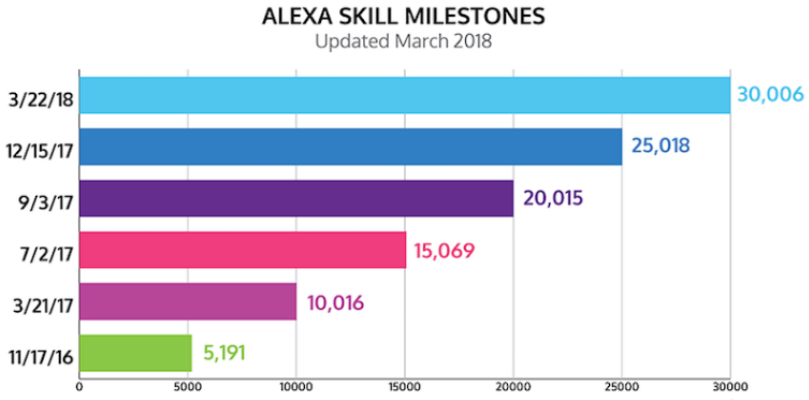
The Insecurity of Home Digital Voice Assistants - Vulnerabilities, Attacks and Countermeasures

Xinyu Lei, Guan-Hua Tu, Alex X. Liu, Chi-Yu Li, Tian Xie

MICHIGAN STATE UNIVERSITY
Computer Science and Engineering

Home Digital Voice Assistant (HDVA) Devices Applications

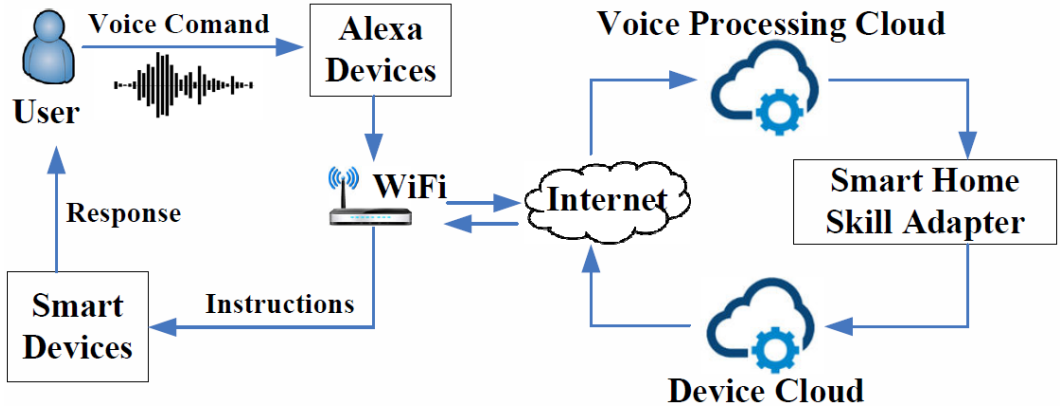


Skills:

- "Alexa, order a laptop from amazon"
- "Alexa, tell smart lock to open my door"
- 50 new skills per day
- Security critical!

Background: Devices and Service Model

Alexa Devices: Echo, Tap, and Echo Dot



Alexa voice service model

Acoustic Attacks

- Fixed location : Alexa devices need to be plugged in to have power supply.
- Weak single-factor authentication: Always listen to “Alexa” to wake up and receive the following voice commands.
- Consequence: More vulnerable to acoustic attacks when owners are absent from home.

V1: Weak Single-factor Authentication

Race	Age	Gender	Result
White African American Asian Hispano	10-30	F	✓
		M	✓
	30-50	F	✓
		M	✓
	50-70	F	✓
		M	✓

- Human Voice: Anyone can pass the authentication by speaking "Alexa" and then command Alexa.

Machines	Speech Speed	Alice	Daisy	George	Jenna	John
Laptop, Desktop	Slow	✓	✓	✓	✓	✓
Mp3 Player, Bluetooth Speaker	Medium	✓	✓	✓	✓	✓
Home Theater System	Fast	✓	✓	✓	✓	✓
Smartphone, Tablet	Very Fast	✓	✓	✓	✓	✓

- Machine Voice: Many machines can produce voice to command Alexa.

(Note voice is generated by online test-to-speech (TTS) system)

V2: No Physical Presence Based Access Control

- No physical presence based access control: Work by detecting the voice even if there is no surrounding users.

- Validation:
  - Alexa can be controlled by the voice command from the Bluetooth speaker 12 meters away.
  - Alexa can accept voice commands larger than 60dB, no matter where the voice comes from.

