

R

D

S



Put Words In My Mouth

Telspace Research Report

by Amy Manià | April 2019

CONTENTS

PART 1 | THEORETICAL EXPLORATION

- Introduction
- The Internet of Things
- Listen Up
- Parrot Fashion
- Speak-Easy
- Join the Choir
- Vocal Variation
- Behind Voice Authentication Software
- What about the Law?
- Mitigation Tactics
- Threat Modelling

PART 2 | DOCUMENTED ATTACKS BY OTHERS

- Documented Vulnerabilities and Attacks

PART 3 | ATTACK AND PROOF OF CONCEPT

- Attack Scenario One
- Attack Scenario Two
- Attack Scenario Three

PART 4 | BIBLIOGRAPHY

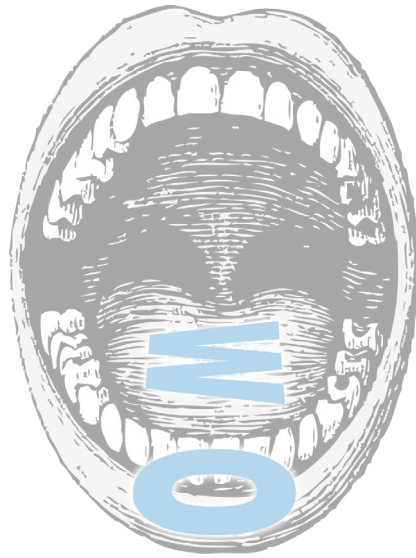
- Bibliography

NOTE | EXTERNAL REFERENCES

For the sake of thorough explanation, some video and audio media files have been included in this submission to supplement the written information discussed in the document.



Where the tag **EXTREF:** appear in the text, it refers to a file that can be watched and/or listened to. **PLEASE NOTE: Some references have not been included in the blog post, as they contain PII or sensitive information.**



R

D

S

PART 1

THEORETICAL EXPLORATION

Cue generic B-Grade film music...

Credits begin...

Camera pans into a stereotypical office building...

A medium-sized enterprise is rapidly expanding and has recently employed several new people. The company holds regular events at the office to entertain their clients. One of the new employees (John, a millennial) purchases a Home Speaker. It's a new brand called XYZ, which was the most affordable option at the store. John brings it to the office, so that music can be played at these events.

The new speaker is placed in the boardroom. It's a big hit at the first function and the boss, Jerry (who was initially sceptical) is now a big fan! He decides to let it stay and it becomes 'part of the furniture'. It takes its place, sitting innocuously in the corner of the boardroom. Jerry frequently uses the boardroom for privacy when conducting personal phone calls. He has started purchasing investment properties and speaks to his private banker every day.

A malicious hacker (Jack) is messing around one day on Shodan, and searches for XYZ speakers publicly visible online. He heard from a 'reliable source' on the dark-web that the XYZ executives wanted to make their product competitively priced, so they cut corners and did not implement any robust security on the device's firmware or software. Hacker Jack receives an extensive list of available devices – John's new Home Speaker, being used in the boardroom, is one of them.

Hacker Jack is easily able to compromise John's device. After some time, he notices the device is running with root privileges, and in several weeks he has completely traversed the company network, copied all of the data and left some backdoors - after all, one never knows when they may need to come back and install some ransomware.

Hacker Jack has overheard countless conversations between Jerry and his banker, and knows he only conducts his finances using this channel. Jack is particularly interested in the amount of money Jerry is able to spend on properties and wants to get his hands on some of it. Using software he's previously configured, any time the speaker in the office is activated by a voice, it records the sound and sends it back to Hacker Jack.

Hacker Jack passes Jerry's voice recordings through a speech-to-text API, and then to a voice-synthesising model which is able to make a clone of Jerry's voice. Hacker Jack writes his attack script, with some malicious banking instructions. Hacker Jack then calls Jerry's personal banker; Jack has scripted 'Jerry' to be in a rush, and Jerry's cloned voice urgently requests that a large sum of money be transferred to a bank account (Hacker Jack's account). The private banker complies as a lot of transactions have been processed lately, plus 'Jerry' kept saying how urgently it needed to be done.

But why stop there? Hacker Jack installs ransomware on the company's system and then sends an email to Jerry with his demands. Jerry complies and empties the company's account to pay Hacker Jack's ransom.

Jerry, stressed out from the recent fraud that took place on his personal account, and the ransomware payment that crippled his business – suffers a heart attack. Jerry is rushed to the hospital and doctors manage to save his life. They implant a new pacemaker into his chest to ensure a similar cardiac event does not happen again. Little does Jerry know that the pacemaker is also vulnerable to attack (save this unfolding drama for the sequel!)

From this scenario, the following observations can be made:

- Jack is a wildly talented lone-hacker, and quite vindictive (poor Jerry).
- This scenario is a little over-the-top, over-simplified in many aspects and slightly unrealistic. However, these types of attack have already either been researched, or proven to be possible.

So, although it sounds like a movie plot, worthy of Nicholas Cage starring as the lead actor, it may stand as a bleak prediction for the future of fraudulent attacks.

The Internet of Things:

Many households and offices around the world now incorporate IoT devices into their daily lives. Devices, such as Home Speakers, have risen in popularity and statistics show a vast number of internet users employ the use of voice-control and smart assistants. Ayinla (2019) states that 57.8 million adults own smart speakers. Statistics released at the RSA Conference stated that over 34% of internet users have expressed interest in a voice-controlled device. Matt Watchinski, vice president of the Global Threat Intelligence Group at Cisco Talos confirmed that the use of IoT devices will only increase. During a keynote, he predicts that “use of IoT devices will explode – perhaps 250 billion (including sensors and the like) that will be active by 2020” (Seals, RSA Conference 2019: How to Be Better, on Trust, AI and IoT, 2019).

However, with security vulnerabilities already explored and successfully exploited by researchers, it is worth asking: **are IoT devices helpful assistants, or a security hazard?**

For quite some time, we have idealised having virtual assistants that make our life easier. Think the likes of KITT from Knight Rider in the 1980s; who, to a large extent, ‘personified’ how artificial intelligence would be portrayed in mainstream media. The suave voice, witty remarks and corresponding flashing light is a concept that has been used in many films since - the most recent and memorable example being J.A.R.V.I.S from Iron Man (Coomes, 2018).

Its therefore natural, that consumers would expect this kind of slick experience in reality; manufacturers are doing their best to supply that demand. Smart Speaker sales have boomed, and more products are being released into the market that can integrate with the Smart Speaker. Items like keyless smart-locks, home-automation tools (that can control lights and appliances) smart coffee machines and vacuum cleaners (Luke, 2018).

“One of the most enticing images of the future has been that of technology that allows ordinary objects to seemingly come to life to cater to our personal needs. A certain version of that future has already materialized in present time. The internet-of-things (IoT) has made possible a new kind of space where that future exists — Complex IoT Environments (Trend Micro Research, 2019). A CIE is a network of IoT devices that allows the user to control many house-hold appliances, all from the convenience of a smartphone application. All of the individual components require a server to manage them. One example is the Perl based FHEM home automation server, which “is used to automate some common tasks in the household like switching on lamps/shutters/heating/ etc. and to log events like temperature/humidity/power consumption”. (FHEM, Unknown)

An excerpt from Trend Micro’s article (Stephen Hilt, 2019):

‘Exposed Automation Servers and Cybercrime published on 5 March 2019 explains that researchers “found more than 6,200 exposed Home



Assistant servers online, most of which were from the U.S. and Europe. Home Assistant has a history feature that shows the operational status of devices and, once accessed, could indicate when the inhabitants are away from home. In some exposed homes, their Home Assistant configuration file contained important credentials, like hard-coded router username and password. It is good to note however, that Home Assistant enforces password protection and most of the exposed home servers were password protected...

Information on the exposed FHEM servers included configuration files and device activities. Configuration files contain a wealth of information, like hard-coded credentials, lists of all devices in the home, and each device's location. Exposed FHEM servers could also show others details from the devices connected to it, like device status, sensor readings, and even electricity usage...

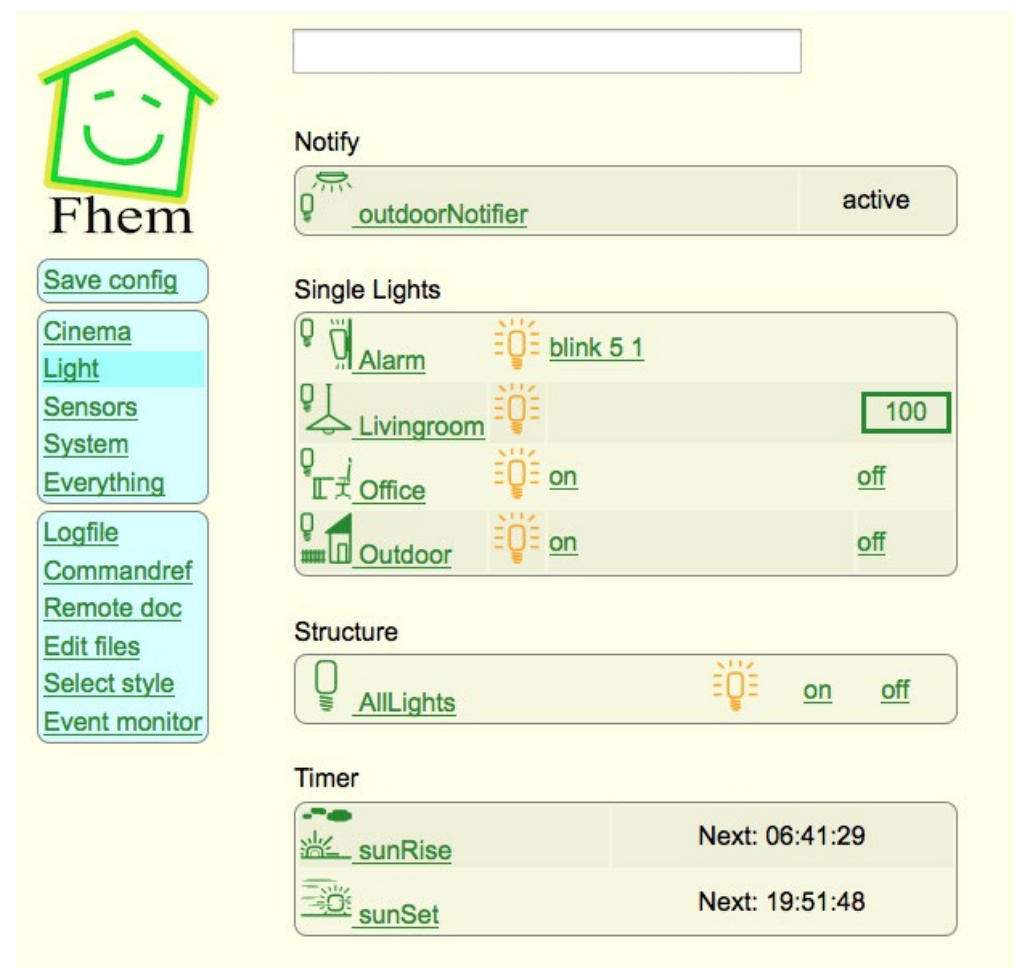
Exposure of automation servers opens smart homes and even smart buildings to several attack scenarios. For open-source automation servers, attackers can reprogram rules which, in turn, lead to a slew of different other attacks — from secretly adding devices to the system to turning off all security setups. Even exposed commercial servers can give attackers physical control over a household by allowing them to interact with controls like alarm systems. Exposed automation servers in buildings and industrial settings could impede business operations should their setups be tampered with. In addition, attackers can monitor and note patterns in resident behaviours using the information readily available in the exposed server.

