 있고, 이 중 28.5%는 AI 음성인식 서비스를 이용하고 있다. 이는 2019년 대비 3.3%가 증가한 수치였다. 사람에게 친숙한 정보전달 방식은 별도의 훈련 없이도 AI 음성인식 서비스와 소통할 수 있으며, 이동 및 업무 중에도 사용할 수 있다는 편리성으로 점차 사용자가 증가할 것으로 예상된다.

하지만 음성인식 서비스에 대한 보안 문제도 지속해서 지적되고 있다.

첫째는 음성인식 서비스 장치 자체의 보안 취약점에 대한 지적이다. 음성인식 서비스는 음성만으로 개인정보 및 음성인식 서비스와 연결된 사물을 제어할 수 있다. 이는 편리하게 다양한 서비스를 사용할 수 있다는 장점이 있으나, 음성데이터를 수집하는 기업으로부터 발생할 수 있는 개인정보 유출 및 프라이버시 침해를 발생한다. 2020년 10월 미국에서 통신망에 침투해 정보를 빼돌리는 소위 '백도어' 보안 문제로 사용을 금지한 화웨이 칩이 사용된 AI 스피커가 군 내부에 설치 및 사용되고 있어 보안 문제로 지적되기도 하였다[5].

둘째는 음성인식(ASR) 기능을 악용하거나 자동화자 식별(Automatic Speaker Verification: ASV) 체계를 우회하는 공격의 위험성이다. 사람의 말소리를 흉내 내는 성대모사나 기계로 말소리를 재연하는 음성합성 및 음성 변환을 통한 사이버 공격과 관련된 연구들이 있다[6, 7]. 사이버 공격의 형태는 네트워크를 이용한 공격뿐만 아니라 음성인식 서비스를 활용한 도청 및 음성데이터 수집도 사이버 위협 및 사이버 공격으로 활용될 수 있다.

본 연구에서는 ASV와 같은 특별한 보안체계를 갖추지 않은 다양한 음성인식 기반 서비스에 대한 사이버 위협 분석을 통해 잠재적 위험성을 알리고 경각심을 제고하는 것을 목적으로 하며, 연구 측면에서는 다음과 같은 공헌도가 있겠다.

첫째, 일반적인 음성인식 서비스에 대한 3단계로 동작하는 사이버 공격 모델을 제안하고 사이버 위협을 확인하는 분석 도구로 활용한다[8, 9].

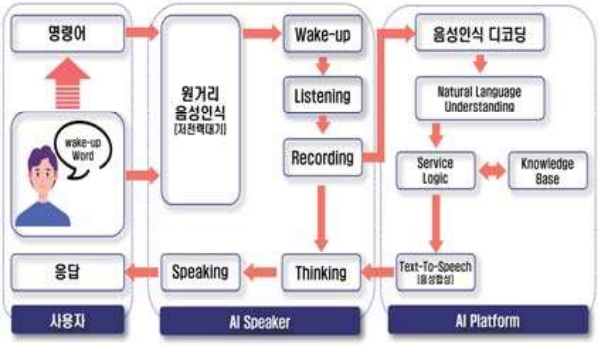
둘째, 음성 합성도구(Prosody)를 활용하여 음성인식 서비스와 AI 스피커가 연결된 가전제품 공기청정기, 무선 청소기 그리고 음성인식 기능이 탑재된 자동차의 공조시스템과 편의장치를 작동시켜 일부 제어권을 확보할 수 있음을 보이고 이를 통해 발생 가능한 사이버 위협 시나리오를 제시한다.

이후 논문의 구성은 다음과 같다. 2장에서는 배경지식 및 관련 연구를 살펴본다. 3장에서는 음성인식 서비스에 대한 공격 시나리오와 공격모델을 제안한다. 4장에서는 실험 결과를 제시하고, 5장에서는 향후 연구 방향과 함께 결론을 맺는다.

2. 배경지식 및 관련 연구

2.1 음성인식 기술

음성인식 기술을 이해하기 위해 AI 스피커의 음성인식 및 처리 절차를 설명한다. 음성인식은 사람과 음성인식 기기 간의 통신 과정이며, 주로 사람의 음성을 통한 정보전달 방식이다.



