Light Commands: Laser-Based Audio Injection Attacks on Voice-Controllable Systems

Takeshi Sugawara, The University of Electro-Communications
Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu, University of Michigan
Usenix, 2020

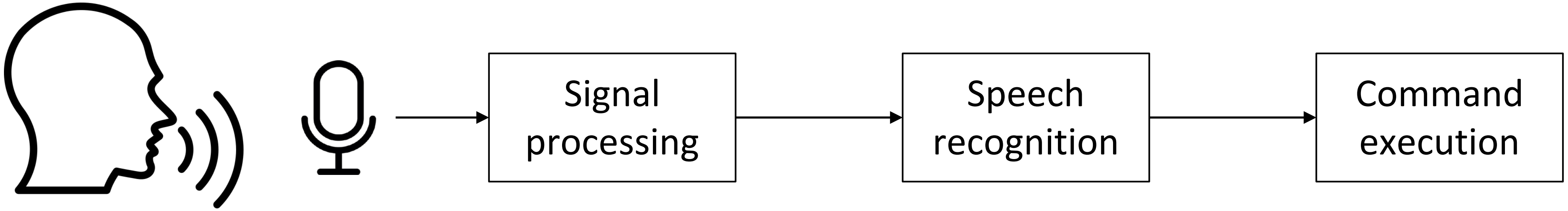
Presenter: Junho Ahn

PPT from author

Voice Controllable Systems (VCSs)

[Source: pandaily.com]

[Source: developers.google.com]



Signal processing

Speech recognition

Command execution

Security Concerns

- The sacrifice of security to improve availability
- Interfacing with 3rd Party Software
- **Blind trust** in the microphone reading

‘100…’ ‘Incorrect…’

‘101…’ ‘Incorrect…’

‘102…’ ‘Incorrect…’

… …

‘777…’ ‘OK…’

Vulnerability

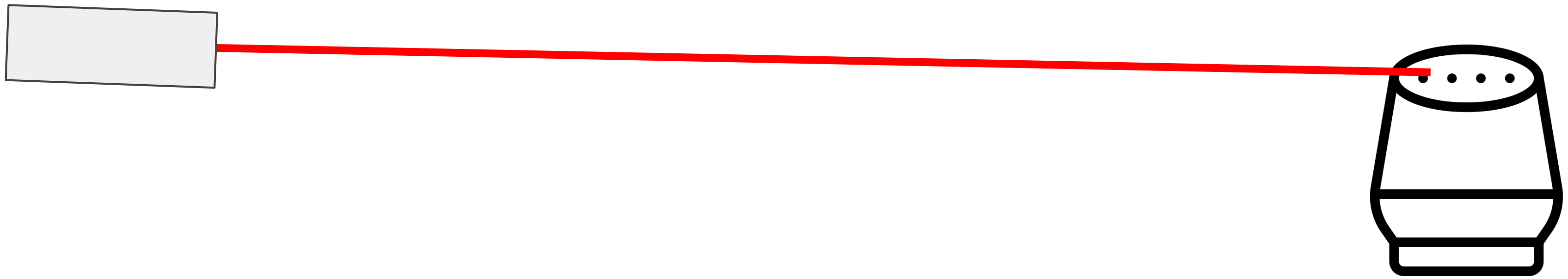
Assumption:

Microphones capture the **acoustic** signal

Vulnerability

Reality:

Microphones capture the acoustic sound and **light signal**



Vulnerability

Questions:

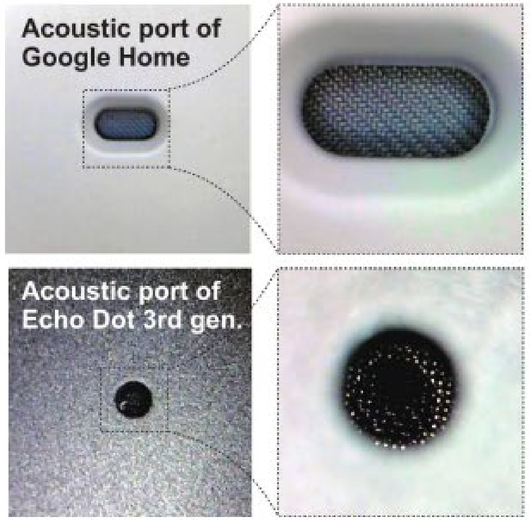
1. How does laser injection affect VCSs?
2. How can we protect VCSs against LASER injection?