Consumers are unwittingly opening themselves up to many attack avenues, by actively installing the IoT devices and/or CIE systems into their homes. Capabilities like the ones mentioned above can allow an attacker to:

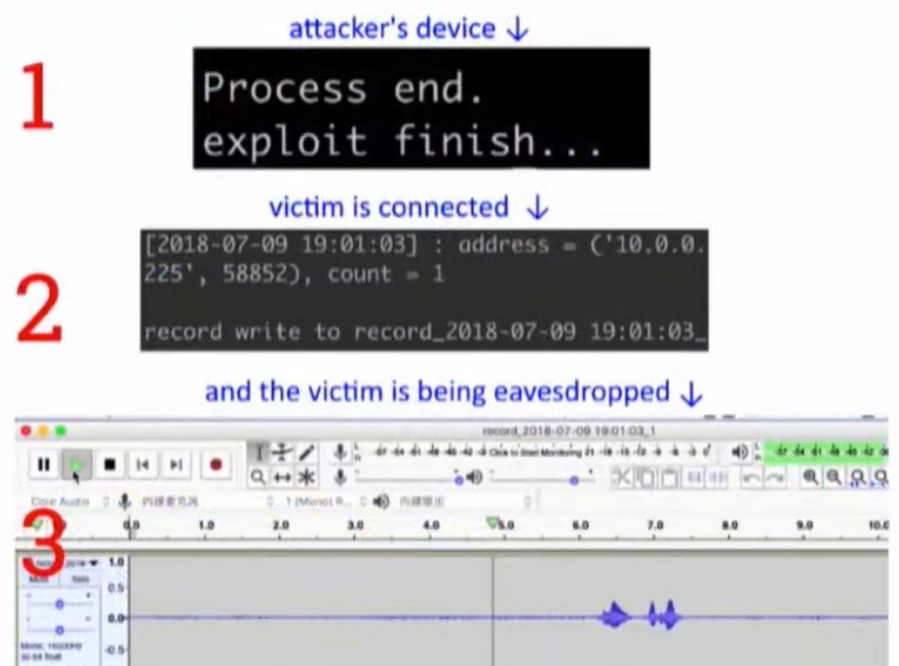
- Know when a resident is home/away
- Override security settings
- Access PII and account credentials.
- Listen to, watch and/or record any activities taking place in the house (where microphones and cameras are present).

Listen Up:

The talk titled “Breaking Smart Speakers: We Are Listening To You” presented by the Tencent Blade Team at DEF CON 26, showed that Home Speaker devices can be compromised, with voice recordings being sent directly to the attacker (Tencent Blade Team, 2018).



3 Steps to Eavesdropping the Target



Although the specific vulnerability explained in this lecture has been patched, other means exist that a threat actor could use to capture the audio data recorded by IoT devices. “Much of the embedded firmware running on these devices is insecure and highly vulnerable, leaving an indeterminate number of critical systems and data around the world at risk” (IoT For All, 2017). Even though patches can be issued, consumers may not update their software/firmware so that they are protected. Testimony of these kind of IoT security breaches are available in countless articles published online about a range of devices. They detail how these devices have vulnerabilities which leak personally identifiable information (PII) or allow remote access – which means this potential data leak goes above and beyond just Home/Smart Speakers.

So - what is possible when you have virtually unlimited access to a person’s voice, and the things they say?

Parrot-Fashion:

In November 2016 at the Adobe Max event, a new prototype software was presented called VoCo. What was dubbed ‘photoshop-for-voice’ demonstrated a program that enables editing and creation of audio. “VoCo takes in large amounts of voice data and breaks it into the distinct sounds that make up spoken language, collectively known as phonemes, before creating a voice model of whoever is recorded. If a word isn’t already in the recording, then the program will use these phonemes to create it from scratch. As with many content-manipulation technologies, over the initial applause and adulation loom questions about the way such tech could easily be used to cause harm and spread fake news” (Powers, 2018). Adobe have released no further information about the software, but this has not stopped competitors from creating software with similar capabilities.

A show on Netflix called *Follow This* tracks the journalist Charlie Warzel, who used an AI application called Lyrebird to produce a clone of his voice (Lyrebird, 2018). He called his mother and used the vocal clone during the conversation – his mom was convinced that she was speaking directly to her son (Warzel, 2018). This exact experiment was successfully replicated by another journalist who interviewed the developers at Lyrebird (Bloomberg, 2018). The name of the application was inspired by the Australian lyrebird, which is known for its ability to mimic any natural or artificial sounds they are exposed to (Puiu, 2019).

Warzel then interviewed Aviv Ovadya regarding what has been dubbed *The Infocalypse* – Ovadya discusses in detail how the early versions of the software used to manipulate media (video and/or audio) now referred to as Deepfakes (Peel, 2019), are already quite advanced, with some options being both free and easily available to the public. These apps allow the user to combine and then superimpose existing images and videos onto source images or videos using machine learning. The combination of the existing and source videos results in a video that can depict a person saying things or performing actions that never occurred in reality (Wikipedia, 2019).



The most popular application ‘FakeApp’ caused controversy on the internet, and the official website has consequently been taken down. However, this move has not prevented the software from being available online. The video below is an example of a Deepfake, where president Donald Trump’s likeness was superimposed over Alec Baldwin’s face (Clark, 2018). This video was shown on the television show *Saturday Night Live*, but has subsequently been blocked in both Canada and the USA (Derpfake, 2018).

Although deepfake applications have generally been used for nefarious purposes so far, there is extensive research actively taking place in the realm of face-swapping applications, TTS (Text-To-Speech) and SST (Speech-To-Text) or Speech Recognition. Google alone have released 10 publications since March 2017 on this topic as part of their ongoing Tacotron project (Google, Date unclear).

The latest paper released in late 2018 by Google was accompanied by software developments that have made huge advances. “Tacotron 2 is a multiple neural network architecture for speech synthesis. It is the combination of the text-to-speech systems (TTSS) WaveNet and Tacotron. It is an end-to-end TTS system with a sequence-to-sequence recurrent network that predicts mel spectrograms with a modified WaveNet vocoder. It can be directly trained from data and can achieve state-of-the-art natural human speech sound quality” (Golden, Unknown).

In simpler terms, it is an AI-powered speech synthesis system that can convert text to speech. Tacotron 2 works on the principle of superposition of two deep neural networks — One that converts text into a spectrogram, which is a visual representation of a spectrum of sound frequencies, and the other that converts the elements of the spectrogram to corresponding sounds (Biswas, 2018). The voice synthesising technology is used for the Google’s voice service. “The robotic voice is a staple in our culture, like Microsoft’s Cortana or Apple’s Siri. As the years have gone by Google’s AI voice has started to sound less robotic and more like a human. And now, it is almost indistinguishable from human (RankRed, 2018). Listen, and decide for yourself using the link below:



<https://google.github.io/tacotron/publications/tacotron2/index.html>

The DEF CON 26 lecture titled: *Your Voice is My Passport* describes the software options to record and manipulate voice data in more detail, Tacotron being one of them (Seymour, 2018). In their talk, John Seymour and Azeem Aqil focused on the voice-authentication feature and explained how it is being used more commonly during verification processes on various personal accounts. A clone of John’s voice was used to test login authentication on Microsoft Azure service – which worked. The two went on to mention that Schwab Bank (in the USA) is an example of an institution using voice authentication. Closer to home, Issue 14 of Banker SA (released in July 2015) mentioned that Investec had invested in this type of technology and that other institutions were implementing a similar system - along with other biometric technologies (Banker SA, 2015). Using a body part to authenticate account login, is both speedy and convenient. “Despite security concerns, South African consumers find biometric technologies... to be much faster and easier than passwords... among the 500 South Africans surveyed, 72% revealed they are interested in using biometrics to verify identity or make payments” (itweb, 2018).

At one point in the DEF CON lecture mentioned above, the speakers reference the movie *Sneakers* (IMDb, Unknown). For those who have not seen the film: the ‘heroes’ require access to a specific office. Access control is strict, with a voice recognition phrase from the office-user being required to enter the room in question. The film depicts the social engineering gymnastics required to collect adequate voice recordings from the victim, in order to carry out the heist. Considering the advances in technology and prevalence of IoT devices, the difficulty of capturing speech from a target would be minimised if the attacker had unlimited access to a range of utterances captured by any vulnerable IoT device with a microphone (such as Home Speakers, baby monitors or security cameras are obvious examples).

IoT devices are seemingly innocuous, and it is this exact quality makes them potentially dangerous. It’s fair to say that most IoT owners do not understand the potential security risks associated with having these devices in their private spaces. It’s also fair to assume that most IoT owners may not be tech savvy enough to update the software or firmware after they have been released or have two-factor authentication available – due purely to ignorance. For example, several Nest devices



(baby monitors) were hacked in January 2019. In an interview the one device owner mentioned that “she didn’t know the camera had speakers and a microphone” (Brice-Saddler, 2019).

These devices are strategically placed in our homes and offices, enabling them to have immediate access to a large range of speech from a user. As biometric/voice logins become more widely used, so does the risk of having them stolen. “Authentication is a must on the technical side, but there’s a human element in play that also needs consideration. Standard ‘don’t tell anyone your PIN/password, don’t use easy-to-guess numbers or phrases’ advice applies, but speaking your code[s] out loud invites more opportunities for exposing your credentials” (Wollerton, 2017). This means that IoT users may unwillingly and unknowingly be sacrificing both their privacy and security, in the name of convenience.

Speak-Easy:

Simple day-to-day errands like searching for a telephone number, navigating to your friend’s GPS coordinates or placing a reminder to call your dentist has changed dramatically over the last few decades - with the majority of these tasks now taking place via a device that is connect to the internet – and increasingly by directing these requests verbally to smart assistants.

As more people are exposed to technological resources, the demand grows and more services become available, with the resulting user-interface engineered with convenience in mind. *“With millennials quickly becoming the largest generation in today’s workforce, these trends may impact how employers and technology companies provide access to devices and applications in the near future... respondents recognised the benefits of biometric technologies like fingerprint readers, facial scans and voice recognition”* (itweb, 2018).

These newer forms of authentication - such as fingerprint, facial, iris and voice recognition - can make unlocking accounts and processing payments more convenient than traditional passwords or PINs. The basis of biometric technology is the ability for a customer to use something they already have (their voice or fingerprint etc.) rather than something they are required to remember (i.e. a password).