V3: Insecure Access Control on Alexa-enabled Devices

- Default names: Alexa-enabled devices can be controlled by speaking their names. Re-name a device is usually not mandatory, so the adversary may directly know the commands to control the device.

Devices	Vendors	Default Names
Garage Door	Garageio	My Door
Smart Bulb	Tp-Link	My Smart Bulb
Smart Plug	TP-Link	My Smart Plug
Smart Switch	WeMo	WeMo Light Switch
Learning Thermostat	Nest	Thermostat

Many devices can be controlled by their default voice commands

How To Deliver Voice to Home Digital Voice Assistant Devices

The adversary can deliver voice to Alexa by many methods without being present nearby.

Two example

All devices that are capable of making the sound can be abused and used!!

- Bluetooth s can connect his/her smartphone to the victim's Bluetooth speaker and then play an MP3 audio file of voice commands.
- Smart TVs: the adversary can steal the victim's home Wi-Fi password and then cast a video with voice commands to the victim's smart TV.

Proof-of-concept Attacks

Two real-world attack cases:

- Home Burglary: Tell Alexa to open a door via smart lock. E.g., “Alexa, tell Garageio to open my door”.

- Fake Order: Tell Alexa to place a fake order on Amazon. The owner may suffer from the financial loss.

Possible Solutions

- Learn the authentication users' voice (V1)
  - Voice may change due to health conditions
- Turn off all audio devices while users leave home (V2)
  - Not convenient for users
- Force users to rename the default Alexa-compatible device name (V3)
  - It may not be practical to ask all vendors and users to do so.

Solution: Physical Presence Based Access Control

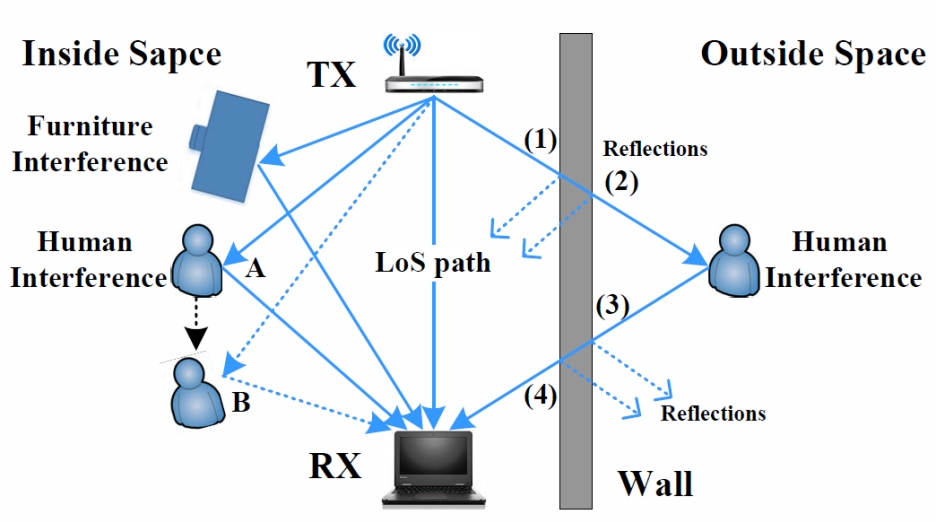
Idea: Virtual Security Button (VSButton):

- Turn on Home Digital Voice Assistant Devices: If physical presence is detected
- Turn off Home Digital Voice Assistant Devices: If no physical presence is detected

Approach: Wi-Fi channel state information (CSI)-based motion detection:

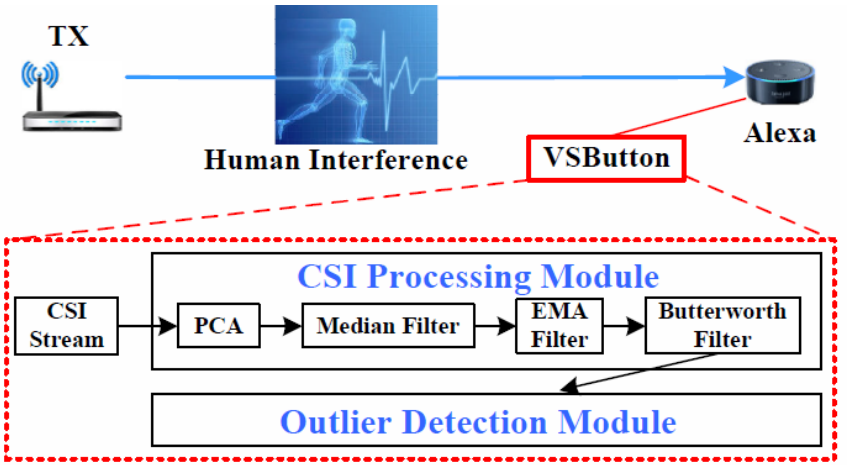
- Significant CSI Variation: Human motions
- Nearly Stable CSI: No motions

