Journal of The Korea Institute of Information Security & Cryptology(JKIISC) ISSN 1598-3986(Print)
VOL.23, NO.4, August 2013 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2013.23.4.617

소리를 이용한 릴레이 공격 공격의 탐지*

김 종 욱,† 강 석 인, 홍 만 표‡
아주대학교

Detecting a Relay Attack with a Background Noise*

Jonguk Kim,† Sukin Kang, Manpyo Hong‡
Ajou University

요 약

NFC, RFID 등의 무선 기술이 발전하면서 기기간의 데이터 전달이 용이해지고 있다. 사용자는 여러 개의 복잡한 비밀번호를 외우고 입력하는 대신 항상 소지하고 있는 카드나 휴대폰을 이용하여 간편하게 자신을 인증함으로써 건물에 출입하거나, 보안 자료에 접근하고, 결제를 할 수 있게 되었다. 그러나 최근 릴레이 공격(relay attack)의 출현으로 편리한 토큰 기반(something you have) 인증의 안전성이 위협받고 있다. 릴레이 공격은 안전한 통신 채널을 가진 두 기기 사이에서도 효과적으로 공격을 성공시킬 수 있고, 공격의 원리가 복잡하지 않아 쉽게 구현이 가능하다. 본 논문에서는 거리 제한 방식(distance bounding)이나 위치 측정 후 비교와 같은 기존 방어 기법과 다른 청각 채널을 통한 릴레이 공격 탐지에 대해 제안한다.

ABSTRACT

Wireless communication technology such as NFC and RFID makes the data transfer between devices much easier. Instead of the irksome typing of passwords, users are able to simply authenticate themselves with their smart cards or smartphones. Relay attack, however, threatens the security of token-based, something-you-have authentication recently. It efficiently attacks the authentication system even if the system has secure channels, and moreover it is easy to deploy. Distance bounding or localization of two devices has been proposed to detect relay attacks. We describe the disadvantages and weakness of existing methods and propose a new way to detect relay attacks by recording a background noise.

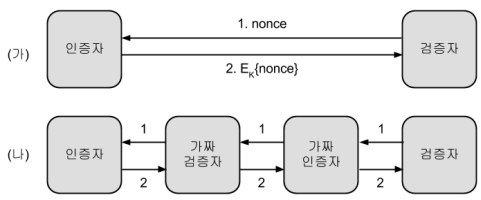
Keywords: Authentication, Man-in-the-middle Attack, Relay Attack, NFC, Audio Fingerprinting

I. 서 론

RFID(radio frequency identification)에 대한 연구에 이어 최근에는 근거리 통신(near field communication)에 대한 연구가 진행되고 있으며 대부분의 스마트폰에 기본 기능으로 탑재되고 있다. 간편한 결제나 인증에 주로 활용되고 있는 근거리 통신은 10cm 내외의 거리에서 통신이 발생한다.

근거리 통신에서는 두 기기의 사이에서 중간자 공격(man-in-the-middle attack)이 발생하기 어려운 것으로 알려져 있다[1]. 두 기기가 물리적으로 극히 가까운 곳에 위치하기 때문에 중간에 위치하여 상대를 속이기 어렵기 때문이다. 근거리 통신 표준에서도 중간자 공격이 발생하지 않을 것으로 가정하여 상호 키 교환 시 디피-헬만 키 교환 기법[2]을 그대로 이용한다[3]. 디피-헬만 키 교환 기법은 안전하지 않은 채널을 통해서 안전하게 키를 교환할 수 있는 방법

접수일(2013년 7월 2일), 수정일(2013년 7월 30일), 게재확정일(2013년 7월 30일)
* 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2011-0011289)
† 주저자, kju@ajou.ac.kr
‡ 교신저자, mphong@ajou.ac.kr(Corresponding author)



〔그림 1〕 (가) 정상적인 인증 (나) 릴레이 공격

이나 상호 미리 보유한 비밀 정보(pre-shared secret)를 가지지 않을 경우 중간자 공격에 취약하다〔4〕. 그럼에도 불구하고 근거리 통신에서는 중간자 공격이 발생할 확률이 거의 없다고 판단하여 이 기법을 그대로 적용한 것이다.

그러나 근거리 통신에서도 두 명의 공격자가 협력하면 중간자 공격이 가능하다는 것이 밝혀졌다〔5〕. 이 공격은 릴레이 공격(relay attack)으로 불리고 있으며 근거리 통신을 이용한 인증 시스템을 속일 수 있는 기법이다. 암호화로 보호된 통신 채널을 직접 공격하지 않고도 효과적으로 공격을 성공할 수 있다. 인증을 요청하는 자(prover, 이하 인증자)와 인증을 허가하는 자(verifier, 이하 검증자) 사이에 두 개의 가짜 기기가 위치하여 두 개체가 서로 정상적으로 통신을 하고 있다고 속이는 것에 목적을 둔다.

릴레이 공격은 인증을 위한 메시지들이 암호화되어 있고 공격자가 암호키를 알지 못하는 상황에서도 단순히 메시지를 전달하는 것만으로 공격이 가능하다. 본 논문에서는 인증자와 검증자는 미리 암호키를 안전한 채널을 통해 공유한 상태를 가정한다. 〔그림 1〕의 (가)는 두 기기의 정상적인 인증 프로토콜을 단순하게 표현한 것이다. 두 기기가 근접 위치에 와서 통신을 시작하면 검증자가 인증자를 확인하기 위해 임의의 값(nonce)을 보낸다. 인증자는 미리 공유하고 있는 암호키(K)를 이용하여 받은 값을 암호화하여 전달한다. 검증자가 복호화 했을 때 자신이 보냈던 값과 일치하면 인증자가 암호키를 알고 있음을 확신할 수 있으므로 인증자를 인증해준다. 〔그림 1〕의 (나)는 릴레이 공격을 보여준다. 그림에서와 같이 두 명의 공격자가 필요하다. 정상적인 인증자와 검증자에 각각 가짜 검증자와 가짜 인증자가 접근한다. 정상적인 인증자와 검증자가 실제로 근접 통신을 수행할 정도로 물리적으로 가까이 위치하고 있지 않을 때, 가짜 인증자는 진짜 검증자가 전달하는 임의의 값을 그대로 전달하여 가짜 검증자가 인증자에게 인증을 요구하는 상황을 만들어낸다. 가짜 검증자가 전달해온 인증 요청 메시지를 인증자가 받은 후 정상적인 인증 메시지를 전달하면 이 메시지를 그대로 검증자까지 다시 전달한다. 이와 같이 릴레이 공격은 암호화된 메시지에 대해 암호학적인 공격을 수행하지 않고도 단순히 메시지를 전달함으로써 공격에 성공할 수 있다.