<그림 1> AI 스피커 구조 및 흐름도

<그림 1>은 AI 스피커의 구조 및 흐름을 나타낸 것이다. 화자가 발화하게 되면 먼저 AI 스피커가 반응한다. AI 스피커는 입력 및 출력을 담당하는 인터페이스로 사용자의 음성을 입력받게 되면 서버로 전달되기까지의 전처리 과정을 AI 스피커가 처리하게 된다.

즉, 사람의 음성 신호에서 정보를 추출하여 컴퓨터가 이해할 수 있는 문자 정보로 바꾸는 과정을 하게 되며, 아날로그 신호인 사람의 음성데이터를 전처리 과정을 통해 디지털 신호로 변환하게 된다. 이후 클라우드에서는 디코딩 과정부터 자연어 이해(Natural Language Understanding: NLU) 및 음성합성(Text to Speech: TTS)을 처리한다. 합성된 음성은 AI 스피커를 통해 출력된다[8].

2.2 관련 연구

음성인식 관련 사이버 위협에 관한 기존 연구들은 음성인식 장치 자체의 보안 취약성과 음성데이터를 수집하고 처리하는 과정에서 발생할 수 있는 개인정보 유출 및 프라이버시 침해 그리고 음성인식의 취약점을 예방할 수 있는 대응 기법에 관한 연구들로 대표적인 연구는 다음과 같다.

Nicholas Carlini 등[8]은 사람이 인식하지 못하는 음역의 명령어와 이해할 수 없는 음성으로 음성인식 서비스를 공격하는 모습을 보여주면서 공격기법을 설명했다. 하지만 ASV 기능이 없는 음성인식 서비스에 대한 제한적 실험이었다.

Takeshi Sugawara 등[9]는 AI 스피커에 탑재된 미세전자기계시스템(Micro-Electro Mechanical Systems: MEMS)이 레이저 빛에 반응하면서 음성명령을 인식하는 진동판이 똑같이 떨리게 되는 방식으로 레이저 빛에 암호화된 명령어를 인식하고 작동하는 공격기법이다.

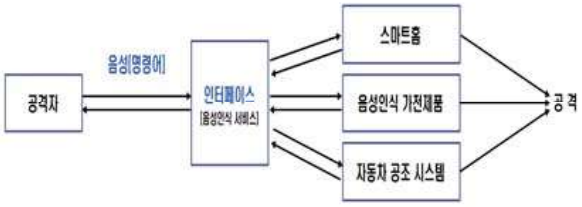
Ji-seop Lee 등[10]은 STRIDE(Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege: STRIDE)와 LINDDUN (Linkability, Identifiability, Information Disclosure, content Unawareness, policy and consent Non compliance: NINDDUN) 등 위협 모델링을 적용하여 AI 스피커의 보안성 평가를 진행하였다. 클라우드 기반 음성인식 시스템은 정보처리 간 노출되는 정보와 사용자와 서버 간의 송·수신되는 패킷을 중간에 공격자가 가로채는 방법으로 공격할 수 있음을 증명하였으며, 극복하는 방안으로 인증서 및 공개키를 적용한다는 한정된 대응 방법을 제시하였다.

Massimiliano Todisco 등[6]은 ASVspoof 2019에서 스푸핑 위협으로부터 ASV 체계를 보호하기 위한 대응책을 마련하는 연구를 진행했으며, 스푸핑 시나리

오와 합성 및 변환 그리고 재생 음성이라는 세 가지 주요 스푸핑 공격에 관한 실험 결과 및 대응 방법을 제시했다.

3. 음성인식 서비스에 대한 공격 모델 제안

본 장에서는 음성인식 서비스 대상 공격 모델을 설명한다. 음성인식 서비스에 대한 보안 문제들은 지속해서 지적받고 있지만 개선되고 있지 않은 상황이다[11]. <그림 2>는 음성인식 서비스에 대한 공격모델을 설명한다. 제안하는 공격모델에는 3가지의 주요 구성 요소가 있다. 1) 공격자, 2) 음성(공격 방법) 3) 공격 대상의 인터페이스이다.