Distinguish the Inside and Outside Human Motions



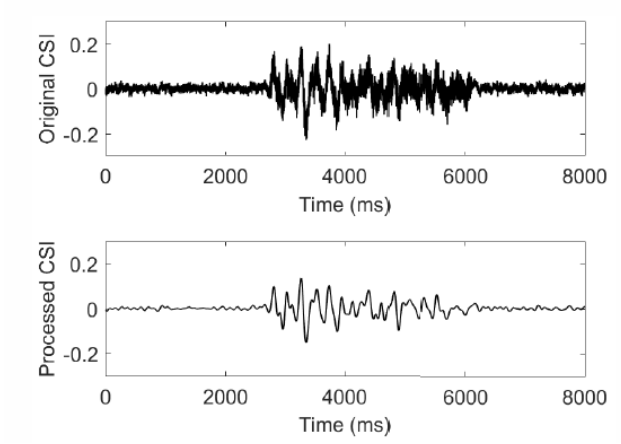
- Outside motions: a small CSI variation
- Inside motions: a significant CSI variation

VSButton Design

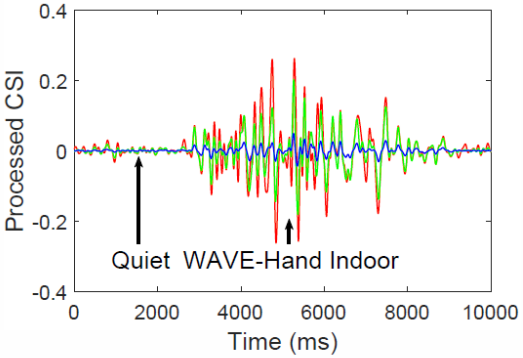


- PCA (Principle Component Analysis) Module : reduce the dimension and remove the noise
- Median and EMA (Exponential Moving Analysis) Filter Module: smooth CSI values
- Butterworth Filter Module: filter out high frequency signal
- Outlier Detection Module : detect human motions.
  - It uses dynamic baseline: adaptive to environment changes.