근거리 통신을 이용한 인증은 빠르고 편리하다는 장점을 가진다. 두 기기의 거리가 가까워지면 자동으로 상호간의 통신이 발생하고 사용자의 개입이 요구되지 않는다. 그러나 이러한 장점은 보안적인 측면에서는 단점이 된다. 사용자(사람)의 개입이 없어 인증 과정에 대한 확인이 불가하기 때문이다. 릴레이 공격을 막기 위해 기존의 연구들은 사람이 인증 과정에 개입하도록 요구하기도 한다. 그러나 이는 근거리 통신이 가지는 가장 큰 장점인 편의성을 낮추게 된다. 본 논문에서는 본래의 근거리 통신과 동일하게 사용자가 전혀 개입하지 않으면서도 릴레이 공격을 탐지할 수 있는 기법을 제안한다.

본 논문은 다음과 같이 구성한다. 2장에서는 릴레이 공격의 특성과 기존의 관련 연구에 대해 살펴보고, 3장에서는 청각 채널을 통해 릴레이 공격을 탐지할 수 있는 방안을 설명한다. 4장에서는 3장에서 제안한 아이디어를 어떻게 실제로 구현할 수 있는지를 논하고, 5장에서는 구현된 소프트웨어를 통해 실험 결과를 보인다. 결론은 6장에 기술된다.

II. 관련 연구 및 릴레이 공격의 특성

릴레이 공격을 방어하기 위해서는 검증자와 인증자가 가까운 거리에 위치함을 증명할 수 있어야 한다〔6〕. 릴레이 공격을 감지할 수 있는 연구로는 패킷이 전달되었다가 돌아오는 시간(round-trip time)을 측정하여 상호간의 거리를 짐작하는 방법인 거리 제한 방식(distance bounding)이 주로 이용되어 왔다〔7〕. 그러나 다음의 이유로 거리 제한 방식은 한계를 가진다. 첫째, 통신 매체가 무엇인가에 따라서 지연 시간은 크게 달라질 수 있기 때문에 지연 시간이 곧 거리를 말해주지 않는다. 환경에 따라 오차가 있을 수도 있다. 통신 시스템에 있어서 트래픽은 항상 일정하게 발생하지 않으며 상황에 따라 지연 시간은 달라질 수 있다. 메시지 분실 또한 발생할 수 있다. 근거리 통신에서는 두 기기만이 서로 통신에 참여하기 때문에 주변 상황에 비교적 영향을 적게 받지만 그렇다고 하여 전혀 무시할 수는 없다. 둘째, 공격자가 유선 채널을 이용하여 빠르게 메시지를 전달할 수 있다〔8〕.

거리 제한 방식 외에 위치를 이용하는 방법도 있다[9]. 검증자와 인증자가 각각 자신의 위치를 측정하고 서로 비교하는 것이다. 직관적인 방법으로 간단하면서도 강력하게 릴레이 공격에 대응할 수 있는 방법이다. 그러나 기기가 자신의 현재 위치를 정확히 알아내는 것은 아직까지는 현실성이 없다. 대표적인 위치 측정 기술인 GPS(global positioning system)는 실내에서 정확도가 상당히 낮으며 오류 발생 빈도도 높다[10]. 기기가 WiFi 모듈을 장착하고 있을 경우 인근 AP(access point)의 SSID(service set identifier) 목록을 비교하는 방법[11]을 활용해볼 수 있으나 검증자와 인증자가 모두 WiFi 기능을 갖추고 있어야 한다. 또한 주변에 AP가 전혀 존재하지 않는 경우나 널리 사용되고 있는 의미 없는 이름(예를 들어, linksys, netgear, iptime, anonymous 등)의 SSID가 많을 경우에는 확인이 쉽지 않다.

사용자의 적극적인 개입을 통해 두 기기의 인접성을 판단하는 기법도 있다[12]. 두 기기는 하나의 버튼을 가지고 있다고 가정한다. 인증을 할 때 사용자가 두 손가락으로 각 기기의 버튼을 동시에 눌렀다가 떼고 이 때 눌려진 시간을 측정한다. 두 기기의 버튼이 한 사람에 의해 눌려졌음을 판단함으로써 같은 위치에 있다고 판단하는 방법이다. 간단한 인터페이스를 통해 판단이 가능하지만 사용자가 적극적으로 개입해야하는 것이 큰 단점이며 두 기기의 버튼을 사용자가 동일한 시간 동안 오차 없이 정확히 눌러주어야만 확인이 가능하다.

두 기기가 근접하여 위치하고 있음을 시각이나 청각 채널을 통해 판단하는 방법도 있다. 인간이 가진 오감(시각, 청각, 촉각, 후각, 미각) 중 현재 기계에 간단히 적용해볼 수 있는 것은 시각과 청각이다. 시각은 카메라를 통해 촬영한 사진이나 동영상을 분석함으로써 적용할 수 있고, 청각은 마이크로 녹음한 데이터를 분석해서 적용할 수 있다.

시각 채널을 통해 기기간의 인증을 하려는 시도로는 대표적으로 '보는 것이 믿는 것' (SiB, seeing is believing) 연구가 있다[13]. 릴레이 공격에 대한 대응을 목적으로 진행된 연구는 아니나 두 기기 간에 공개키를 안전하게 전달하기 위해 기기간의 인접성을 확인하는 과정을 거친다. 키를 전달하고자 하는 측에서 바코드(barcode)를 제공하고 정보를 얻으려하는 측에서 카메라를 이용해서 이것을 촬영함으로써 정보를 전달함과 동시에 현재 나와 마주하고 있는 기기가 내가 제공한 정보를 취득했음을 확인할 수 있다. 통신채널에서 전달받은 내용과 카메라(시각 채널)를 통해 받은 두 값을 비교해봄으로써 전달받은 데이터가 현재 마주하고 있는 기기로부터 온 것임을 확신할 수 있게 된다. 카메라로 전달되는 데이터는 본래의 데이터의 확인을 위한 짧은 길이의 정보(일반적으로 본 메시지의 해시 결과)만 전달된다.