A MasterCard survey showed that “65% of South Africans would favour biometrics over passwords” and that “seven out of ten say biometrics, such as fingerprint or voice recognition technology, would be an easier way to access their accounts. Biometrics have the potential to provide a secure and seamless experience. Voice and facial features are difficult to fake or recreate. It’s easier for customers to use than information they have to remember – you have a voice and face already. Banks like Investec Private Bank have embraced biometrics.” (Lourie, 2017).

It is not only private companies that make use of biometrics, government institutions are also jumping onto the voice authentication trend: *“The Australian Tax Office (ATO) has collected voiceprints from 3.4 million Australians (or one in seven citizens) to authenticate their identity in interactions with its call centre or mobile app... Voiceprints from the database could be shared with other government agencies, with user consent, the ATO has said, and the Department of Human Services currently uses it to identify citizens calling its Centrelink”* (Burt, 2018). This is merely one example of a large scale entity making use of voice-based authentication. Instances like this one will end up as enormous repositories of Personally Identifiable Information, that could be vulnerable to theft or attack if poor security policies are in place.

Having vast amounts of personal data all stored in a single location is a threat actor’s dream-come-true. Criminals *“could break into your Amazon account and see all your Alexa interactions. [The hacker] would suddenly know a lot more about you”* (Wagenseil, 2018). These kind of “threats could come from financially motivated cybercrime gangs, state-sponsored spies and even the manufacturers themselves, who collect an increasing amount of personal data via our smart devices” (Edwards, 2017).

Join The Choir:

User-friendliness has a large impact on biometrics popularity. Research in this field indicates that uptake is only expected to increase: *“The use of biometrics technology for online authentication is picking up momentum in the cybersecurity industry. In fact, up to 90% of organizations will be implementing biometrics by 2020 according to a Spiceworks report”* (Capps, 2018).

Evidence of this can be easily observed, with various forms of biometric authentication having already been implemented by countless companies around the world; with South Africa being no exception to this trend.

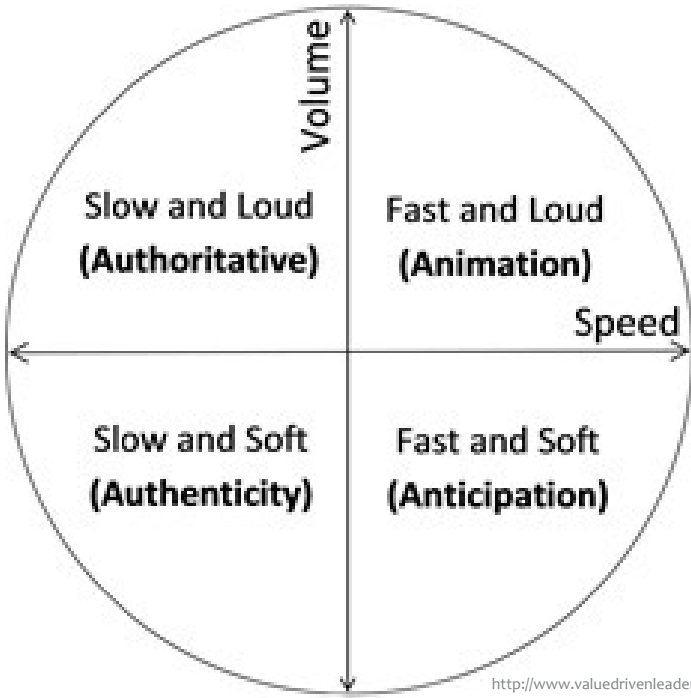
Investec and Vodacom are a few, very prominent companies who have introduced these systems, which notably include voice authentication: *“Vodacom, a mobile operator in South Africa, recently introduced Voice Password for customers using the My Vodacom app, and for those calling in to the Vodacom call centre. Instead of requiring customers to tap in a lengthy PIN or password string, or endure a string of security questions, Vodacom now provides voice biometrics so that customers can speak a simple passphrase to verify their identity”* (Nuance Communications, 2014).

This trend has caught on with South African banking institutions too - *“Ammar Faheem, solutions expert for digital banking and payments at Gemalto Africa, says: Local banks are adopting the technology just as well as any other banks worldwide. Within the next five years, more than 91% of SA’s banks plan to have fingerprint scanning, 84% face recognition, and voice recognition is to be used by 74%”* (Gool, 2018).

Vocal Variation:

Speech recognition capabilities have come a long way: “In 1952, ‘Audrey’ was invented by Bell Laboratories which could only understand numbers. But in 1962, the ‘shoebox’ technology was able to understand 16 words in English. Later, voice recognition was enhanced to comprehend 9 consonants and 4 vowels...By 2001, speech recognition technology development had hit a plateau, until Google came along. Google invented an application called ‘Google Voice Search’ for iPhones which utilized data centers to compute the enormous amount of data analysis needed for matching user queries with actual examples of human speech. In 2010, Google introduced personalized recognition on Android devices which would record different users’ voice queries to develop an enhanced speech model. It consists of 230 billion English words” (Kikel, 2018).

These advances in technology have trickled down into services accessible by the average person and companies that require these kinds of products. Websites for TTS and STT products explain that the software has many capabilities - for instance - that it can cancel out background noise to capture only what is being said, that it can distinguish between different languages and transcribe accordingly, and recognise different voices during a conversation/phone-call and produce a dialogue-style transcription.



<http://www.valuedrivenleaders.com/uncategorized/vocal-variety-in-preaching-an-important-part-of-influence/>

Google Cloud

Why Google

Solutions

Products

Pricing

Getting started

Q

Docs

Support

Sign in

AI & Machine Learning Products

Contact sales

Try free

Cloud Text-to-Speech features

Multilingual

Supports 100+ voices across 20+ languages and variants, with more to come soon.

WaveNet Voices

Exclusive multilingual access to DeepMind WaveNet voices that provide the most natural-sounding speech.

Text and SSML Support

Customize your speech with SSML tags that allow you to add pauses, numbers, date and time formatting, and other pronunciation instructions.

Speaking Rate Tuning

Customize your speaking rate to be 4x faster or slower than the normal rate.

Pitch Tuning

Customize the pitch of your selected voice, up to 20 semitones more or less than the default output.

Volume Gain Control

Increase the volume of the output by up to 16db or decrease the volume up to -96db.

Audio Format Flexibility

Choose from a number of audio formats including mp3, Linear16, and Ogg Opus.

Audio Profiles

Optimize for the type of speaker from which your speech is intended to play, such as headphones or phone lines.

rev.ai

Use Cases

Pricing

Documentation

API Features

Automatic Speech Recognition

Automatic Speech Recognition (ASR) converts spoken word to text with best-in-class accuracy.

Punctuation & Capitalization

Automatically punctuate (commas, question marks, periods, etc.) and capitalize for an easy-to-read transcript.

Speaker Diarization

Recognize multiple speakers and attribute text to each.

Timestamp Generation

Receive a timestamp for each word.

Custom Dictionaries

BETA

Customize vocabulary with names, industry-specific terminology, and more to increase transcript accuracy.

Live Streaming

COMING SOON

Transcribe speech to text in real-time.

TRY IT FREE

TALK TO AN EXPERT

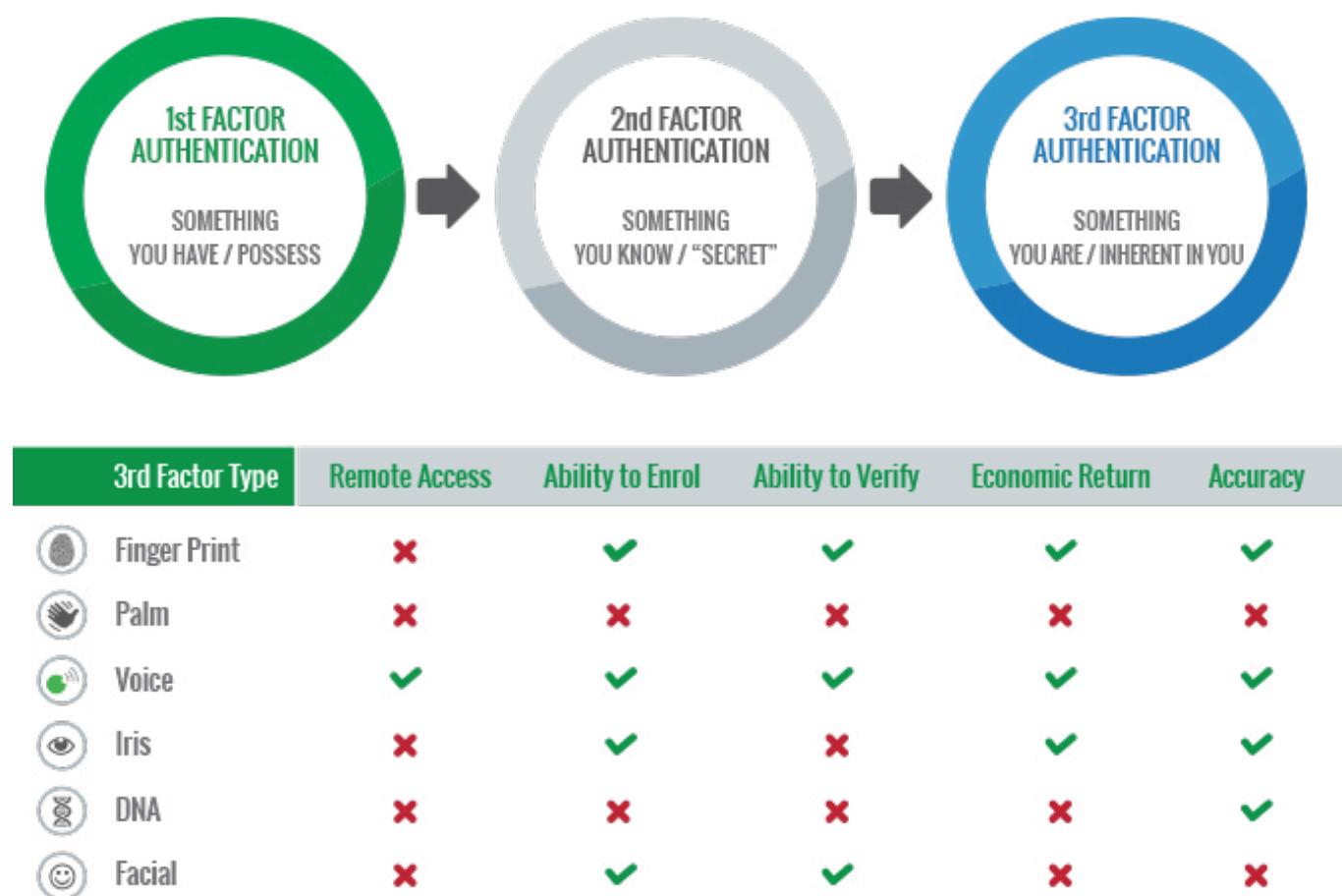


Image from the Google Cloud website and rev.ai website shown on the previous page list the software’s capabilities.

Voice recognition and authentication systems can allegedly identify the voice of a user even when the person in question is ill. “[Investec] are asked ‘what happens if the client has a cold?’ but voice biometrics continue to function, irrespective of language, cold, accent etc. It uses the characteristics of the voice, which is unique for each individual, rather than actual words. “We have implemented ‘free speech’ technology, which means not what is said, but rather the person talking is used in the authentication process. The technology is very mature and considers such instances. In addition, the authentication takes place in real-time conversation with a bank-er, so the responses would need to make sense relative to the discussion taking place” (Banker SA, 2015).

