Dolphin Attack on Smart Home systems

Presented by : Aadithya Venkatanarayanan and Nrithya Theetharappan

Motivation

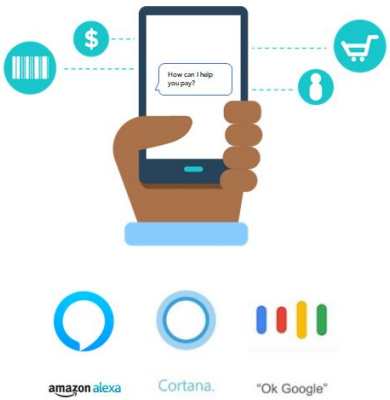
- With the rise of Artificial Intelligence, Machine Learning and Natural Language Processing, Speech Recognition systems have increasingly gained popularity as Human Computer Interaction Methods
- Several Smart systems have turned into Voice Control Systems
- They have access to security critical applications like door locks and other sensitive data from the environment or smartphones
- How secure are these Systems?
- How viable is it to attack these systems?

Prior Work

- Prior work focuses on attacking these systems using commands hidden in sounds that are incomprehensible to humans [3],[4]
- Another class of “silent” attacks involve the use of intentional Electromagnetic interference through wireless coupling with headphone cables, to control the Speech Recognition Systems [5]
- Another interesting, but unintentional, attack on the initial class of Alexas was through radio signals. It was found that Amazon Echos all over the country were always listening and got triggered by news articles on them, to increase thermostat values and shop online [6]
- Voice Commands are also embedded in songs and made to attack voice control systems [7]
- Another attack, quite close to the dolphin attack, utilized 2 signals in the inaudible range to exploit microphone non linearities and attack VCS [8]

What is Dolphin Attack

- A completely inaudible Attack on the Intelligent Speech Recognition (SR) systems developed by Google, Amazon etc.
- Accomplished by playing ultrasonic sound (inaudible) in the vicinity of any device with these SR systems [1]
- The non linearities of the microphones in mobile phone's receiver demodulate the original low frequency components used to modulate the ultrasound carrier
- These low frequency harmonics are picked up by the SR systems
- Wide range of devices are vulnerable to this attack [2]



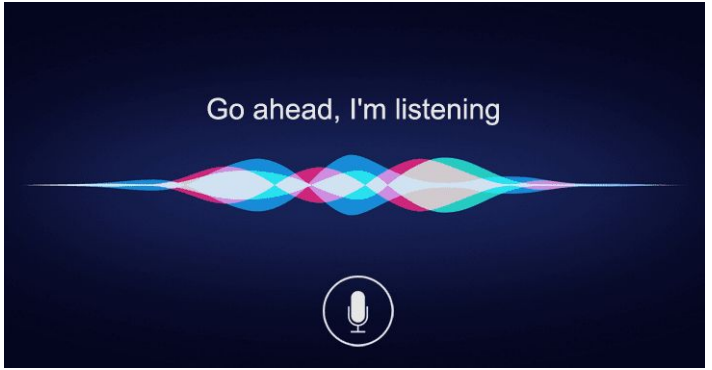
Types of attacks

Two types of attacks are possible:

Benchtop attack : makes use of a vector signal generator for amplitude modulation and a custom speaker for ultrasound transmission. The attack feasibility is tested using this setup by the original author's.

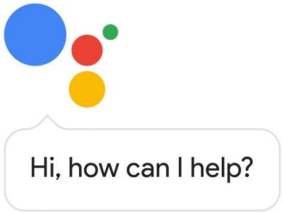
Portable Attack : It makes use of smartphone as the source of modulated signal and transmission is via an ultrasonic transducer. This attack is used to evaluate the feasibility of an attack on the go.

Our Attack



Why Portable attack?

- Low Cost
- Can be implemented with COTS devices
- Table top attack expensive and not easy to carry everywhere



Theoretical Analysis

- All phones and smart-home devices use MEMS microphone
- Analog circuitry involve non-linearity
- This can be modeled as :

$$S_{out}(t) = A*\sin(t) + B*(\sin^2(t))$$

- Amplitude Modulation Function used :

$$\sin(t)=m(t)*c(t) + c(t), \text{ where } c(t)= \cos(2*pi*F_c*t)$$

- Because of the nonlinear component, the following harmonics are produced at the receiver side:

$$f_m,\ 2(f_c-f_m),\ 2(f_c+f_m),\ 2f_c,\ 2f_c+f_m, \text{ and } 2f_c-f_m$$

Analysis(cntd)