<그림 2> 음성인식 서비스에 대한 공격모델

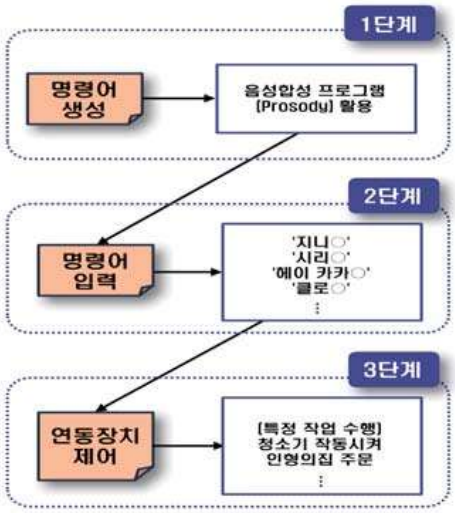
지금부터 제안하는 모델의 작동 방법에 대하여 단계적으로 설명한다. 먼저 공격자는 공격 대상이 되는 음성인식 장치에 명령어 입력을 위해 접근할 수 있다고 가정한다.

<그림 3>은 음성인식 서비스를 대상으로 하는 공격모델의 단계별 공격 절차를 구분한 것이다.

1단계에서 공격자는 음성인식 서비스를 공격하기 위해 명령어를 생성한다. 이때, 음성인식 서비스가 음성합성 프로그램을 이용하여 생성된 합성음성에도 작동하는 것을 입증하기 위해 음성합성 프로그램 Prosody[12]를 이용하여 명령어를 생성했다.

2단계는 공격자는 공격 대상의 음성인식 서비스에 접근하고, 생성된 명령어를 스마트폰 또는 노트북의 음악 재생플레이어를 이용하여 재생 공격한다.

3단계는 음성인식 서비스와 연결된 장치 또는 가전제품들을 제어하고, 기타 명령에 따라 특정 작업을 수행할 수 있다[13, 14].



<그림 3> 공격모델의 단계별 공격 절차

3.1 공격 시나리오

이번 절에서는 ASV 기능이 포함되지 않은 음성인식 서비스 대상 사이버 공격 수행 방법과 제안 모델이 어떻게 적용될 수 있는지 공격 시나리오를 통해 설명한다.

<그림 4> 음성인식 서비스 공격 시나리오

<그림 4>는 음성인식 서비스에 대한 음성합성 공격을 수행하는 공격 시나리오를 나타낸다. 공격자는 합성된 음성으로 음성인식 인터페이스인 AI 스피커에 명령어를 입력한다. 입력된 명령어를 통해 AI 스피커와 연결된 가전제품 또는 스마트홈(smart home) 시스템을 제어할 수 있다. 같은 방법으로 자동차에 탑재된 음성인식 서비스에 명령어 입력이 가능하고 명령어를 통해 자동차 공조시스템 또는 편의장치를 조작할 수 있다. ASV 기능이 포함되지 않은 음성인식 서비스는 사이버 공격을 방어할 수 없으며, 보안에 매우 취약하다.

4. 실험 결과 및 분석

4.1 실험 목적과 방법

실험의 목적은 ASV 기능이 포함되지 않은 음성인식 서비스를 대상으로 한 사이버 공격 위협을 실험을 통해 보이는 것이다. 음성인식 서비스에 보안인증 기능이 없을 때 발생할 수 있는 위협을 통해 보안인증 강화할 필요가 있다고 판단된다면 제안 모델 및 실험의 목적은 달성됐다고 판단할 수 있다.

실험의 환경은 <그림 5>와 같이 3단계로 나눠서 구성하였다. 명령어 생성단계 및 입력단계를 위해 Intel i7 10th. RAM 16GB, LG Gram 노트북을 활용하였으며, 명령어 생성을 위해 음성합성 프로그램 Prosody를 사용하였다.