SiB와 동일한 목적을 가지고 청각 채널을 이용하여 두 기기의 인접성을 확인한 연구도 있으며 대표적으로 L&C(loud-and-clear)[14]가 있다. 시각 채널을 이용할 경우, 기기가 카메라를 가지고 있어야 하며 시각 장애가 있는 사용자는 활용이 불가능하고 적정한 수준의 조도가 요구되는 약점을 지적한다. 그리고 역설적이게도 높은 보안이 요구되는 시설에서는 카메라 사용이 금지되는 경우가 많다. L&C는 보낼 메시지에 대한 해시 결과로부터 문장을 생성하여 스피커로 내보내고 사람이 이를 인식하게 한다. 두 기기가 모두 소리를 낸 후 비교하거나 한 쪽에서는 소리를 다른 한 쪽에서는 화면에 문장을 보여주어 비교하게 한다. 문장에서 사용되는 단어는 최대한 다르게 발음되는 것을 선택하여 사용하기 때문에 사람이 쉽게 구분할 수 있다는 주장이다.

본 논문에서 제안하는 기법 또한 소리를 이용한 채널을 추가하여 릴레이 공격을 차단한다. 그러나 L&C와 다르게 본 논문의 기법은 사용자의 개입이 전혀 요구되지 않는다. L&C는 SiB가 시각 채널로 전달한 정보를 단지 소리로 변환하여 보내는 수준인 것에 비해 본 논문에서 제안하는 기법은 두 기기가 위치하는 공간에서 발생하는 소음을 두 기기가 각각 녹음하여 비교함으로써 동일한 위치에 있음을 파악한다. 제안하는 기법은 기존의 시각 및 청각 채널을 이용한 논문과 비교하여 다음과 같은 장점을 가진다.

첫 번째로 사용이 편리하다. SiB의 경우 카메라를 이용하여 바코드를 사람이 직접 촬영해야 한다. 현재 자동 초점 기술 등이 발전하여 비교적 빠르게 촬영을 할 수 있으나 사람이 직접 기기의 카메라를 구동시키고 렌즈를 바코드 근처로 가져가서 촬영하는 것은 여전히 번거롭다. L&C의 경우도 사람의 개입이 반드시 필요하며 사용자가 직접 문장을 듣고 비교하는 과정을 거쳐야한다. 반면 제안하는 기법은 본래의 근거리 통신을 이용할 때와 같이 근처에 기기를 가져가면 자동으로 녹음이 진행되고 위치를 파악하므로 기존의 근거리 통신이 가지는 사용성(usability)을 전혀 해하지 않는다. SiB에서와 같이 적정 수준의 조도가 요구되지도 않으며 L&C에서와 같이 비교적 조용한 주위 환

경이 요구되지도 않는다. 오히려 주위 소음은 본 제안 기법의 인증 성공률을 더욱 높여준다. 두 번째로 공격자 입장에서 릴레이가 더욱 어렵다는 점이다. SiB의 경우 성능 좋은 카메라가 충분히 먼 거리에서 이차원 바코드를 정확히 인식할 가능성이 있으며, L&C에서도 성능 좋은 마이크가 녹음할 가능성이 있다. 두 경우 모두 공격자가 인식한 정보는 원본과 일치할 가능성이 높기 때문에 정확히 공격자 간에 인증 정보 전달이 가능해진다. 반면 제안하는 기법에서는 두 기기가 동일한 공간에 근접하여 위치하더라도 두 기기가 각각 녹음한 결과가 차이를 가질 수밖에 없기 때문에 릴레이 공격을 위한 공격자 간의 정확한 정보 전달이 쉽지 않다. 다시 말해서, 제안하는 기법이 적용되어 있을 경우 공격자는 릴레이 공격을 성공하기 위해 어떤 공간에서 발생하는 소음을 다른 공간에서 그대로 재현해야만 한다.

III. 소리를 이용한 릴레이 공격의 방어

소리를 이용한 릴레이 공격의 방어는 다음 순서로 진행된다.

1. 인증 과정이 시작되면 두 기기가 동시에 녹음
2. 증명자는 녹음한 것을 분석한 후 결과를 검증자에게 전송
3. 검증자는 자신이 녹음한 후 분석한 결과와 증명자가 보내온 것을 비교
4. 일치 여부를 판단 후 유사한 소리라고 판단하면 인증, 아니면 거부

검증자와 증명자가 같은 위치에 있다면 주변에서 발생하는 소리가 같을 것이라는 것이 본 논문의 핵심 아이디어다. 근접 통신 방식으로 동작하는 자동차 열쇠가 있다고 가정하자. 자동차는 도로에 주차되어 있고 차주는 도로변 카페 건물 내에 있다. 두 명의 공격자가 한 명은 자동차 근처에 한 명은 차주 근처 접근하여 릴레이 공격을 시도한다. 이 때 카페 내에서 발생되는 배경 소음(사람들의 말소리, 음악 소리 등)과 도로변에서 발생하는 배경 소음(자동차 소음, 경적 소리 등)은 확연히 다를 것이다. 따라서 인증 과정에서 검증자와 인증자가 동시에 짧은 시간 녹음을 한 후 인증자가 인증 정보를 보내면서 녹음된 정보도 함께 암호화해서 보낸다. 검증자는 인증 정보를 검증함과 동시에 녹음된 정보도 비교함으로써 두 기기가 물리적으로 인접한 곳에 위치하고 있는지 여부를 판단할 수 있게 된다.

더욱 적극적으로 청각 채널을 활용할 수 있는 방법은 두 기기가 녹음을 하면서 동시에 완전히 다른 임의의 소리를 재생하는 것이다. 인증 시마다 소리가 발생한다면 사용자에게 불쾌감을 주거나 주변 환경에 피해를 줄 수 있으므로 유의해서 사용해야 한다. 그러나 두 기기가 원격에 위치하면서도 둘 다 조용한 공간에 있거나 유사한 배경 소음을 가진 장소에 있을 경우 비슷한 녹음 결과가 얻어질 공산이 있으므로 고의적으로 소리를 발생시켜서 이를 방해할 수 있다. 또한 소리가 재생되면 릴레이 공격에 대한 일차적인 방어가 가능하다. 릴레이 공격이 성립하기 위한 가장 큰 조건 중 하나는 공격자(가짜 검증자)가 몰래 인증자의 기기에 접근하여 인증 정보 요청에 대한 응답을 하도록 유도할 수 있어야 한다는 것이다. 릴레이 공격의 시나리오에서는 주인의 주머니에 있는 진짜 증명자와 가짜 검증기를 몰래 접촉시키는 시나리오가 많은데[7], 소리가 발생한다면 증명자 기기의 주인이 공격을 알아챌 가능성이 높다.