The level at which these systems can fully comprehend speech is debatable, especially in tricky sound environments. Although Voice Recognition software has im-proved vastly, devices may still struggle to correctly identify individuals, proof of this can be observed when using Smart Speakers: “Even if you use Voice Match [on a Google Home Speaker], your friends or kids can still engage with your device; they just won’t be able to get certain sensitive information. Google Home will support only six unique Voice Match users. It also isn’t perfect, as similar-sounding voices may trigger your device” (Long, 2018).

Behind Voice Authentication Software:

Most providers who offer voice authentication as a service, describe how reliable and safe this specific biometric method is. The OneVault website lists the applica-tions where voice authentication can be used (with the bold items being of specific relevance to this research report’s proof of concept):

- Internal & external password reset
- Employee time & attendance recording
- **Customer call identification & authentication**
- Internet & telephone system logon
- Cellular PUK resetting

- Criminal identification & solving crimes
- **Legally binding voice signature**
- Monitoring parolees or people under house arrest
- Immigrant tracking
- **Making/authorising payments** (OneVault, 2013-2019)

The image from the OneVault website (shown on the previous page) indicates that voice authentication is the most reliable form of biometrics to use, ticking both the practical and security requirements.

On the opposite side of the spectrum, there has been outcry since consumers became aware that their requests made to a voice assistant are kept indefinitely. In the case of the Google's Assistant and Amazon's Alexa, recordings need to be manually deleted – which is discouraged by the manufacturer, as this information aids in the training of speech recognition and natural language understanding systems, so removing the data could also affect the product's accuracy. Most users are probably under the false assumption that this happens via machine learning, but there's still a manual (aka human) factor while processing data.

Consumers are expected to understand that “...these devices are designed to listen. This includes recording and learning the tone of your voice and improving voice recognition and features for the virtual assistant” (Etienne, 2018). However, the manufacturer's do not explicitly mention that humans are used to accurately transcribe these voice recordings. Amazon have a generic disclaimer in their FAQ that says “We use your requests to Alexa to train our speech recognition and natural language understanding systems” which is a rather vague description of what happens behind the scenes.

In an article published on 11 April 2019, Amazon was ousted to employ thousands of people to transcribe voice recordings directed to the smart assistant (Durden, 2019). The article explains that an Amazon spokesperson defended their practise, saying “we only annotate an extremely small sample of Alexa voice recordings in order improve the customer experience” and that “Employees do not have direct access to information that can identify the person or account as part of this workflow” - however, there will always be some kind of data to link the recordings with the device they came from , to the accounts it is linked to with - which could compromise privacy.

Google's home assistant records all request, as well as the coordinate location where it came from. So not only does Google know what you are saying, they know where you are when you said it. To use the Google Assistant or Google Home, one has to use their Google credentials, a significant amount of your personal information falls under the umbrella of just one company. This may be preferable to scattering your data among dozens or hundreds of third-party apps, but centralization does present some challenges. “If you trust Google to take good care of your data in general, having it in one place versus all over the place is good,” said Jeff Wilbur, director of the non-profit Online Trust Alliance. “The danger, when it's all centralized, is if someone, somehow, gets access to your Google account, they have a rich set of stuff to look at — your voice queries, payment history and search history” (Long, 2018).

As George Avetisov, CEO of HYPR said “The ‘warehousing’ of personally identifiable information needs to end, since it can (and has) resulted in, ‘a catastrophic data breach’” (Armerding, 2017).

What About The Law?

It's an obvious assumption, that, like any personal data or personally identifiable information (PII), your voice is something that belongs to you, and biometrics are specifically defined in the South African *Protection of Personal Information Act* (The Banking Association South Africa, Unknown):

‘Personal information’ means information relating to an identifiable, living, natural person, and where it is applicable, an identifiable, existing juristic person, and may include the following:

- *information relating to the race, gender, sex, pregnancy, marital status, national, ethnic or social origin, colour, sexual orientation, age, physical or mental health, well-being, disability, religion, conscience, belief, culture, language and birth of the person*
- *information relating to the education or the medical, financial, criminal or employment history of the person*
- *any identifying number, symbol, e-mail address, physical address, telephone number, location information, online identifier or other particular assignment to the person*
- ***the biometric information of the person (Biometric information includes a technique of personal identification that is based on physical, physiological or behavioural characterisation including blood typing, fingerprinting, DNA analysis, retinal scanning and voice recognition.***

A voice is used to create verbal contracts. Moreover, in international and specifically South African law, “verbal agreements are no less binding than written agree-

ments. Proving these agreements and the exact terms thereof on the other hand is another matter (Malan Lourens Viljoen Incorporated, 2017) - unless the conversation has been recorded.

“Whatever the superficial attractions of IoT devices, this means there is nothing to reassure you that you won’t get more than you bargained for. That’s not to say every device out there is a risk, but consumers need to know what they are welcoming into their homes, and understand that any insecure embedded device they connect to the internet is a potential target for attacks...At the moment manufacturers don’t need to provide any guarantees of the safety of their equipment beyond electrical compliance” (Wickes, Unknown).

One would think that there would be legislation in place to protect the consumer, yet it is frightening to discover that there are currently no industry standards for IoT device security. It is clear that technology moves a lot faster than the legislation process – with potentially disastrous results and no recourse for victims.

Although a document that defines and regulates security standards in IoT devices has been proposed and drafted in Europe, it is still considered a work-in-progress and is a far way off from actual implementation – leaving a large window of opportunity open for threat actors to take advantage of vulnerabilities.

In the meantime, OWASP have put together a list of items to assist IoT manufacturers to help build better secured products (OWASP, 2017). However, this document is also labelled as a draft, and serves merely as a guide to cover the fundamentals of security. If manufacturers are not forced to comply with certain standards, guidelines like this have very little impact in reality.

There have been many recent examples of vulnerable IoT devices lately, one recent case revealed a “...series of both unauthenticated and authenticated remote code-execution vulnerabilities have been uncovered in a variety of Grandstream products for small to medium-sized businesses, including audio and video conferencing units, IP video phones, routers and IP PBXs. According to Trustwave Spider-Labs research...compromising these devices can allow an attacker to start scanning, installing remote access trojans and attacking other machines on the network that would otherwise be inaccessible; or install arbitrary applications. Attackers can also use the vulnerabilities to gain access to cameras and microphones to turn them into listening devices. ‘The most notable aspect of the vulnerabilities is what you can do simply by using the programs that get shipped on the device... This is pretty bad for places such as boardrooms or executive offices where confidential conversations frequently happen. An attacker can silently eavesdrop on a confidential conversation without anyone knowing the device has actually been compromised. As all the devices are running with root privileges, you have access to do pretty much whatever you want’” (Seals, Bugs in Grandstream Gear Lay Open SMBs to Range of Attacks, 2019).

When considering the above, the ‘movie plot’ introduction to this research paper starts to seem a lot less like a fiction, and a lot more like a potential attack vector -



VOICESIGN VOICE BIOMETRIC E-SIGNATURES
Speak on the Dotted Line
Legally binding equivalent of a hand-written signature.
Proven voice biometric technology – hundreds of thousands of voice e-signatures to date.
Significant reduction in use of paper, lower costs and environmentally friendly.
Integrates into any call flow to increase closure rates and shorten sales cycles.

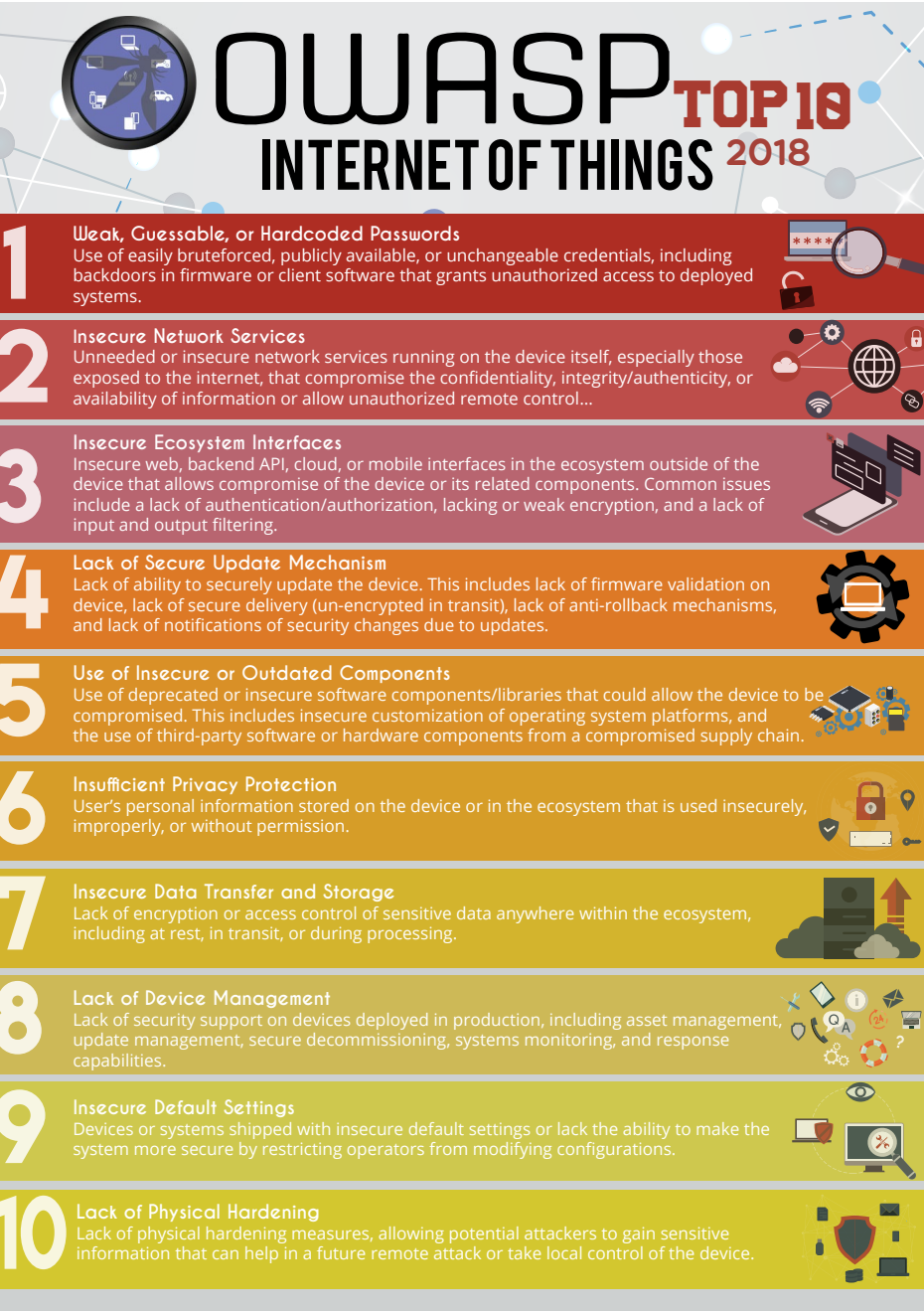


Image from VoiceVault, an American voice authorisation-based legal service (voicevault, Unknown).