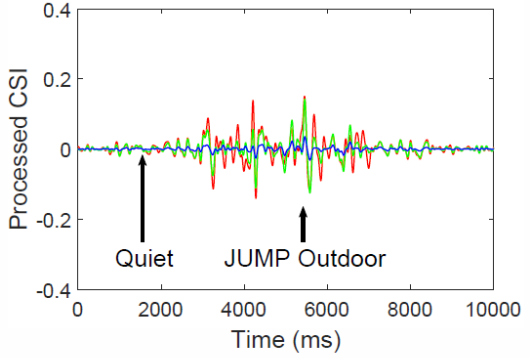
CSI Comparison



Comparison between original/processed CSI



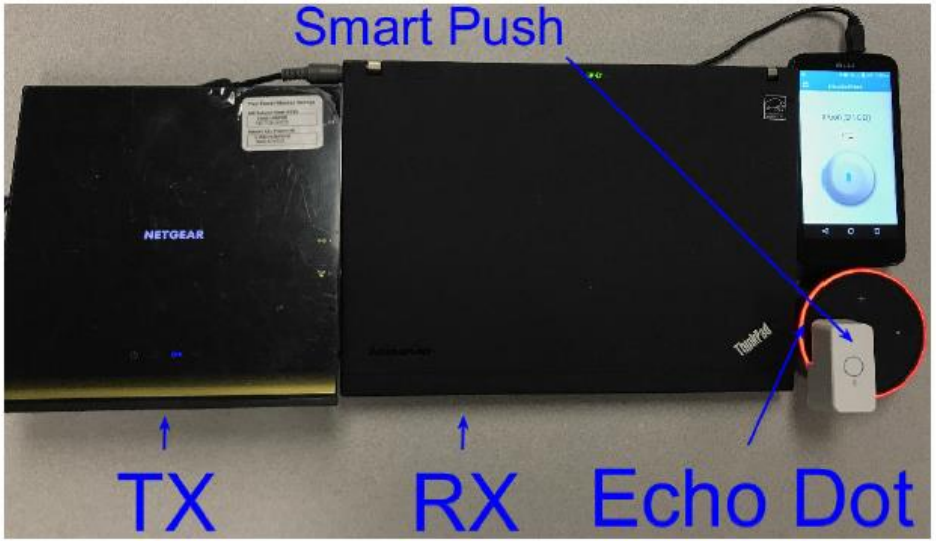
(a) Indoor motion CSI variation



(b) Outdoor motion CSI variation

Comparison between indoor and outdoor CSI variation

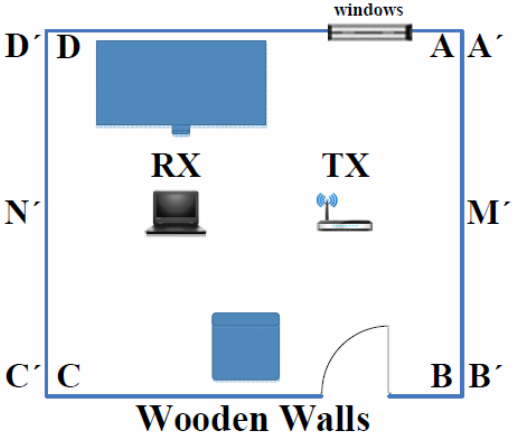
Prototype Implementation



- Transmitter (TX): Home Wi-Fi router
- Receiver (RX): Laptop for processing the CSI
- Controller: Smart push to turn on/off Alexa

- Software updates: RX and controller can be integrated to the Alexa devices by only software upgrades.

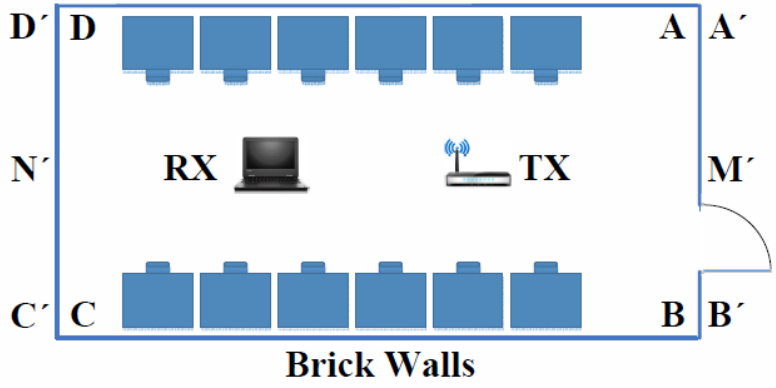
Evaluation – Square room with wooden walls



- Configuration: Square room with wooden walls.
- CSI Variation: The strongest CSI variation of the movements outside is smaller than the weakest one among the movements inside.

Square Room	Indoor locations				Outdoor locations					
locations	A	B	C	D	A'	B'	C'	D'	M'	N'
WAVE-HAND	**0.312**	0.315	0.401	0.409	0.041	0.043	0.049	0.051	0.092	0.063
SIT-DOWN-STAND-UP	0.345	0.349	0.423	0.430	0.060	0.062	0.069	0.071	0.121	0.089
JUMP	0.401	0.407	0.451	0.459	0.069	0.071	0.084	0.086	**0.241**	0.099
DO NOTHING	0.025	0.021	0.022	0.024	0.028	0.026	0.021	0.022	0.023	0.025

Evaluation – Rectangle room with brick walls



- Configuration: Rectangle room with brick walls
- CSI Variation: The strongest CSI variation of the movements outside is smaller than the weakest one among the movements inside.

Rectangle Room	Indoor locations				Outdoor locations					
locations	A	B	C	D	A'	B'	C'	D'	M'	N'
WAVE-HAND	**0.147**	0.150	0.180	0.183	0.020	0.022	0.025	0.027	0.035	0.030
SIT-DOWN-STAND-UP	0.181	0.184	0.216	0.217	0.024	0.026	0.028	0.029	0.039	0.033
JUMP	0.254	0.255	0.287	0.288	0.029	0.029	0.032	0.033	**0.042**	0.035
DO NOTHING	0.022	0.021	0.022	0.027	0.028	0.026	0.021	0.022	0.020	0.025

Current Limitations of VSButton

- Motions not from human: VSButton may activate HDVA service due to motions from pets.
- Tradeoff between security and convenience: By increasing the motion detection threshold *t*, VSButton has a smaller false positive rate.
- Physical invasive attack: VSButton cannot resist invasive attack (e.g, the adversary breaks the windows and enters the victim's room). If invasive attack occurs, the adversary is able to do many things much eviler than attacking the HDVA.

Conclusion

- HDVAs: Amazon Alexa and Google Home
- Identify three vulnerabilities
- Solution: VSButton based on physical presence
  - Accurate detection in both laboratory and real-world home setting

Questions & Answers