본 논문에서는 인증 시 증명자와 검증자 역할을 수행하는 기기가 다음과 같은 기능을 가지고 있다고 가정한다.

- 소리를 녹음할 수 있다
- 데이터를 전송할 수 있다
- 음향 데이터를 분석할 수 있는 계산 능력을 가지고 있다.
- (선택적) 녹음을 하면서 동시에 소리를 재생할 수 있다

본 프로토콜의 수행에 있어서 가장 중요한 부분은 녹음된 두 소리 데이터의 일치 여부를 판단할 수 있는 방법이다. 녹음 데이터의 분석에 짧은 시간과 적은 자원을 사용하면서도 비교 성능이 높은 효율적인 기법이 요구된다. 비교하는 방법은 두 가지가 있을 수 있다. 첫 번째는 Shazam이 제안했던 음악 검색에서 사용되었던 방식이다[15]. 녹음된 결과에 푸리에 변환(fourier transform)을 시간별로 적용하고 이를 통해 얻은 특이점(peak)들만 비교하는 방식으로, 완전히 일치하지 않고 유사한 두 음원을 비교할 때 유리한 점을 가지고 있다. 본 논문이 제안하는 방식도 두 기기가 독립적으로 녹음을 하고 두 기기가 가지는 각각의 위치와 배경 소음이 발생하는 위치가 가지는 관계

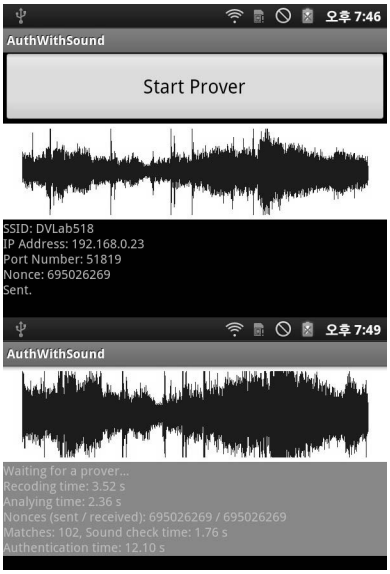
가 모두 다르기 때문에 두 기기의 녹음 결과가 정확히 일치하는 것은 불가능하다. 녹음된 결과를 그대로 비교하는 것은 주변 녹음 환경과 녹음된 데이터의 품질, 음량의 크기 등에 크게 영향을 받는다. 따라서 음원에서 특이점을 찾아내고 이를 비교하는 방식이 많이 이용되고 있다[16]. 두 번째 사용할 수 있는 방법은 두 기기가 음원 샘플을 동일하게 보유하고 이 샘플을 활용하여 임의의 멜로디를 연주한 후 서로 녹음하여 멜로디를 추출하고 비교하는 방법이다[17]. 이 방법은 양쪽 기기가 동일한 샘플을 가지고 있다는 가정이 있어 좀 더 정확한 인식에 도움이 될 수 있다. 그러나 본 논문에서는 앞서 언급한 대로 배경에서 발생하는 소리를 최대한 활용하기 때문에 첫 번째를 방식을 선택한다.

IV. 구현

인증자와 증명자 기기는 안드로이드 OS 2.3.3 (gingerbread)가 설치된 스마트폰(삼성 갤럭시탭 SHW-M180S, 1.0 GHz)을 사용한다. 스마트폰은 3장에서 정의한 인증 기기가 갖추어야 할 조건을 모두 갖추고 있다. 마이크와 스피커를 통해 녹음과 재생이 가능하고 계산 능력과 네트워킹을 통한 데이터 전달이 가능하다. 계산 능력은 녹음된 소리 데이터를 분석하기 위한 푸리에 변환을 위해 필요하다.

인증자와 검증자 사이에는 미리 공유된 암호화 키가 있다고 가정하며, 공격이 없는 상황에서, 인증자와 검증자는 다음과 같은 방식으로 통신하는 것으로 설계한다.

1. 인증자가 검증자에게 인증 시작을 요청
2. 검증자는 인증자에게 응답하면서 임의의 숫자(nonce)를 하나 보내주며 녹음을 시작 (설정에 따라 녹음과 동시에 소리를 발생시킬 수 있음)
3. 인증자도 검증자의 응답을 받고 녹음을 시작 (인증자 역시 설정에 따라 녹음과 동시에 소리를 발생시킬 수 있음)
4. 인증자는 녹음한 자료를 분석하고 결과를 전송 (이 때 결과는 암호화되며 검증자가 보내준 임의의 숫자도 함께 암호화 됨)
5. 검증자 또한 녹음한 음원 자료를 분석하고 인증자로부터 받은 메시지를 복호화
6. 인증자가 보내온 임의의 숫자가 검증자 자신이 보냈던 숫자와 일치하지 않으면 인증 실패
7. 동일한 숫자를 받았을 경우 검증자 자신이 분석한 결과와 인증자로부터 받은 분석 결과를 비교
8. 분석 결과가 두 녹음이 유사하다고 판단하면 인증 성공, 그렇지 않으면 인증 실패



[그림 2] 동일한 배경 음악을 동일한 위치에서 증명자와 인증자가 녹음한 결과. 유사한 패턴을 보이지만 같은 소리 데이터가 녹음되는 것은 아님을 알 수 있다.

임의의 숫자는 검증자가 임의로 생성하게 되며 지금 인증을 요청하는 인증자와 서로 동일한 암호키를 가지고 있는지 점검할 수 있게 해준다.