<그림 5> 실험 수행 환경

마지막 명령어 입력단계에서 사용한 음성인식 서비스는 AI 스피커 GIGA Genie2를 사용하였으며, AI 스피커와 연결된 공기청정기와 무선 청소기를 특정 작업 수행 가능 여부를 입증하기 위해 사용하였다. 또한, 음성인식 기능을 탑재한 자동차 중 현대의 2020 그랜저를 대상으로 했다.

<그림 6> 실험 세부 수행 절차

실험 단계별 수행 결과는 다음과 같으며 세부 수행 절차는 <그림 6>과 같이 설명할 수 있다.

- **(1단계) 명령어 생성단계:** 공격자는 음성인식 서비스에 명령어 입력을 위해 음성합성 프로그램 Prosody를 이용하여 Wake-Up Word(예: 지니야)와 음성인식 이후 특정 작업(예: 공기청정기 작동시켜줘, 무선 청소기 작동시켜줘)을 수행하기 위해 명령어를 추가로 생성한다.
- **(2단계) 명령어 입력단계:** 음성인식 서비스에 명령어를 입력하기 위해 스마트폰 또는 노트북을 이용하여 명령어를 저장하고 음악 재생플레이어를 이용하여 명령어를 재생한다.
- **(3단계) 음성인식 서비스 작동 단계:** 음성인식 서비스 작동 단계는 최초 음성인식 이후 '공기청정기 작동시켜줘', '무선 청소기 작동시켜줘'와 같은 특정 작업을 위한 명령어를 추가로 입력하고, 작동 여부를 확인한다. 이후, 확

보한 제어권을 활용하여 공격자의 의도에 따라 해당 음성인식 서비스 또는 장치를 악의적으로 작동시켜 공격을 수행한다.

4.2 실험 결과 및 분석

실험 수행 결과, ASV 기능이 포함되지 않은 음성인식 서비스는 합성된 음성으로도 명령어 입력 및 연결된 장치들을 제어할 수 있었다.

- **(1단계) 명령어 생성단계:** 공격자는 음성인식 서비스를 공격하기 위해 <그림 7>과 같이 음성합성프로그램 Prosody를 사용하여 성우 선택한다. 이후 텍스트를 입력해서 Wake-Up Word(예: 지니○) 및 추가로 특정 작업을 수행하기 위해 명령어를 생성한다. 무료로 생성할 수 있는 음성합성 프로그램은 공격자가 원하는 명령어를 쉽게 생성할 수 있다.

<그림 7> 실험 수행 결과: 정상 작동

- **(2단계) 명령어 입력단계:** 공격자는 음성인식 서비스를 공격하기 위해 <그림 8>과 같이 생성된 명령어를 재생하여 공격한다. AI 스피커의 경우 Wake-Up Word에 정상적으로 반응을 하였으며, 자동차도 재생된 음성에 대해 정상적으로 반응하였으나, 물리적 버튼을 눌러야만 음성인식이 가능했다.

이는 ASV 기능이 없는 음성인식 서비스는 명령어 공격을 방어할 수 없고 보안에 취약하다는 것이다. 공격자가 음성인식 서비스에 접근만 할 수 있다면 공격 대상의 음성인식 서비스를 이용 다양한 공격이 가능한 것이다.

<그림 8> 음성인식 서비스 명령어 입력

- **(3단계) 음성인식 서비스 작동 단계:** 공격자는 특정 작업을 수행하기 위해 Wake-Up Word 입력 이후 추가적인 명령어를 <그림 9>와 같이 입력했으며, 정상적으로 AI 스피커와 연결된 가전제품들이 작동하였다.

<그림 9> 공격 수행 절차

이번 실험에서 AI 스피커와 연결된 가전제품은 공기청정기와 무선 청소기로 한정하여 진행했으나, <그림 10>과 같이 AI 스피커와 블루투스 또는 와이파이로 연결된 가전제품 또는 스마트 기기들도 제어할 수 있다.