OWASP IoT 10, retrieved from: <https://www.owasp.org/images/1/1c/OWASP-IoT-Top-10-2018-final.pdf>

especially when considering that a ‘voice’ can be used to create a contract (voicevault, Unknown).

Your voice can essentially be stolen from you via a device that has no data-security standards, and be used to complete a legally binding transaction – where it would be very difficult to prove that it in fact, it was not ‘you’ on a recorded phone call.


Mitigation Tactics:

It seems that phreaking is still alive and well, and can often go relatively undetected, as it is not a popular method of attack. The article titled ‘Your timely reminder: Not all hacking requires a computer’ explores the phreaking challenge (organised by the renowned Telephreak members TProphet and Lion Templin) that took place at the 2017 Def Con event. The purpose of the challenge was to explore how effective phreaking attempts can be, either in place of, or to supplement active network hacking (Baldwin, 2017).

As a modern society that seems to be relying increasingly more heavily on technology, it would be easy to assume that hacking attempts only take place in the ‘tech realm’. While that may be true for the majority of them, it seems that old methods die hard and phreaking is still a common attack vector: “Recent research from Pin-drop has confirmed this, revealing that call centre fraud attacks have increased 113% in the past year. A shocking amount of fraud is now occurring over the phone, due to this access point having been neglected. The industry had primarily focused its efforts on cyber defences, but businesses are now turning to voice biometrics to help them tackle fraud” (Gaubitch, Unknown).

As social engineering tactics become stealthier, the development of Deepfake technology advances, Speech Recognition, STT (Speech To Text) and TTS (Text To Speech) API’s improve, developers will have the difficult task of trying to prevent fraud that could take place using any voice based/interactive service.

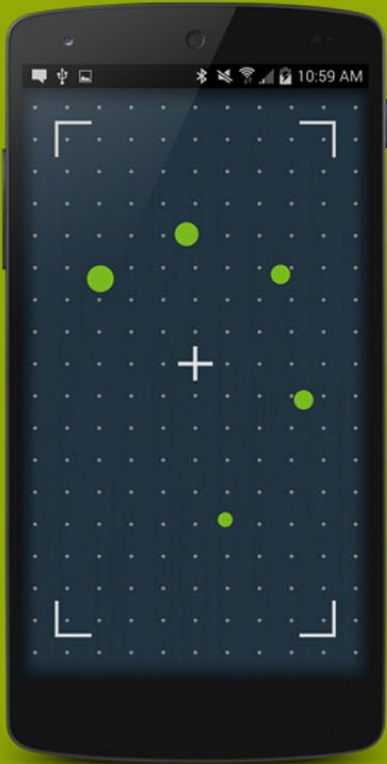
Some researchers have already attempted to mitigate potential threats by using spoof detection. One of these systems aims to discern if the source of the audio is from a speaker or a human (Traynor, 2018). However, using some fairly simple modification to the delivery method – this mitigation system could be overcome. There would also be (rare) instances where a customer is mute due to a medical condition and has to use a speech API to communicate verbally.



Ear Biometrics

The ear as a biometric indicator is as unique as a fingerprint, less invasive to use than an iris or retinal scan, and as natural and intuitive to use as voice recognition.

LEARN MORE



A second mitigation technique is an ear biometrics application. “The ear as a biometric indicator is as unique as a fingerprint, less invasive to use than an iris or retinal scan, and as natural and intuitive to use as voice recognition” (Descartes Biometrics, 2018). The scan takes place when the user puts their smartphone up to their ear during a phone call. However, this would easily be bypassed if a customer used their device on speaker-phone (while driving), or made the call with a head-set (in a noisy environment), certain hairstyles and jewellery could also affect the reliability of the app. This service would also not be available to anyone without a smartphone.

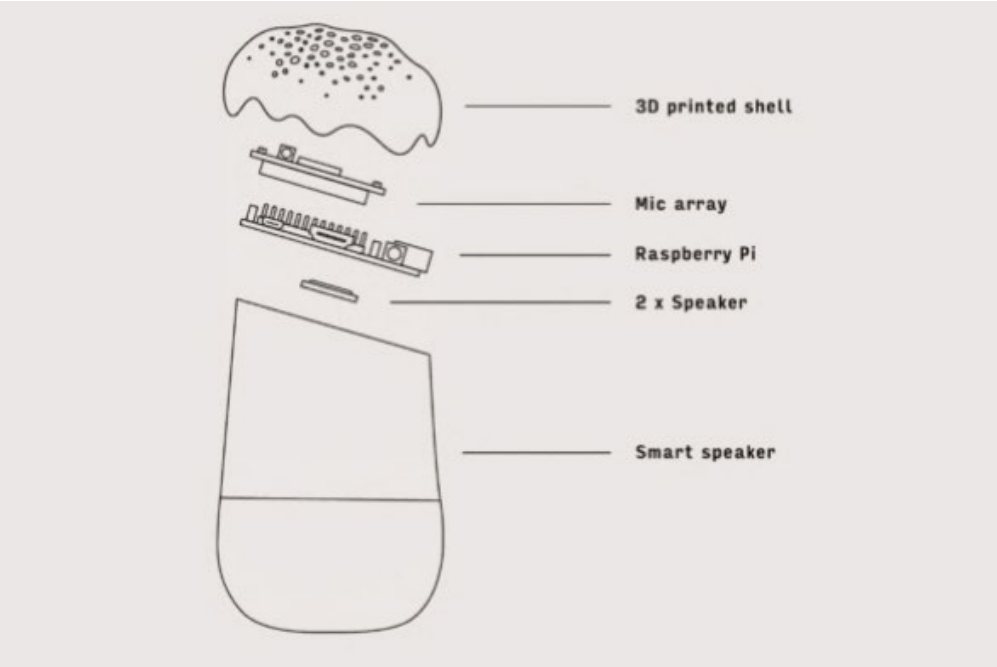
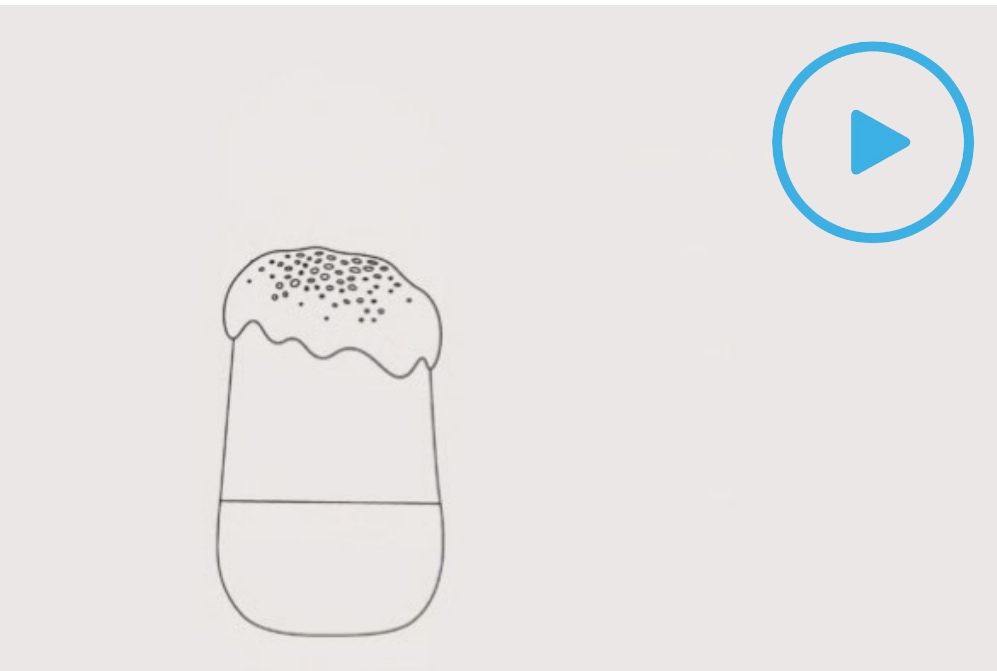
For Smart Speakers specifically, a potential mitigation is a device called the Alias. It covers the Home Speaker’s microphone and plays ‘white noise’ so that the device cannot eavesdrop. The Alias is given a unique ‘activation’ name or phrase by the owner and it recognises that phrase based on local machine learning only (so there is no cloud connection involved). When the Alias device is called, it “whispers” *Hey Google* (or equivalent wake-up command) so that the Home Speaker activates, it then stops playing the white noise so that the user can issue commands and communicate with the base device. (Wilson, 2019). The obvious drawbacks are having to purchase **another** device, that may also hinder the Home Speaker’s functionality. Unfortunately, this product is purely a concept device and does not exist for actual purchase.

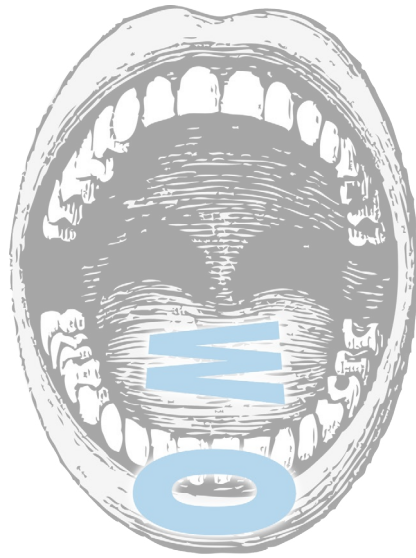
Threat Modelling:

Although not strictly a mitigation tactic, it is possible to assess your threats and go through a security planning process known as threat modelling, by answering the questions below from the Electronic Frontier Foundation (THE ELECTRONIC FRONTIER FOUNDATION, n.d.):

- 1. What do I want to protect? (asset)
- 2. Who do I want to protect it from? (adversary)
- 3. How bad are the consequences if I fail? (a spectrum from mild to disastrous)
- 4. How likely is it that I will need to protect it? (also, a spectrum)
- 5. How much trouble am I willing to go through to try to prevent potential consequences? (very little or a lot)

For example, the data (assets) from an average middle-class citizen, will differ greatly from that of a politician for example. The adversaries aiming to attack the ‘average Joe’ will likely be opportunistic; and not nearly as skilled or organised as a covert operation targeting an opposing government official. For reasons like these, every single individual’s threat model will be unique.





R

D

S

PART 2

DOCUMENTED ATTACKS BY OTHERS

Documented Vulnerabilities and Attacks:

IoT Attacks:

Hacking a Smart Lock by voice

published on 9 November 2018 by CNET:

“Hey Google, unlock the front door” - that is literally how easy it now is to control a new smart lock available on the market. Most smartlock manufacturers require a (voice) passphrase to be said, but there are ways to get around this authentication step, by passing the command directly to a hub that executes the command without authentication.