가장 중요한 과정은 8번으로 두 음원이 얼마나 유사한지 판단하는 부분이다. 단순히 녹음된 데이터를 비교하는 것 - 녹음된 결과의 수치를 순서대로 비교하는 것 - 으로는 유사성 판단이 불가능하다. 두 음원은 정확히 동일한 시점부터 녹음되지 않는다. 네트워크로 신호를 받고 녹음을 시작하므로 시간 차이가 발생하며 스마트폰은 본 인증 과정만을 수행하는 전용 기기가 아니므로 다른 작업과의 우선순위 차이에 의해 상황에 따라 다소 늦은 시각에 녹음과 재생이 시작될 수도 있다. 시간차의 문제를 차치하여도 녹음된 결과가 달라질 수 있다. 가까운 위치에서 녹음을 동시에 하므로 두 기기가 유사한 녹음을 할 것이지만 - 사람은 이를 쉽게 구분할 수 있을지 몰라도 - 기기 입장에서는 상당히 다른 신호로 인식할 수 있다. [그림 2]는 주변에서 음악이 재생되고 있고 인증자와 증명자 두 기기가 5cm 정도 떨어져 있을 때 녹음된 결과를 시각화 한 것이다. 같은 주변 배경음을 녹음한 결과이므로 유사

〔그림 3〕 배경 소음이 통제된 방에서 인증자(위)와 검증자(아래)가 녹음한 결과 (두 기기는 인접하여 녹음을 진행)

한 패턴을 관찰할 수 있으나 두 기기의 마이크 위치와 음원의 발생 위치에 따라 녹음된 결과가 단순 비교할 수 있을 정도로 동일하지는 않음을 볼 수 있다.

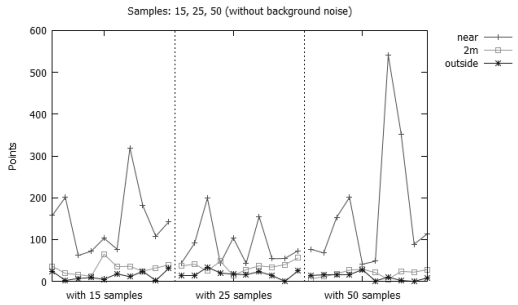
3장에서 설명한 바와 같이 두 녹음 결과의 유사성을 판단하기 위해 Shazam의 아이디어를 이용한다〔15〕. 본 구현에서는 샘플링 주기(sampling rate) 32,000Hz, 단일 채널 (mono), 16 비트로 녹음한다. 두 음원 자체의 유사성을 비교하는 것이므로 다중 채널(stereo)보다는 단일 채널을 이용해서 적정한 수준의 고음질로 녹음한다. 푸리에 변환은 녹음된 결과에서 차례대로 512 바이트 단위로 적용하였다. 푸리에 변환의 결과를 시간 순으로 이차원으로 나열한 것을 스펙트로그램(spectrogram)이라고 한다. 스펙트로그램을 만들고 주변의 값보다 특이하게 높은 값을 가지는 점을 추출한다. 본 논문에서는 주변 값들과의 차이가 가장 큰 특이점 순으로 150개의 점을 추출한다. 이렇게 얻어낸 위치를 별자리 지도(constellation map)라고 한다. 추출한 모든 점을 차례로 기준점(anchor point)으로 정하고 기준점을 중심으로 일정 시간 이내에 그리고 주파수 값이 크게 다르지 않은 지역(target zone)을 설정한다. 이 범위에 위치한 각각의 점(target point, 목표점)에 대해 해시(hash)값을 얻어낸다. 해시는 〔기준점의 주파수, 목표점의 주파수, 두 점의 시간 차〕이며 추가로 기준점의 시간도 기록한다. 기준점을 중심으로 목표점이 너무 많을 수 있으므로 하나의 기준점에서는 최대 15개의 목표점만 얻어내고 해시를 계산한다. 본 구현에서는 150개의 점을 얻어내고 각각 최대 15개의 해시를 얻게 되므로 얻을 수 있는 해시의 수는 최대 2,250개가 된다. 증명자는 얻어낸 모든 해시를 검증자에게 전달한다.

검증자도 마찬가지로 녹음을 하고 위와 같은 분석을 하여 해시를 얻은 후 증명자로부터 받은 해시와 비교한다. 일치하는 해시가 존재한다는 것은 유사한 특징점이 존재한다는 것이다. 해시에는 특정 시간 값이 담기는 것이 아니라 두 점의 시간차가 담겨 있기 때문에 두 음원의 시간 동기화는 필요하지 않다. 해시가 일치할 경우 해시에 추가적으로 기록해두었던 기준점의 시간 값을 비교한다. 해시가 일치하는 모든 경우에 대해서 시간차를 모두 기록한 뒤 가장 많이 발생한 시간차 값을 찾고 해당 시간차의 빈도를 얻어낸다. 두 녹음 결과가 비슷했다면 해시가 많이 일치할 것이고 일치한 해시들이 우연히 겹친 것이 아니라 시간 차이까지 비슷하게 가지게 될 것이다. 따라서 시간차의 빈도를 두 녹음 결과가 유사한 정도를 나타내는 값으로 이용할 수 있다. 이와 같이 비교를 통해 얻어낸 결과(일치한 해시들 중 가장 많이 발생한 시간 차이 수)를 '점수'라고 칭한다. 두 기기가 인증 과정을 수행하고 점수가 높으면 근접한 것으로 판단하여 인증에 성공하게 된다.

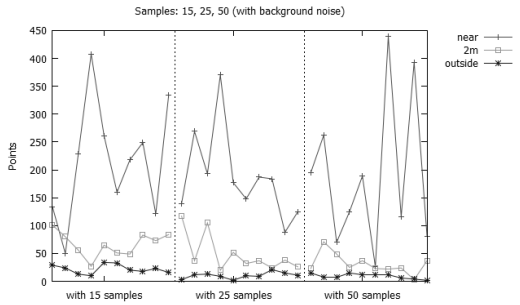
두 기기가 녹음을 어느 정도의 길이로 진행해야 하는지 또한 중요하다. 인증 시간이 짧을수록 사용성이 높아지기 때문이다. 근거리 통신을 통한 인증은 편의성을 위해 사용되는 것이므로 가급적 짧은 녹음으로 두 기기의 위치를 빠르게 판단해야 한다. 그러나 녹음하는 시간이 짧을수록 비교할 수 있는 데이터의 양이 적어지므로 정확도를 높이기 위해서는 가급적 길게 녹음하는 것 유리하다. 적절한 녹음 길이는 다음 장에서 실험을 통해 알아본다.