<그림 10> AI 스피커와 연결된 가전제품

<그림 11>과 같이 자동차도 물리적 버튼을 누르고 추가 명령어를 입력했을 때 에어컨 및 공기 청정 모드를 작동할 수 있으며, 음성으로 차량을 창문을 내리거나 올리는 명령이 가능했다.

<그림 11> 음성명령으로 자동차 공조시스템 제어

4.3 예상 공격 시나리오 및 토의

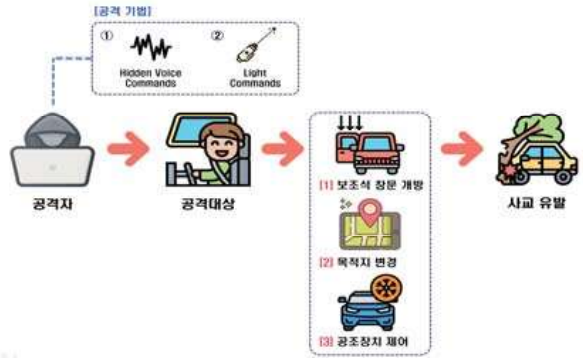
제안 공격기법으로 확보한 음성 서비스에 대한 제어 권한으로 공격자가 수행 가능한 공격 시나리오와 위험 상황의 예는 다음과 같다.

상황 1: 공격자는 공격 대상에게 금전적 손해를 입힐 수 있다. 음성인식 서비스를 이용하여 물품 구매 또는 가전제품을 작동시켜 전력 소비를 유발한다.

상황 2: 공격자는 공격 대상에게 물질적 손해를 유도할 수 있다. 음성인식 서비스와 연결된 장치 중 난방시설 또는 전열기를 작동시켜 기기의 과부하를 유도하여 화재를 유발한다.

현재 AI 음성인식 기술은 지속해서 발전 중이며, 적용 범위가 넓어지고 있다. 하지만 서비스에 대한 보안인증 및 개인정보 보호를 위한 노력이 필요하다.

실험 결과 합성된 음성으로 음성인식 서비스를 쉽게 정상 작동시킬 수 있다. 이를 통해 우리는 기존 연구를 접목한 공격 시나리오를 제시할 수 있다. 자동차의 경우 AI 스피커와 달리 물리적 버튼을 누르고 음성을 입력해야 하지만 기존 연구 [8, 9]와 같이 사용자가 인식하지 못하는 상황에서 사이버 공격을 수행한다면 <그림 13>과 같은 위협이 될 수 있다. 사람이 인지하지 못하는 음역의 명령어 공격과 레이저 빛을 이용한 공격은 공격자에게 있어 사용자가 공격을 인지하지 못하는 상황에서 공격할 수 있다는 장점이 있다. 이러한 공격기법은 사용자가 운전 중 사고를 유발할 수 있는 공격모델을 구성할 수 있다.



<그림 11> 기존 연구를 적용한 공격 시나리오

국방 분야에서 현재 음성인식 기능을 사용하고 있는 차세대 전투기 F-35의 경우 전투기의 임무 수행 간 사람이 인지하지 못한 음역의 명령어 공격 또는 레이저를 이용한 공격을 할 수 있는 위의 공격 시나리오를 적용해 본다면 군의 임무 수행에 있어 위협으로 다가올 수 있을 것이다.

5. 결 론

본 연구에서는 4차 산업혁명 시대에 주목받고 있는 기술인 음성인식을 적용한 서비스를 AI 스피커와 자동차 그리고 이와 연결된 공기청정기, 무선 청소기 그리고 자동차 내 편의시설 및 공조시스템으로 한정하고 실험을 통해 공격 시나리오 및 모델을 제안했다.

또한, 기존 연구를 통해 공격 가능성이 입증된 공격기법을 적용한 공격 시나리오를 제시했다. 이는 향후 국방 분야에 다양하게 적용될 수 있는 음성인식 기술에 대해 보안 취약점을 사전에 인식하고 보안 강화를 위한 방법을 구상하기 위함이며, 보안 위협을 낮추기 위한 제안이다. 하지만 조금 더 다양한 공격기법과 이를 대응 하기 깊이 있는 연구가 필요하다.