- Thus the fundamental component f_m, should be reconstructed at the receiver and recognized by the SR systems
- However, because it is transmitted by modulating on an ultrasonic carrier, the attack commands remain inaudible to the user

Initial Attack Setup

- The portable attack is carried out using a Samsung S7 as the source of the modulated attack signal
- The audio signal is transmitted to a Class D audio amplifier by means of a stereo 3.5 mm audio jack
- The amplifier is powered by a 4.7V battery.
- The amplifier output is connected to a 40kHz ultrasonic transducer
- The victim phone is placed at varying distances from the transducer and it’s SR system is tested

Tools utilized

Hardware

- 3.5 mm Audio Jack pigtail-mono and stereo
- SparkFun Class-D audio amplifier TPA2005D1
- Ultrasonic transducer (UTR-1440K-TT-R)
- Samsung S7 Edge as Attack phone
- Multiple Smartphones as Victim
- 5V AA Battery

Software :

- Matlab
- Audacity
- Mbed Online compiler

Component Selection

S7 Edge :

- The attack requires (carrier frequency - baseband frequency.) to be greater than 20 kHz.
- The minimum sampling rate should be twice this value.
- Most smart phones only support a maximum sampling rate of upto 48 kHz, restricting the transmitted signal to a frequency of 24 kHz
- This does not give us a wide range to work with.
- Fortunately, the Samsung S7 Edge supports a sampling rate of 192 kHz and lends itself well to the attack

Component Selection

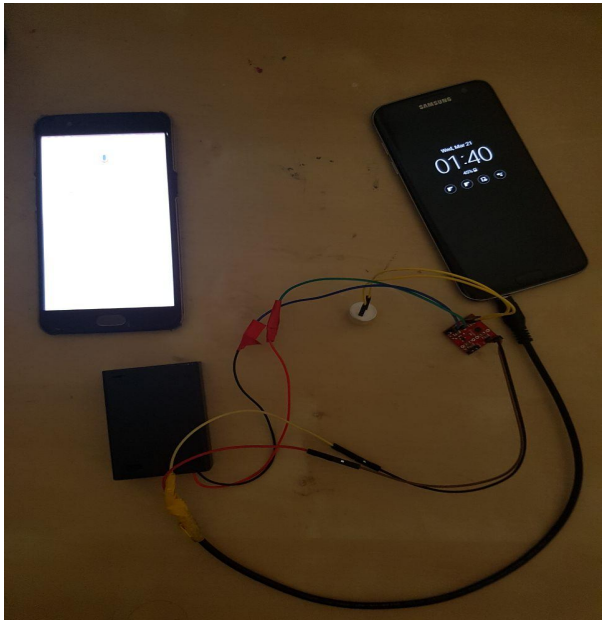
Ultrasonic Transducer

- The Samsung phone provides a sufficient sampling rate, however it restricts it's output frequencies to the audio range
- Thus a narrow band transducer is utilized for transmission of the attack signal
- This particular carrier is chosen as it was the most widely available as opposed to other frequency ranges like 23 or 25 kHz.

Initial Attack Steps

- Record human voice and Import to MATLAB
- High frequency modulation of the audio in MATLAB
- Play the audio via the audio jack hooked up from the phone to the transducer via the amplifier
- Utilized Oneplus 5 and Xiaomi Redmi Note 4 as the initial test devices
- Also tested on a Google Home Mini
- Attack did not succeed in the initial phase of testing

Image of initial setup



Challenges

Audio Jack

- The audio jack output was analysed on an oscilloscope
- Seemed to limit the output that was played from the phone to less than 15 kHz.
- Tested by playing individual sound frequencies (Sine waves) and also sweeping audio frequencies to ascertain the limit
- The Stereo audio jack had a cutoff at 14 kHz and the Mono audio jack at 16 kHz.
- The same tests were performed using apple earphones to obtain the same results

Challenges

Non Linearity Modeling

- Nonlinearity model for the microphone at the receiver seemed to hold for the speaker on the attack side as well
- This hypothesis was tested by connecting a probe from the laptop to the oscilloscope and playing the high frequency signal
- Components within the audible range were observed
- This caused frequency components in the audible range, that hindered the proper recognition of commands, at the receiver end

Testing Video

1) https://drive.google.com/open?id=1DrHbMIyVrRouhfZZqh9IUGmbUZB9vmuu

Tools Utilized

Hardware

- 3.5 mm Audio Jack pigtail-mono and stereo
- SparkFun Class-D audio amplifier
- Ultrasonic transducer
- STM32 Nucleo
- Scope Analyzer
- Battery