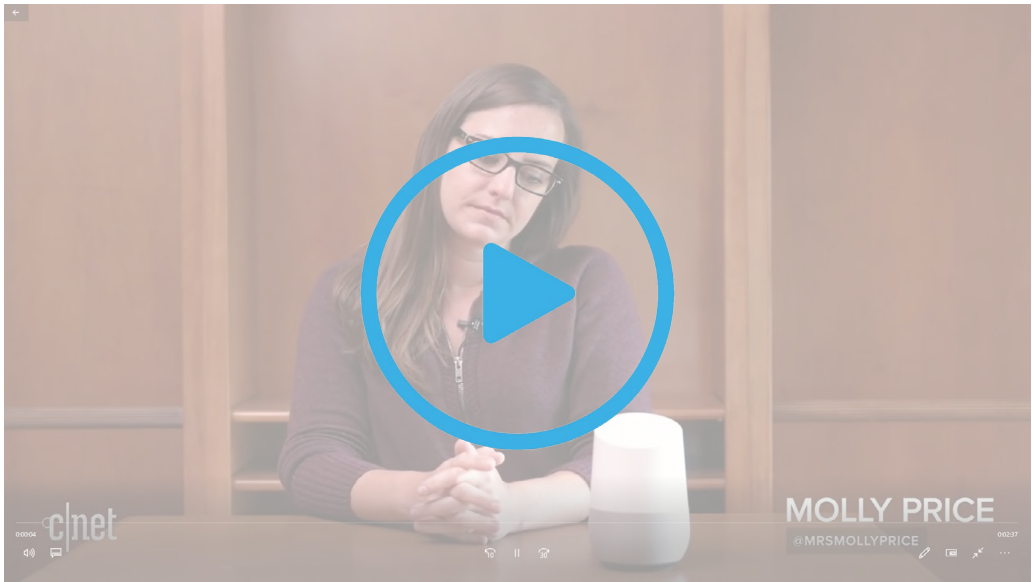
This is incredibly convenient, but also opens smart-lock users up to unauthorised entry. The one explored in this video shows an attacker outside the building, who uses an audio transducer device against a window in line-of-sight to the smart speaker. The attacker then plays a voice recording to the smart speaker, requesting that it unlocks the door, and the smart speaker obliges.

Unlike a conventional speaker, an audio transducer vibrates the entire surface it is placed on, turning the entire thing into a speaker. This can be placed on a window and the sheet of glass will become a speaker (Price, 2018) and (CNET, 2018).

Excerpt from Security Predictions for 2018 Paradigm Shifts,

published on 5 December 2017 by Trend Micro:

The massive Mirai and Persirai distributed denial-of-service (DDoS) attacks that hijacked IoT devices, such as digital video recorders (DVRs), IP cameras, and routers, have already elevated the conversation of how vulnerable and disruptive these connected devices can be. Recently, the IoT botnet Reaper, which is based on the Mirai code, has been found to catch on as a means to compromise a web of devices, even those from different device makers. We predict that aside from performing DDoS attacks, cybercriminals will turn to IoT devices for creating proxies to obfuscate their location and web traffic, considering that law enforcement usually refers to IP addresses and logs for criminal investigation and post-infection forensics. Amassing a large network of anonymized devices (running on default credentials no less and having virtually no logs) could serve as jumping-off points for cybercriminals to surreptitiously facilitate their activities within the compromised network (Trend Micro, 2017).



**Excerpt from The Need for Better Built-in Security in IoT Devices,
published on 27 December 2017 by Trend Micro:**

“Publicly available personally identifiable information (PII) – These can come from legitimate sources such as online search tools or social media, as well as from data breach information made public. In the case study, we found 727 unique email addresses that could be loaded into open source intelligence tools such as Maltego. We also saw several email accounts connected to previously reported breaches such as River City Media, LinkedIn, and last.fm.

The URI led to a site which was used to populate the applications that control the device and also contained information exposed without requiring authentication. It also included information about tracks currently being played, libraries connected to the device, devices used to control the speakers, devices that were in the same network as the speakers, as well as email addresses associated with audio streaming services synced with the device.

This exposure is not limited to home users. In a workplace scenario, an exposed device which identifies and lists down other IoT devices connected to the same network can give an attacker plenty of information to work on. Bad actors could find machines such as printers with existing vulnerabilities and use that to gather further information or as an entry point.”

Voice-Authentication Style Attacks:

**The Alleged Can-You-Hear-Me Scam,
published by The Standard on 7 May 2017**

“...residents have been warned about a scam involving telephone voice recognition. The advice is if you receive a phone call and somebody asks, “can you hear me”, don’t say anything and hang up. The scam has arrived in Australia after being used in the United States and Britain.

The scammer may ask several times “can you hear me?”, to which people would usually reply “yes”. The scammer is then believed to record the “yes” response and end the call.

That recording of the victim’s voice can then be used to authorise payments or charges in the victim’s name through voice recognition. Because it is the person’s own voice authorising transactions, it makes it hard to dispute later if a victim claims they were scammed” (Thomson, 2017).

The extent of the risk, as explained on CSO:

“... Fingerprints are no longer an entirely hack-proof method of authentication – they can be spoofed. That will soon be true of your voice as well. The risk goes well beyond recent warnings from the Federal Communications Commission (FCC) and Better Business Bureau (BBB) about spam callers trying to get a victim to say the word “yes,” which they record and then use to authorize fraudulent credit card or utility charges, or to “prove” that the victim owes them money for services never ordered.

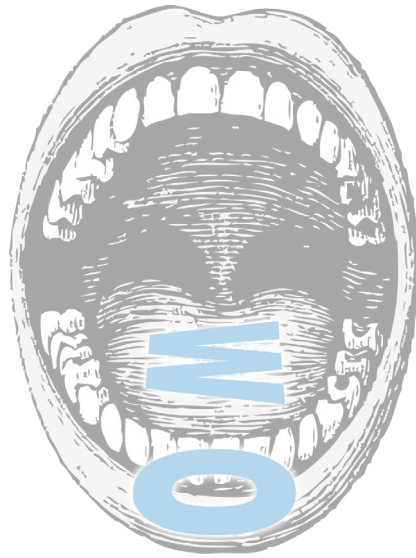
This technology is aimed at ‘cloning’ an individual’s voice accurately enough to make him or her say anything you want. The potential risks are obvious: If your phone requires your voice to unlock it, an attacker with some audio of your voice could do it.

It is not perfect yet. But it is already remarkably close. A demonstration last fall at Adobe Max 2016 of the company’s VoCo, nicknamed ‘Photoshop for voice’, turned a recording of a man saying, “I kissed my dogs and my wife,” into “I kissed Jordan three times.” The audience went crazy. The pitch for the product: “With a 20-minute voice sample, VoCo can make anyone say anything.”

Once perfected, there are numerous possibilities for mischief – well beyond simply creating comedic videos spoofing the voices of your favourite celebrities. Besides undermining voice-based verification, leading to identity theft or other fraud – Santander Bank was running ads just this past week on voice verification– it could eliminate the use of voice or video recordings as evidence in court.

Lyrebird itself, in a brief ethics statement on its website, acknowledges its product, “could potentially have dangerous consequences such as misleading diplomats, fraud and more generally any other problem caused by stealing the identity of someone else.”

“Because voice transformation technologies are increasingly available, it is becoming harder to detect whether a voice has been faked,” said Jan van Santen, a mathematical psychologist at the university...” (Armerding, 2017)



R

D

S

PART 3

ATTACK AND PROOF OF CONCEPT

Hypothesis:

The technology available to spoof voices has already been used to make deepfake videos of famous individuals – who are arguable easier targets, due to the amount of content available online.

Although intensive work is involved to create an attack – is it possible for a threat actor to manipulate recorded data captured by IoT devices and use it during an attempt to impersonate the victim or authenticate their account(s) – without their knowledge or permission.

Attack Scenario 1 – Attack a Target’s Friends/Family:

Procedure:

1. Identify Targets

By listening in on an IoT device, a threat actor could get a good idea of who their target speaks to most frequently. This would include a parent, friends and family members.

2. Collect Data

The voice recordings from the target will need to be collected by downloading all the information identified in the Step 1. The recorded audio from the target has two purposes: Firstly, it allows the threat actor to gather critical information, this can include anything from the target’s name, the names of friends/family, who the target speaks to most often or reaches out to when they need assistance. The second use is to gather adequate high-quality material to be able to make a believable spoof the target’s voice.

3. Process Data

The audio data then needs to be broken up into smaller sections - each string of talking cannot be more than 10 seconds in duration. Any interjections (like “um”) need to be removed from the recordings.

The words spoken then need to be transcribed. Currently the transcription process is quite arduous, but I predict that this process would be expedited as Speech-To-Text API’s improve.

This data is then fed to a model for training.

4. Prepare Attack

The threat actor could desire any number of outcomes, like attacking the target’s nearest and dearest. They would need the name and contact details of friends and family. If the threat actor only had a name, they could find the person’s contact number from TrueCaller or a social media profile. The Attacker also requires a service where their call identity is hidden or spoofed.

5. Carry Out Attack

There are many unscrupulous avenues an attacker could take:

- Example 1: The attacker could use the Target’s spoofed voice to explain that they have a problem, and need money transferred urgently. The attacker could then provide their own banking details or ask for cash to be dropped off at a certain location.
- Example 2: The attacker could request help, so that the friend/family member comes to meet them. This individual could then be used to open up the building (if they have keys to that property), or could be robbed of their possessions.
- Example 3: An attacker could use the spoofed voice to make prank calls, or say embarrassing phrases, to ruin the target’s reputation

Attack Scenario 1 - Proof of Concept:

Using my Lyrebird synthesised voice (please refer to the detailed explanation in *Attack 2 Proof of Concept*). I prepared several ‘attack scripts’ to use on my unsuspecting friends and family. I found that the synthesised audio was most believable when kept quite short and simple:

Proof of Concept 1A:

EXTREF: I sent a WhatsApp voice-note to FRIEND, a friend of 26+ years. The voice-note asks her to call me. The voice-note was sent while FRIEND was online, and she responded almost immediately by calling.

She was not aware that the voice-note was synthesised and not my real voice!

Proof of Concept 1B:

I created an attack script to send to my father. This script said I had a problem, that I needed R200 and requested he make an EFT into my account.

My dad responded two hours later (unfortunately I missed the call) but he left a voicemail stating that he could drop the cash off at my house.

My father was not aware that the voice-note had been synthesised and that it was not me speaking.

Disclaimer:

No friends or family were harmed during the making of this Proof of Concept.

Conclusion:

Short phrases, with simple requests managed to successfully fool two people that have known me most of my life!

I have no doubt that a criminal could think of many more instances that this attack vector could be used in. That said, in principal, this type of attack uses basic social engineering tactics, executed via the target’s voice. The target’s family/friends may think they are helping in good faith – instead they become the victims of a scam by being manoeuvred into a potentially dangerous or fraudulent situation.

Attack Scenario 2 – Attack Target’s Bank Account:

Procedure:

1. Identify Targets

A. The Primary Target

By compromising vulnerable IoT devices, the attacker can identify suitable targets. These would be people who conduct any business or personal transactions via telephone calls, and say important information out loud (ID number, account number etc.).