V. 실험 결과

실험으로 증명하고자 하는 것은 짧은 시간 동안의 녹음으로도 두 기기가 서로의 인접성을 파악할 수 있다는 것이다. 본 실험에서는 두 기기의 위치를 다음 세 가지로 구분한다. 첫 번째는 완전히 근접한 위치이다. 근거리 통신이 발생하는 10센티미터 이내로 한다. 두 번째는 2미터 정도 떨어진 거리이다. 단, 이 때 두 기기는 같은 방안에 위치한다. 두 번째 위치를 실험에



〔그림 4〕 샘플 수 15, 25, 50개에 따른 위치별 점수 변화 (주변 소음이 최대한 차단된 상황, 각 실험 10회)



〔그림 5〕 샘플 수 15, 25, 50개에 따른 위치별 점수 변화 (증명자 주변에서 음악 소리가 지속적으로 발생하는 상황, 각 실험 10회)

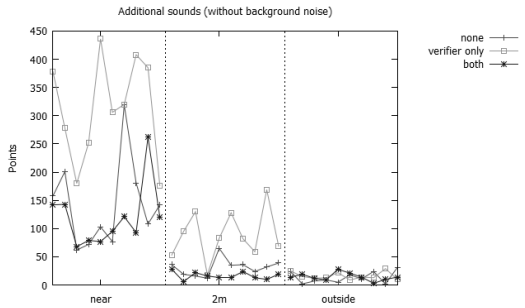
추가하는 이유는 두 기기가 같은 방에 있을 경우 충분히 유사한 배경 소음을 녹음할 수 있기 때문이다. 방 안에 어떤 음악이 흐르고 있다면 2 미터 떨어져 있다고 하더라도 같은 방에 위치하므로 유사한 녹음 결과를 얻을 가능성이 있다. 세 번째는 증명자가 방 밖에 위치하는 것이다. 이 환경에서는 상당히 다른 녹음 결과를 얻을 공산이 크다. 본 논문이 제안하는 방식은 두 기기가 완전히 근접하여 위치하고 있음을 감지하는 것이 목표이다. 즉, 기기의 위치를 위와 같이 달리했을 때 녹음 결과에 주목할 만한 차이가 있어야 한다.

본 실험에서는 실험이 진행되는 방에서 어떤 소음이 발생하는가가 중요한 환경 변수가 된다. 주변 소음이 가지는 임의성이 너무 높기 때문에 두 가지의 통제된 환경을 가정한다. 아무런 주변 소음이 없는 환경과 낮은 임의성을 가지는 배경 소음이 존재하는 환경이다. 본 논문이 제안하는 방식은 주변 소음을 판단의 근거로 삼기 때문에 주변에 아무런 소리가 발생하지 않는 상황은 제안하는 방식에서 최악의 시나리오가 된다. 이론적으로 두 공간이 완전히 다른 위치에 있더라도 두 공간 모두 소리가 존재하지 않으면 구분이 불가능할 것이다(물론, 공간이 진공상태가 아닌 이상 소리의 완전한 제거는 불가능하다). 본 실험에서는 주변 소음을 최대한 차단하기 위해 사람의 말소리나 음악 소리가 전혀 없는 방에서 실험을 진행했다. 이와 같은 상황을 가정한 이유는 최악의 상황에서도 두 기기의 위치를 파악할 수 있음을 보이는 것이 중요하기 때문이다. 〔그림 3〕에서 보듯이 녹음된 결과를 보면 소리의 파형(waveform)이 가지는 크기가 거의 없음을 알 수 있다. 그럼에도 불구하고 위치 구분은 높은 확률로 가능했다. 〔그림 4〕는 샘플 수 15, 25, 50개에 대해 두 기기의 위치를 달리하여 인증을 10회 시도한 결과이다. 두 기기는 어떤 소리도 발생시키지 않았으며 생활 소음이 극히 제한된 환경에서 녹음을 시도하였음에도 근처에 있을 때 더 높은 점수를 기록했다.

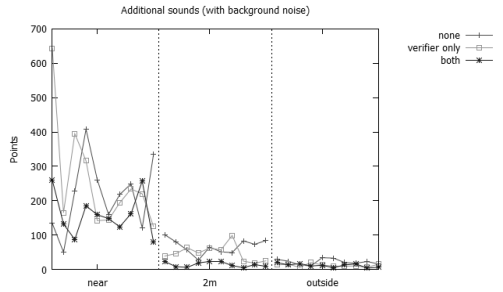
인증에 소요되는 시간을 줄이는 것도 중요하다. 근거리 통신은 두 기기 간에 빠르고 편리하게 데이터를 교환하기 위해 사용하기 때문에 녹음에 많은 시간을 투자하는 것은 바람직하지 않다. 〔그림 4〕에서 여러 샘플 수로 실험을 진행한 것은 짧은 녹음으로도 두 녹음 결과를 잘 비교할 수 있는가를 확인하기 위함이었다. 많은 샘플을 이용하여 비교할수록 결과가 좋아지는 것은 당연하겠으나 〔그림 4〕는 결과에 큰 차이가 없음을 보인다. 4장에서 설명한바와 같이 샘플 하나는 4096 바이트의 오디오 데이터이다. 샘플 수 15, 25, 50은 임의로 설정한 값이며 각 샘플 수에 따른 인증 시간 은 〔표 1〕에 정리되어 있다. 시간 값은 총 60회 실험 결과에 대한 중앙값(median)이다. 녹음과 음원 분석에 소요되는 시간은 당연히 추출한 샘플 수에 비례한다. 분석에 소요되는 시간은 거의 동일하다. 그 이유는 음원 분석 후 얻어내는 해시의 수가 샘플 수와 무관하기 때문이다. 4장에서 언급한 대로 샘플 수와 관계없이 150개의 특이점을 추출하고 각 특이점을 기준으로 최대 15개의 해시를 얻어낸다. 전체 인증에 소요되는 시간은 통신 지연 시간이 포함된 것이다. 본 실험에 사용된 스마트폰이 근거리 통신을 지원하지 않

〔표 1〕 샘플 수에 따른 인증 과정에서의 시간 소요 변화 (시간 단위: sec)

샘플 수	녹음	음원 분석	분석 결과 비교	전체 인증
15	1.3	0.84	1.84	4.795
25	1.95	1.3	1.825	5.955
50	3.57	2.42	1.785	8.82



〔그림 6〕 두 기기가 녹음만 하는 경우, 증명자만 소리를 발생시키는 경우, 두 기기가 모두 소리를 발생시키는 경우에 대해 두 기기의 위치에 따른 점수 (주변 소음 억제된 환경. 샘플 수 15개)



〔그림 7〕 두 기기가 녹음만 하는 경우, 증명자만 소리를 발생시키는 경우, 두 기기가 모두 소리를 발생시키는 경우에 대해 두 기기의 위치에 따른 점수 (증명자 근처에서 음악이 재생되어 배경 소음을 만들어내는 환경에서 실험 진행. 샘플 수 15개)

아 무선 랜을 이용하여 두 스마트폰이 데이터를 주고받았다. 샘플 수 50개의 경우 총 소요 시간이 대략 9초로 인증 시 다소 오래 걸린다고 볼 수 있으며, 25개 이하의 경우 수용할만한 시간으로 보인다.