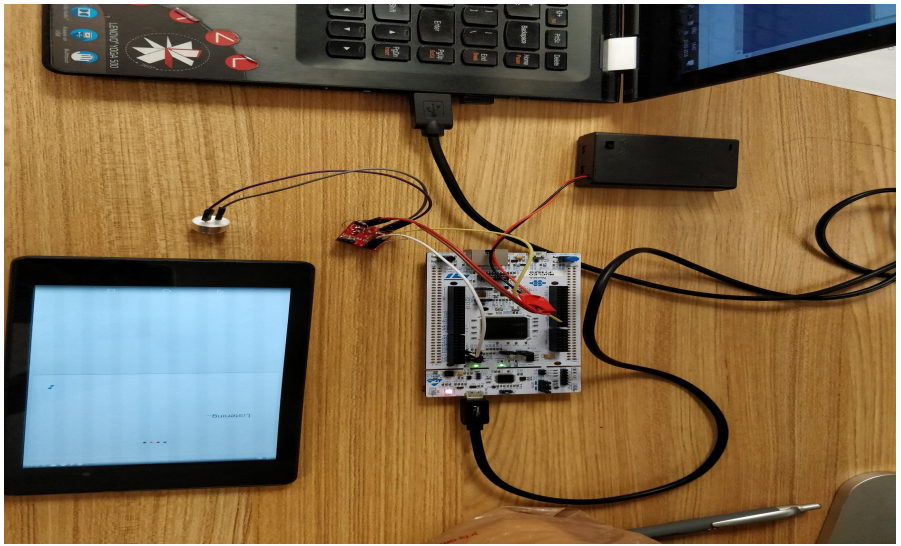
Software :

- Matlab
- Audacity
- Mbed Online compiler

Revised attack Setup

- The phone and audio jack were thus eliminated as the source of the attack signal
- Audio was played out of a STM32 Nucleo microcontroller with a 12 bit DAC
- The output of the DAC is connected to the audio amplifier
- The audio amplifier is used to drive the ultrasonic transducer
- The victim is placed at varying distances from the transducer

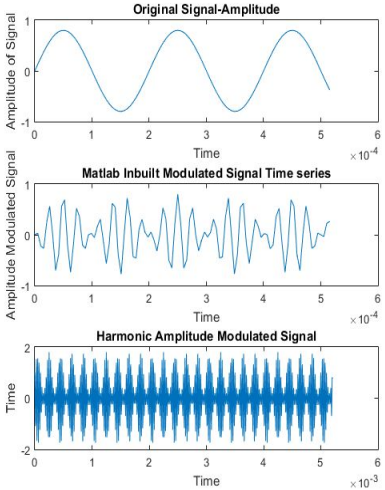
Image of Revised Attack

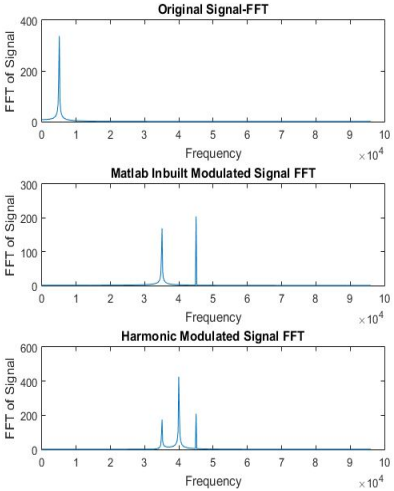


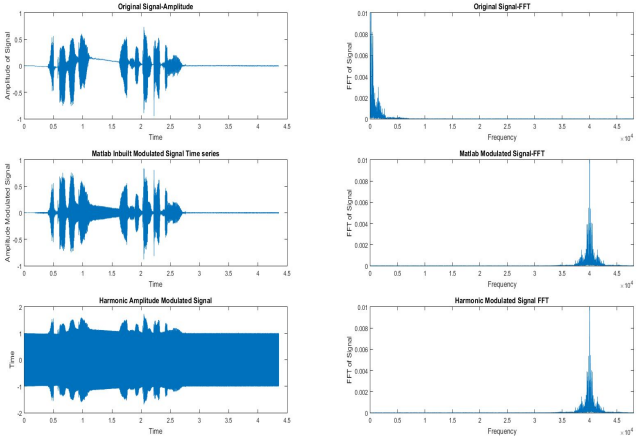
Revised Attack Steps

- The viability of the attack path was tested using montone signals
- Audacity used as the tool to generate monotone sine waves
- The audio signal, sampled at 96khz was modulated on MATLAB
- The wav file was converted to an array of samples stored in a C file using a tool called WAVtoCode Converter
- Mbed online compiler was used for generating the binary
- The generated binary was flashed in the flash memory of the Nucleo board
- The waveform of the signals at the transmitting and receiving end were generated and analysed using matlab and audacity respectively