B. The Secondary Target

The second target is the service provider the attacker knows the primary target uses. This would need to be an institution where transactions that take place as telephone conversations are the norm.

2. Collect Data

The voice recordings from the primary target will need to be collected by downloading all the information identified in the Step 1A. The recorded audio from the target has two purposes: Firstly, it allows the threat actor to gather critical information, this can include anything from the target’s name, identity number, to their address etc. The second use is to harvest enough high-quality material to be able to make a believable spoof the target’s voice.

3. Process Data

The audio data then needs to be broken up into smaller sections - each string of talking cannot be more than 10 seconds in duration. Any long pauses or interjections (like “um”) need to be removed from the recordings.

The words spoken then need to be transcribed. Currently the transcription process is quite arduous, but I predict that this process would be expedited as Speech-To-Text API’s improve.

This data is then fed to a model for training.

4. Prepare Attack

While the model is trained, the attack script can be prepared. Once the model is complete, the actual phrases can be saved as output. It would be quite believable for a threat actor to pre-empt a basic transactional conversation, by preparing certain key phrases in the victim’s voice.

Some information that would need to be generated includes:

A. Who is making the call?

- Attacker generates phrases including the victim’s name, identification number and other personal information like physical addresses etc. that would be needed to identify themselves or answer security questions.

B. What is the purpose of the call?

- *Attacker prepares sentences detailing the funds transfer/withdrawal required.*

5. Carry Out the Attack

The attacker is then able to make the call to the secondary target (B), pretending to be the first target (A). The attacker then plays the primary target's 'voice', that (for example) request a certain amount of money to be transferred from one account to another. If the transfer is completed, then the attack is successful.

6. The Secondary Target - **BANK**:

Marketing information from the **BANK** website:

Telephone Banking is a phone call away

This service allows any customer to bank remotely, via speaking to a consultant or to an interactive assistant.

Telephone banking as a service allows customers to transfer funds between accounts, or pay any other account (from the same bank or a different bank).

All that is required is for a customer to register by calling in, and once complete, they will be able to complete telephone transactions.

Attack Scenario 2 - Proof of Concept:

The following steps were followed on the practical portion of the research:

A. Lyrebird

The Lyrebird website contains audio samples from Donald Trump and Barack Obama – accessible at: <https://lyrebird.ai/vocal-avatar-api>. The ethics statement released by Lyrebird explains that this was done to raise public awareness regarding the technology (and its associated risks).

I created a personal Lyrebird account in March 2019 and proceeded to complete 300 short verbal recordings. The synthesised output had a 'machine-like' twang, but definitely sounded like my voice. If some background noise was present when the generated audio was played, it would be difficult for someone who knows me to differentiate between the synthesised audio and real live-spoken words.

When selecting and preparing phrases, I took the following approach:

- Spelling words correctly often generated a slightly American sounding accent. By manipulating the spelling, I was able to make the pronunciation sound more South African.
- Adding in pauses (with punctuation) as well as *umms* and *errs* made the speech sound *slightly* more realistic.
- If generating phrases with numbers (like Social security/Identity, Account or Cellphone number), I would switch between synonyms where possible (like inputting 'oh' instead of 'zero') and group digits together - to read as 'double one' instead of 'one, one' (which sounded very mechanical).

B. Tacotron

Although Google did not officially release their Tacotron software, certain versions of it were easily available on several GitHub profiles. Even the manufacturer NVIDIA, released a version where processing times could be improved by enabling a graphics card to be used during the machine learning process.

Setting up Tacotron (and all its dependencies) is a lot more time intensive than using Lyrebird. My results from this application can be improved; this was due to time constraints – as good quality audio output is definitely possible. In the longer term, mastering Tacotron far outperforms the audio that is generated with Lyrebird.

BANK ATTACK

Attack Day 1: Setting Up Telephone Banking

I called through to the BANK enquiries line on 11 April.

Finding the telephone banking section was a bit tricky. The one consultant I got through to told me that I could not do telephone banking, only cellphone banking (via USSD) or internet banking (via the smartphone application).

On a second call, I was able to locate the automated telephone banking section. To register for telephone banking I was required to enter my identity number, the last 6 digits of any of my bank cards that had a PIN, and the PIN for that card. It was a quick and easy process to complete.

I was then successfully registered, with a reference number provided by the automated service. However, no notification was SMSed or emailed to me after I had registered for telephone banking.

Attack Day 2: Testing Telephone Banking

I called BANK on 12 April 2019.

This call was different to the ones on 11 April, because the Interactive assistant answered my first call and asked me what I would like to do. This worked well, but I had to unexpectedly end the call due to circumstances beyond my control. Subsequent calls made later that day provided me with the standard automated menu - none of these calls went through to the Interactive assistant (even though I called the exact same number).

EXTERNAL REFERENCE: In video 1, you can hear the virtual assistant.

EXTERNAL REFERENCE: In video 2, I get through to a different section of the self-help menu and am asked to enter an “access number” if I have registered for telephone banking – this access number had not provided when I set up the Telephone Banking several days prior.

Note: For a transfer of funds, the account holder’s ID number and card number are required. Information a threat actor could believably have access to.

Attack Day 3: Attack Testing

16 April 2019.

I called BANK eleven times during the hour of testing and was still unable to get through to the Interactive assistant and I suffered many network related issues (**EXTERNAL REFERENCE:** Video 3, the call would go silent when navigating the menu, and then drop).

EXTERNAL REFERENCE: Video 4, When I did get through to a consultant, she struggled to hear the audio I was playing from the laptop.

The two consultants that I did “speak” to could not adequately hear the audio, and both of them ended the call. I noted that there was no correspondence was sent to me saying that attempts had been made to access my account.

Following my series of unsuccessful calls, I began to investigate other means to make the attack possible:

- Using a paid, online telephone service like Skype, I was able to make calls with a number (e.g. 0100350XXX) that was in no way directly linked to my identity (confirmed by searching on the TrueCaller application). This would make tracing or blocking the call more difficult from the bank’s side.
- The research will continue to explore the best methods to deliver audio that is generated from a computer, and needs to be clearly audible on a phone call.

Conclusion:

As I was able to successfully fool friends and family with the initial Proof of Concept, this type of attack can be used against a financial institution (with or without voice authentication and/or recognition capabilities).

As far as safety concerns go - voice-based verification and other biometrics definitely have a place in the authorisation process, as long as it is coupled with other means of verification.

The era of biometrics was heralded as ‘the end of passwords’, however I see it as quite the opposite.

A physical token (such as a verified device), a verbal pass-phrase, as well as a password should be used during authentication. All three factors combined would be a formidable feat for any hacker to spoof. It is highly unlikely that a threat actor would be able to gain access to all devices and information, unless the victim is forced to make a call under duress.

Please note that many of the external references have not been included in the blog post, as vast amounts of PII and sensitive information are contained in them.



Attack Scenario 3 – Attack a Bank’s Reputation

Procedure:

1. Identify the Target

A Bank’s relationship with its customers is built on trust and as expected, that institution’s entire reputation is based off its perceived trustworthiness.

In the age of online content, anyone with a social media account can post information. Often, the more salacious a post seems, the more attention it will garner. Most social media platform users are not discerning of whether a post is factual or not. Although all large social media platforms have pledged against fake news, it is incredibly hard for both humans and AI to detect what is true and what is not.

Now imagine if a *clickbait-titled deepfake video* post was released on social media – showing a Bank CEO at a media briefing, where a statement is released that they are supporting a controversial political party prior to a major election; or releasing a statement that huge amounts of money had been stolen and affected customers would not be compensated.

Regardless if the information contained in the deepfake video is completely false, this would have a large impact on the bank’s corporate image. There would be a significant damage done to the Bank’s reputation, before they could respond to set the record straight.

2. Collect Data

A prominent CEO, like PERSON from BANK, would be an ideal target. PERSON has a large online presence and collecting both images and/or video footage of their appearance is as simple as typing their name into Google or YouTube.

3. Process Data

Although the official FakeApp website was taken down (and the official version of the software with it) I was able to both find and successfully install the FakeApp software on my computer. Finding the software was done via a google search - I did not have to resort to ‘the dark web’ to find it. It is fairly simple to install, as long as you have an internet connection and the appropriate hardware required to install and run the application.

4. Prepare Attack

Preparing the attack would take some time, mostly because the neural network deep learning process required to make Deepfake videos possible is lengthy.

However, the actual process is quite straightforward – as FakeApp is able to extract faces of the victim from photographs and video. This face is then transposed onto the sample footage – and voilà!

5. Carry Out the Attack

Posting this Deepfake video on several social media pages would be the start of the process. The attackers could even pay to promote the initial post until the post gains momentum of its own.

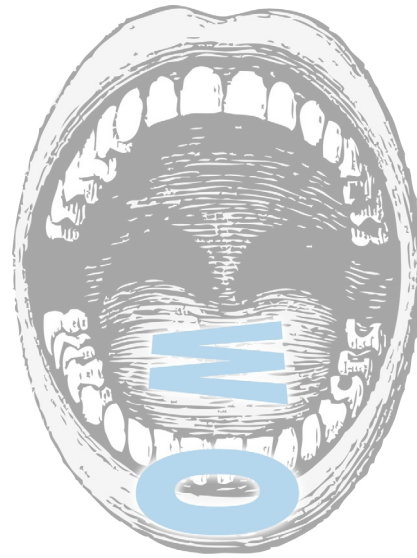
No compromise of the bank’s website, or social media accounts is necessary. The attack plays out completely independently of any of the bank’s online footprint.

Conclusion:

A scenario like this would be an absolute PR nightmare for any bank, anywhere in the world. Additionally, large institutions are unlikely to be prepared for attacks of this nature and would be on the back foot should anything remotely similar to this kind of attack occur.

To safeguard against this, there could be a way for banks to cross reference any press release/ statement/special offer or advert that is released by the institution. For example, a reference number could be made visible on any of these items, that links back to a page on the bank’s website.

This way, customers can cross reference the official website for evidence that the information seen in an ad, image or video is indeed legitimate. In the event that a Deepfake is released, there will be an incorrect or missing reference on the website – revealing it as a counterfeit.