향후 연구계획은 다음과 같다. 우선 본 연구에서는 ASV 기능이 포함되지 않은 음성인식 서비스를 대상으로 실험을 진행하였고, 사이버 공격 위협에 대해 언급하였다. 다음 연구에서는 음성인식 서비스의 보안 강화를 위해 ASV 기능을 포함한 음성인증 시스템이 아닌 AI 스피커 자체적으로 사용자와의 대화를 통해 사용자를 인지하고 확인할 방법을 연구할 계획이다. 이는 사용자 인증을 위해 생체인증을 이용하는 것에서 사용자의 대화 습관 및 패턴을 습득하고 누적된 데이터를 이용한 화자 인식 방법을 강구 할 예정이다.

참 고 문 헌

[1] 대한민국 국방부, *국방개혁 2.0*, 국방개혁실, 2019

[2] Seongwoo Kim, Chulsu Shin, BongGyu Kim, "A Study on Fighter Airplane's Voice Command Recognition System Design and Verification Environment," The Korean Society for Aeronautical & Space Sciences, pp.327-331, 2012.11.

[3] JungWon Kim, YouJin Song, YongjunSung, SejungMarina Choi, "AI Speaker for the Elderly : Functional and Emotional Evaluation of AI Speaker," Journal of Media Economics & Culture 18(4), pp.7-35, 2020.11.

[4] Ministry of Science and ICT, 「*2020 Internet Use Survey Results*」, 2021.

[5] TV CHOSUN 뉴스, "[단독] 軍 AI 스피커 4만여대에 美 금지한 화웨이칩...軍 "몰랐다"
(검색일: 2021.11.09.)

[6] Serife Kucur Ergünay, Elie Khoury, Alexandros Lazaridis and Sébastien Marcel, "On the vulnerability of speaker verification to realistic voice spoofing," IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS), pp.1-6, 2015

[7] Massimiliano Todisco, Xin Wang, Ville Vestman, Md Sahidullah, Hector Delgado, Andreas Nautsch, Junichi Yamagishi, Nicholas Evans, Tomi Kinnunen, and Kong Aik Lee, "ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection," arXiv preprint arXiv:1904.05441v2, 2019.

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

[9] Takeshi Sugawara, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu, "Light Commands: Laser-Based Audio Injection Attacks on Voice-Controllable Systems," 29th USENIX Security Symposium, 2020

[10] Ji-seop Lee, Soo-young Kang, Seung-joo Kim, "Study on the AI Speaker Security Evaluations and Countermeasure," Journal of the Korea Institute of Information Security & Cryptology, Vol.28, No.6, 2018

[11] 이투데이, "[2021 국감] AI 스피커 가입자 1600만 명 넘었는데…보안인증 '0'건"
(검색일: 2021.11.9.)

[12] 음성합성 프로그램 Prosody 관련 링크, 인터넷 홈페이지
(검색일: 2021.11.9.)

[13] 경향신문, "'음성 쇼핑' 생소하다고요? 미국선 일상이래요"
(검색일: 2021.11.9.)

[14] Song, Ji Sung · Kang, Song Hee, "A Study on the Service Usability of IoT Smart Devices," JOURNAL OF THE KOREAN SOCIETY DESIGN CULTURE 25(2), p.332,

저 자 소 개

홍천호(E-mail: hongcheonho@gmail.com)
2006년 침례신학대학교 졸업(학사)
현재 국방대학교 컴퓨터공학/사이버전협동 석사과정
관심분야 : 네트워크 보안, 디지털 포렌식 등

조영호(E-mail: youngho@kndu.ac.kr)
1998년 공군사관학교 졸업(학사)
2006년 연세대학교 졸업 (석사)
2013년 미국 University of Maryland, College Park 졸업 (박사)
현재 국방대학교 컴퓨터공학/사이버전협동전공 주임교수
관심분야 : 네트워크 보안, 스테가노그래피, 신뢰메커니즘, 블록체인, 디지털 포렌식, 적대적 AI 등