Plots from matlab and audacity

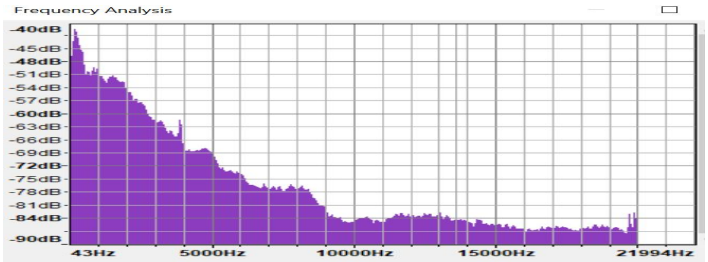






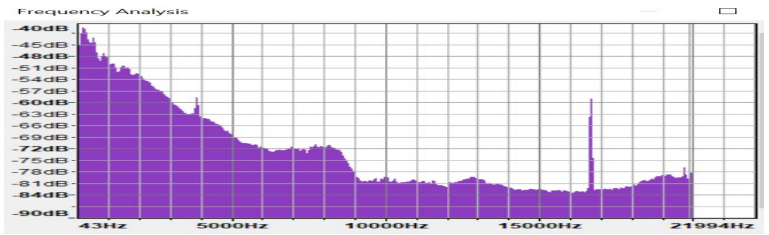
Plots from matlab and audacity

3khz monotone

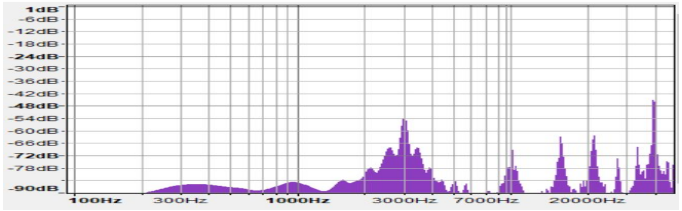


30khz

30khz monotone



Received voice signal



Observations

- In the case of the low frequency monotones, the spectrum of the modulated signals was observed at the carrier frequency and it's two sidebands at the transmitter side
- However, the expected harmonic on the receiver side could not be distinguished by plotting the spectrum of the signal recorded on the victim microphone
- This was because of ambient noise in the range of a few hertz to almost 10khz
- When the frequency of the baseband signal is increased beyond the audible range, definite harmonics are observed on the receiver side
- This is tested for different frequencies, receivers and at different locations and found consistent
- However, the harmonics are not at the expected frequencies
- Also, at 96khz, the reconstruction of the DAC output was not satisfactory, so that sampling rate was increased to 192khz

Voice commands

- Encouraged by the initial non linearity from the speaker and the observed sidebands in the receiver side, we moved on to modulating voice commands
- The amplifier gain was increased by adding resistors to increase the transmission range
- However, due to the higher number of samples and the limited memory on the board, we were only able to flash 1-2 second voice signals.
- When the received signal was recorded on the phone and plotted, a few non linear components were observed at baseband ranges
- However, they were not recognized by the Speech Recognition Systems

References

[1] Zhang, Guoming, et al. "Dolphin Attack: Inaudible voice commands." *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017.

[2] http://www.bbc.com/news/technology-41188557

[3] Carlini, Nicholas, et al. "Hidden Voice Commands." *USENIX Security Symposium*. 2016.

[4] Vaidya, Tavish, et al. "Cocaine noodles: exploiting the gap between human and machine speech recognition." *WOOT* 15 (2015): 10-11.

[5] Kasmi, Chaouki, and Jose Lopes Esteves. "IEMI threats for information security: Remote command injection on modern smartphones." *IEEE Transactions on Electromagnetic Compatibility* 57.6 (2015): 1752-1755

[6]https://www.theguardian.com/technology/shortcuts/2017/jan/09/alexa-amazon-echo-goes-rogue-accidental-shopping-dolls-house

References

[7] Yuan, Xuejing, et al. "CommanderSong: A Systematic Approach for Practical Adversarial Voice Recognition." arXiv preprint arXiv:1801.08535 (2018).

[8] Roy, Nirupam, Haitham Hassanieh, and Romit Roy Choudhury. "BackDoor: Sounds that a microphone can record, but that humans can't hear." GetMobile: Mobile Computing and Communications 21.4 (2018): 25-29.