R

D

S

PART 4

BIBLIOGRAPHY

1. Bibliography:

- Armerding, T. (2017, May 16). *Vocal theft on the horizon. Using your voice for authentication is about to get more risky, thanks to voice-spoofing technology.* Retrieved from <https://www.csoonline.com/article/3196820/vocal-theft-on-the-horizon.html>
- Ayinla, R. (2019, February 1). *17 Statistics and Facts About Voice Assistants.* Retrieved from <https://codesmithdev.com/17-statistics-that-explain-the-rise-of-voice-assistants/>
- Baldwin, R. (2017, August 16). *Your timely reminder: Not all hacking requires a computer.* Retrieved from <https://www.engadget.com/2017/08/16/your-timely-reminder-not-all-hacking-requires-a-computer/>
- Banker SA. (2015, July). *Banker Sa: Edition 14.* Retrieved from <https://www.banking.org.za/docs/default-source/publication/banker-sa/banker-sa-14.pdf?sfvrsn=8>
- Biswas, J. (2018, January 10). *Behind Tacotron 2: Google's Incredibly Real Text To Speech System.* Retrieved from <https://www.analyticsindiamag.com/tacotron-2-google-ai-text-to-speech-system/>
- Bloomberg. (2018, June 5). *This AI Can Clone Any Voice, Including Yours.* Retrieved from <https://www.youtube.com/watch?v=VnFC-s2nO-tl>
- Brice-Saddler, M. (2019, January 28). *Family says hacked Nest camera warned them of North Korean missile attack.* Retrieved from <https://www.denverpost.com/2019/01/28/north-korean-missile-attack-false-warning/>
- Burt, C. (2018, February 16). *Australian Tax Office has collected 3.4 million voiceprints as government embraces biometrics.* Retrieved from <https://www.biometricupdate.com/201802/australian-tax-office-has-collected-3-4-million-voiceprints-as-government-embraces-biometrics>
- Capps, R. (2018, April 9). *Physical and passive biometrics: finding the right security balance.* Retrieved from <https://www.biometricupdate.com/201804/physical-and-passive-biometrics-finding-the-right-security-balance>
- Clark, B. (2018, February 21). *Deepfakes algorithm nails Donald Trump in most convincing fake yet.* Retrieved from <https://thenextweb.com/artificial-intelligence/2018/02/21/deepfakes-algorithm-nails-donald-trump-in-most-convincing-fake-yet/>
- CNET. (2018, November 9). *Hacking a smart lock by voice.* Retrieved from <https://www.youtube.com/watch?v=2CkXTR-zyHI>
- Coomes, K. (2018, June 21). <https://www.digitaltrends.com/home/the-best-ai-assistants/>. Retrieved from <https://www.digitaltrends.com/home/the-best-ai-assistants/>
- Derpfake. (2018, February 19). Retrieved from https://www.youtube.com/watch?time_continue=1&v=hoc2RISoLWU
- Descartes Biometrics. (2018). *Biometrics.* Retrieved from <http://www.descartesbiometrics.com/ergo-app/>
- Durden, T. (2019, April 11). *Busted: Thousands Of Amazon Employees Listening To Alexa Conversations.* Retrieved from <https://www.zerohedge.com/news/2019-04-10/global-network-amazon-employees-listening-alexa-conversations>
- Edwards, S. (2017). *Digital Voice Assistants: The New Front in the War on IoT Hackers.* Retrieved from <http://blog.trendmicro.co.uk/digital-voice-assistants-the-new-front-in-the-war-on-iot-hackers/>
- Etienne, S. (2018, July 27). *How to hear (and delete) every conversation your Google Home has recorded.* Retrieved from <https://www.theverge.com/2018/7/20/17594802/google-home-how-to-delete-conversations-recorded>
- FHEM. (Unknown). *FHEM Homepage.* Retrieved from <https://www.fhem.de/>
- Gaubitch, N. (Unknown). *Why Voice Verification is the Future of Authentication.* Retrieved from <https://www.infosecurity-magazine.com/opinions/voice-verification-authentication/>
- Golden. (Unknown). *Tacotron 2.* Retrieved from https://golden.com/wiki/Tacotron_2
- Google. (Date unclear). *Tacotron (/täkō, trăn/): An end-to-end speech synthesis system by Google.* Retrieved from <https://google.github.io/tacotron/>
- Gool, K. (2018, March 14). *Balancing safety and convenience with biometrics.* Retrieved from <https://www.itweb.co.za/content/KA3W-wqdlnQoqrydZ?>
- IMDb. (Unknown). *Sneakers.* Retrieved from <https://www.imdb.com/title/tt0105435/>
- IoT For All. (2017, May 10). *The 5 Worst Examples of IoT Hacking and Vulnerabilities in Recorded History.* Retrieved from <https://www.ietfforall.com/5-worst-iot-hacking-vulnerabilities/>
- itweb. (2018, February 6). *SA consumers ready to say goodbye to passwords.* Retrieved from <https://www.itweb.co.za/content/dgp-45MaGR6ovX9l8>
- Kikel, C. (2018, July 6). *A Brief History of Voice Recognition Technology.* Retrieved from <https://www.totalvoicetech.com/a-brief-history-of-voice-recognition-technology/>
- Long, E. (2018, April 28). *5 Ways to Secure Your Google Home Device.* Retrieved from <https://www.tomsguide.com/us/secure-google-home,news-27076.html>
- Lourie, G. (2017, February 15). *How voice and facial recognition can secure your online identity.* Retrieved from <https://www.thesolution-slab.co.za/how-voice-and-facial-recognition-can-secure-your-online-identity/>
- Luke. (2018, May 7). *7 Awesome Smart-Devices to make the most out of your Smart Setup.* Retrieved from <https://smartspeakers.co.za/7-smart-devices-smart-setup/>
- Lyrebird. (2018). Retrieved from <https://lyrebird.ai/>
- Malan Lourens Viljoen Incorporated. (2017, October 4). *THE RISKS OF VERBAL AGREEMENTS.* Retrieved from https://www.mlvlaw.co.za/Resources/2017-12_The_Risks_of_Verbal_Agreements.pdf

- Malinga, S. (2018, February 7). *Millennials more likely to use biometric authentication*. Retrieved from <https://www.itweb.co.za/content/KWEBBvyakmOvmRjO?>
- Nuance Communications. (2014, November 18). *Organizations Around the World are Using Nuance Voice Biometrics to Make Passwords and Security Questions a Thing of the Past*. Retrieved from <https://www.nuance.com/about-us/newsroom/press-releases/global-voice-biometrics-adoption.html>
- OneVault. (2013-2019). *Where are voice biometrics applied?* Retrieved from <http://onevault.co.za/where-are-voice-biometrics-applied/>
- OWASP. (2017, February 14). *IoT Security Guidance*. Retrieved from https://www.owasp.org/index.php/IoT_Security_Guidance
- Peel, A. (2019). *FakeApp*. Retrieved from <https://www.malavida.com/en/soft/fakeapp/>
- Powers, B. (2018, February 27). *Adobe is Developing Photoshop for Your Voice*. Retrieved from <https://medium.com/s/story/adobe-is-developing-photoshop-for-your-voice-f39f532bc75f>
- Price, M. (2018, November 6). *Are you setting your smart lock up for intrusion?* Retrieved from <https://www.cnet.com/news/how-you-might-be-setting-your-smart-lock-up-for-intrusion/>
- Puiui, T. (2019, March 13). *The amazing Lyrebird can not only mimic other birds, but also chainsaws, theme songs, and car alarms — anything, basically*. Retrieved from <https://www.zmescience.com/ecology/animals-ecology/the-amazing-lyrebird/>
- Radu, M. (2015, July 20). *Knight Rider KITT Car Gets a Futuristic Makeover with Racing Spoilers*. Retrieved from <https://www.autoevolution.com/news/knight-rider-kitt-car-gets-a-futuristic-makeover-with-racing-spoilers-97981.html>
- RankRed. (2018, January 3). *Google Develops Voice AI That Is Indistinguishable From Humans | Tacotron 2*. Retrieved from <https://www.rankred.com/google-develops-voice-ai-tacotron-2/>
- Robert. (2018, April 9). *Physical and passive biometrics: finding the right security balance*. Retrieved from <https://www.biometricupdate.com/201804/physical-and-passive-biometrics-finding-the-right-security-balance>
- Seals, T. (2019, March 25). *Bugs in Grandstream Gear Lay Open SMBs to Range of Attacks*. Retrieved from <https://threatpost.com/grandstream-bugs-smbs-attacks/143141/>
- Seals, T. (2019, March 5). *RSA Conference 2019: How to Be Better, on Trust, AI and IoT*. Retrieved from <https://threatpost.com/rsa-trust-ai-iot-keynotes/142505/>
- Seymour, J. (2018, October 22). *DEF CON 26 - delta zero and Azeem Aqil - Your Voice is My Passport*. Retrieved from <https://www.youtube.com/watch?v=2uoOkIUB43Q>
- Stephen Hilt, N. H. (2019, March 5). *Exposed IoT Automation Servers and Cybercrime*. Retrieved from <https://blog.trendmicro.com/trendlabs-security-intelligence/exposed-iot-automation-servers-and-cybercrime/>
- Tencent Blade Team. (2018, October 22). *DEF CON 26 - HuiYu and Qian - Breaking Smart Speakers We are Listening to You*. Retrieved from <https://www.youtube.com/watch?v=3sLCoXaqvMg>
- The Banking Association South Africa. (Unknown). *Protection of Personal Information Act (POPI Act)*. Retrieved from <https://www.banking.org.za/what-we-do/market-conduct/regulatory-framework/popia>
- THE ELECTRONIC FRONTIER FOUNDATION. (n.d.). *Your Security Plan*. Retrieved from <https://ssd.eff.org/en/module/your-security-plan>
- Thomson, A. (2017, May 7). *Police issue warning over phone voice scam ‘can you hear me?’*. Retrieved from <https://www.standard.net.au/story/4644792/police-warn-of-phone-voice-scam/?cs=72>
- Traynor, P. (2018, June 18). *Twitter*. Retrieved from <https://twitter.com/patrickgtraynor/status/1008685716384157696>
- Trend Micro. (2017, December 5). *Security Predictions for 2018*. Retrieved from <https://www.trendmicro.com/vinfo/us/security/research-and-analysis/predictions/2018>
- Trend Micro Research. (2019, March 5). *SECURING SMART HOMES AND BUILDINGS: Threats and Risks to Complex IoT Environments*. Retrieved from <https://www.trendmicro.com/vinfo/us/security/news/internet-of-things/threats-and-risks-to-complex-iot-environments>
- voicevault. (Unknown). *VOICESIGN Voice Biometrics E-Signatures*. Retrieved from <https://voicevault.com/products/voicesign/>
- Wagenseil, P. (2018, April 23). *How to Make Sure Alexa, Google Home Don't Hear Too Much*. Retrieved from <https://www.tomsguide.com/us/alexa-google-home-privacy,news-27038.html>
- Warzel, C. (2018, August 27). *I Used AI To Clone My Voice And Trick My Mom Into Thinking It Was Me*. Retrieved from <https://www.buzzfeednews.com/article/charliwarzel/i-used-ai-to-clone-my-voice-and-trick-my-mom-into-thinking>
- Wickes, J. (Unknown). *What is the Standard for IoT Security?* Retrieved from <https://www.infosecurity-magazine.com/opinions/standard-iot-security/>
- Wikipedia. (2019, February 12). *Adobe Voco*. Retrieved from https://en.wikipedia.org/wiki/Adobe_Voco
- Wikipedia. (2019, April 1). *Deepfake*. Retrieved from <https://en.wikipedia.org/wiki/Deepfake>
- Wilson, M. (2019, January 14). *This is the first truly great Amazon Alexa and Google Home hack*. Retrieved from <https://www.fastcompany.com/90290703/this-is-the-first-truly-great-amazon-alexa-and-google-home-hack>
- Wollerton, M. (2017, March 30). *Control a smart lock with your voice: Good idea or bad idea?* Retrieved from <https://www.cnet.com/news/controlling-locks-with-your-voice-good-idea-or-bad-idea/>

THANKS FOR
READING!

FIN