생활에서 발생하는 여러 가지 소음을 실험에 추가하는 것이 정확한 실험이겠으나 소음 발생은 임의성이 높으므로 본 실험에서는 검증자 주변에서 음악을 재생하는 것으로 소음을 시뮬레이션 한다. 대신 약간의 임의성을 추가하기 위하여 음악은 임의로 재생하였다. 이와 같이 배경 소음을 가정한 상황에서 〔그림 4〕와 동일한 실험을 진행하였으며 〔그림 5〕가 그 결과이다. 증명자 주변에서 소리가 발생할 경우 검증자가 다른 위치에 있으면 두 기기가 다른 소리를 녹음하게 되므로 확연히 낮은 점수를 얻음을 볼 수 있고, 같은 방 안에서 2미터 떨어져 있는 경우에는 다소 혼란이 발생하는 것을 확인할 수 있다. 떨어져 있지만 동일한 음악 소리에 접근할 수 있기 때문이다. 그러나 두 기기가 근접해 있을 경우에는 더욱 높은 점수를 전반적으로 얻게 되기 때문에 구분이 가능함을 확인할 수 있다.

결론적으로 두 기기가 모두 조용한 공간에 위치하거나 혹은 주변 소음이 존재하는 경우 모두 기기 간의 인접 여부를 파악할 수 있음을 알 수 있다. 〔표 2〕는 각 10번의 실험값(〔그림 4〕와 〔그림 5〕)에 비교를 위해 중앙값을 정리한 것이다. 배경 소음이 있을 경우 근접 여부는 확연하게 구분이 가능하며, 배경에 소음이 전혀 없는 상황에서도 전반적으로 구분이 가능하다.

두 기기가 녹음에만 의존하지 않고 외부 환경변수에 대한 의존성을 약화시키고 탐지의 정확도를 높이기 위해 기기가 직접 소리를 재생했을 때 결과의 변화를 실험해본다. 기기가 소리를 발생시키는 경우는 다음 두 가지로 구분해볼 수 있다.

1. 검증자 쪽에서만 소리를 재생하고 인증자는 소리를 발생시키지 않는 경우
2. 검증자와 인증자가 모두 소리를 재생하는 경우

〔그림 6〕은 주변 소음을 통제한 상황에서 실험한 결과이다. 검증자만 소리를 발생시켰을 경우가 가장 점수가 높으나 주변이 조용한 경우 2미터 정도 떨어진 위치에서도 유사한 소리를 녹음할 가능성이 있다. 주변이 조용한 경우에는 두 기기가 모두 소리를 발생시키는 것이 가장 확실하게 구분이 가능함을 알 수 있다. 그러나 증명자만 소리를 발생시키는 상황에서도 비교적 잘 구분을 할 수 있다.

〔그림 7〕은 주변에서 음악이 재생될 경우에 기기가 소리를 발생시킴으로써 얻는 결과이다. 이 경우에도 두 기기가 모두 소리를 발생시킨다면 확실히 구분이 가능함을 알 수 있다. 단, 두 기기가 모두 소리를 발생

〔표 2〕 두 기기가 소리를 재생하지 않고 녹음만 하여 비교하는 경우 (10회 실험 결과의 중앙값, 단위: 점수)

주변 상황	두 기기의 위치	15개 샘플	25개 샘플	50개 샘플
배경 소음 없음	근접	125	64	101
	2m	33.5	37	21.5
	다른 방	10	16.5	12
배경 소음 있음 (주변에서 음악 재생)	근접	223	180.5	157
	2m	69	37	24.5
	다른 방	21.5	10.5	10

〔표 3〕 두 기기가 소리를 재생하는 경우의 점수 변화 (10회 실험 결과의 중앙값, 샘플 수 15개, 단위: 점수)

주변 상황	두 기기의 위치	두 기기 모두 소리 발생하지 않음	증명자만 소리 발생	두 기기 모두 소리를 발생
배경 소음 없음	근접	125	306	107.5
	2m	33.5	82.5	15
	다른 방	10	13.5	14
배경 소음 있음 (주변에서 음악 재생)	근접	223	207	153.5
	2m	69	47	12.5
	다른 방	21.5	11.5	12.5

시키는 것은 그만큼 사용자의 불편함을 유발할 수 있기 때문에 신중하게 결정해야 한다.

〔표 3〕은 각 10회의 실험(〔그림 6〕과 〔그림 7〕)결과를 한 눈에 비교하기 위해 중앙값을 정리한 것이다. 두 기기가 모두 소리를 발생시키는 경우 근접성을 확실히 구분할 수 있음을 알 수 있다. 증명자만 소리를 발생시키는 경우는 배경 소음이 거의 없는 경우에 잘못된 결과를 얻을 확률을 낮출 수 있다.

실험 결과, 주변 소음의 여부와 크게 상관없이 두 기기의 인접성 여부를 녹음을 통해 비교적 정확히 판단할 수 있음을 알 수 있었다. 게다가 두 기기는 아주 짧은 녹음(약 1.3초)과 분석 시간(약 0.8초)만으로 공격 탐지를 진행하고, 증명자도 공격 여부 판단에 약 1.85초의 짧은 시간만을 소요한다. 사용자 편의성이 다소 저하되더라도 두 기기가 직접 소리를 발생시키면서 위치를 판단할 경우 더욱 정확성을 높일 수 있음도 확인했다.

VI. 결 론

릴레이 공격은 근거리 통신으로 두 기기가 인증하려 할 때 인증 프로토콜의 암호화된 채널을 공격하지 않고 단순히 메시지를 전달하는 방식으로 원거리에 위치한 두 기기간의 정상적인 인증을 유도할 수 있다. 릴레이 공격을 방어하기 위해서는 인증을 시도하는 정상적인 두 기기가 인접한 위치에 있다는 것을 증명할 수 있어야 한다. 메시지 전달 지연 시간을 측정하거나 위치를 측정하는 방법이 제안되었으나 이들 방법만으로 정확히 두 기기의 물리적 거리를 측정하기가 어렵다. 본 논문은 릴레이 공격을 방어하기 위해 청각 채널을 추가함으로써 탐지 확률을 더욱 높이고자 했다. 두 기기가 같은 공간에서 아주 근접한 위치에 있다면 주변에서 발생하는 소리를 녹음했을 때 아주 유사한 녹음 결과를 얻을 것이기 때문에, 이를 비교하면 두 기기가 동일한 곳에 있다고 판단하는데 좋은 근거가 된다. 본 논문에서는 스마트폰을 이용하여 녹음된 결과를 분석하고 비교하는 방법을 제시하고 실험을 통해 녹음을 통한 두 기기의 인접성을 증명할 수 있음을 보였다. 주변 소음이 거의 없는 악조건 속에서도 인접성 판단을 비교적 정확히 할 수 있었으며 전체 인증 시간이 5초를 넘지 않는 설정에서도 정확히 두 기기의 거리를 판단함을 실험을 통해 검증하였다.

참고문헌

〔1〕 E. Haselsteiner and K. Breitfuß, "Security in near field communication (NFC)," Workshop on RFID Security RFIDSec. 2006.

〔2〕 W. Diffie and M.E. Hellman, "New directions in cryptography," IEEE Transactions on Information Theory, vol. 22, no. 6, pp. 644-654, Nov. 1976.

〔3〕 Standard ECMA-386, "NFC-SEC-01: NFC-SEC Cryptography Standard using ECDH and AES," 2nd edition June 2010.

〔4〕 W. Diffie, P.C. Van Oorschot, and M.J. Wiener, "Authentication and authenticated key exchanges," Des. Codes Cryptography, vol. 2, no. 2, pp. 107-125, June 1992.

〔5〕 G.P. Hancke, "Practical attacks on proximity identification systems," IEEE Symposium on Security and Privacy, pp. 298-333, May 2006.

〔6〕 F. Stajano, F. Wong, and B. Christianson, "Multichannel protocols to prevent relay attacks," In Proceedings of the 14th international conference on Financial Cryptography and Data Security (FC'10), Springer-Verlag, pp. 4-19. Jan. 2010.

〔7〕 S. Brands and D. Chaum, "Distance-Bounding Protocols (Extended Abstract)," Advances in Cryptography - EUROCRYPT '93, LNCS 765, pp. 344-359,

May 1993.

[8] P. Thevenon, O. Savry, and S. Tedjini, "On the weakness of contactless systems under relay attacks," Software, Telecommunications and Computer Networks (SoftCOM), 19th International Conference, pp. 1-5, Sept. 2011.

[9] L. Francis, G. Hancke, K. Mayes, and K. Markantonakis, "Practical NFC peer-to-peer relay attack using mobile phones," In Proceedings of the 6th international conference on Radio frequency identification: security and privacy issues, Springer-Verlag, pp. 35-49. June 2010.

[10] T.D. Le, T.M. Doan, H.N. Dinh, and N.T. Nguyen, "ISIL: Instant search-based indoor localization," IEEE Consumer Communications and Networking Conference(CCNC), pp. 143-148, Jan. 2013.

[11] K. Bauer, H. Gonzales, and D. McCoy, "Mitigating Evil Twin Attacks in 802.11," Performance, Computing and Communications Conference (IPCCC), pp. 513-516, Dec. 2008.

[12] S. Kang, J. Kim, and M. Hong, "Relay attack prevention with a single button for near field communication," Proceedings of the FTRA AIM 2013, pp. 86-87, Feb. 2013.

[13] J. McCune, A. Perrig, and M.K. Reiter, "Seeing-is-Believing: Using Camera Phones for Human-Verifiable Authentication." In Proceedings of the IEEE Symposium on Security and Privacy, pp. 110-124, May 2005.

[14] M.T. Goodrich, M. Sirivianos, J. Solis, G. Tsudik, and E. Uzun, "Loud and Clear: Human-Verifiable Authentication Based on Audio," 26th IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 1-10, July 2006.

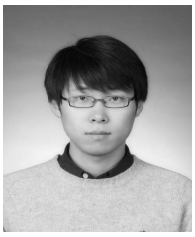
[15] A.L. Wang, "An industrial-strength audio search algorithm," Proceedings of the 4th International Conference on Music Information Retrieval(ISMIR), pp. 7-13, Oct, 2003.

[16] 이동현, 임민규, 김지환, "오디오 Fingerprint를 이용한 음악인식 연구 동향," 한국음성학회, 4(1), pp. 77-87. 2012년 3월

[17] P. Smaragdis and J.C. Brown, "Non-negative matrix factorization for polyphonic music transcription," Applications of Signal Processing to Audio and Acoustics, pp. 177-180, Oct. 2003.

〈저자소개〉

김 종 욱 (Jonguk Kim) 학생회원
2004년 2월: 아주대학교 정보및컴퓨터공학부 학사
2006년 8월: 아주대학교 정보통신전문대학원 석사
2006년 9월~현재: 아주대학교 정보통신전문대학원 박사과정
〈관심분야〉 키 관리, 익명 통신, 인증, 개인정보보호



강 석 인 (Sukin Kang) 학생회원
2008년 2월: 아주대학교 정보및컴퓨터공학부 학사
2008년 2월~현재: 아주대학교 대학원 컴퓨터공학전공 석박사통합과정
〈관심분야〉 네트워크 보안, 인증, 키 관리

홍 만 표 (Manpyo Hong) 종신회원
1981년: 서울대학교 계산통계학 학사
1983년: 서울대학교 계산통계학 석사
1991년: 서울대학교 병렬처리 전공 박사
1983년~1985년: 울산공과대학 전임강사
1985년~현재: 아주대학교 교수
〈관심분야〉 정보보호, 악성코드, DDoS, 스마트그리드 보안, 클라우드 보안