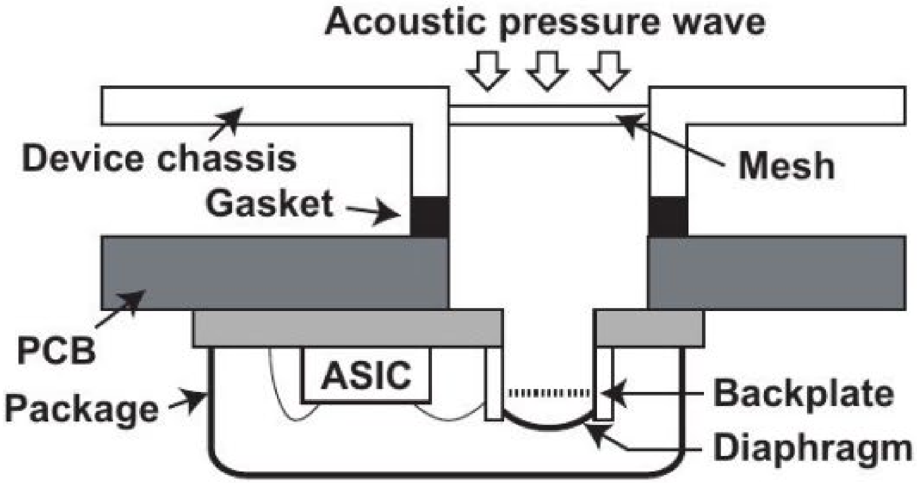
Introduction

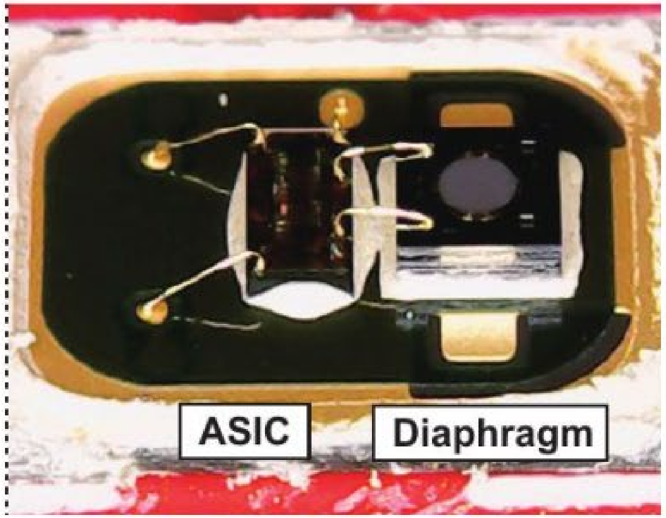
- MEMS microphone basic
- VCS command injection via light procedure
- Evaluation
- Countermeasures

MEMS Microphones

- Used in most VCSs
- The diaphragm and backplate work as a capacitor
- When diaphragm moves, it causes a change in capacitance
- The ASIC converts the capacitive change to voltage

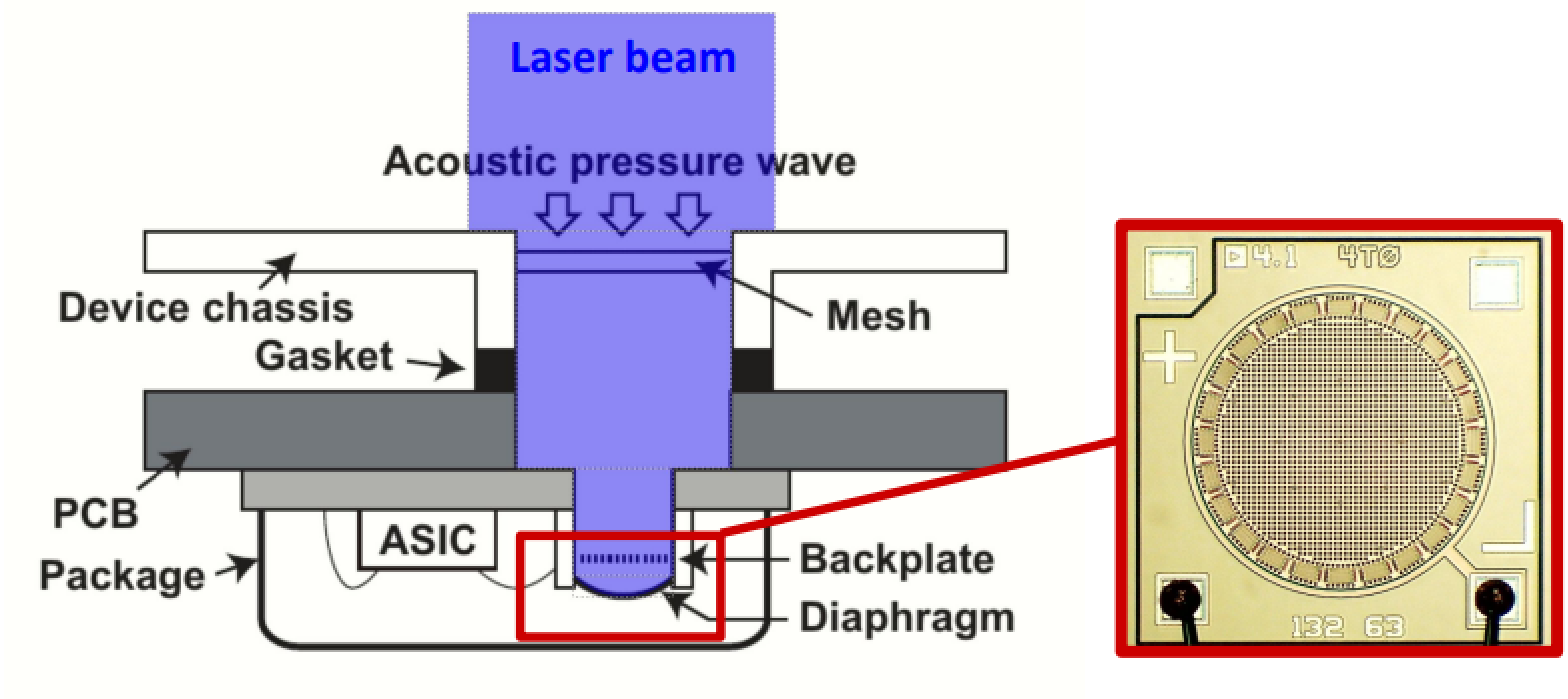

Acoustic port of Google Home
Acoustic port of Echo Dot 3rd gen.





MEMS Microphones

- MEMS microphones exhibit light sensitivity
- Output voltage affected by light **irradiance**
- Inject signal by modulating optical power

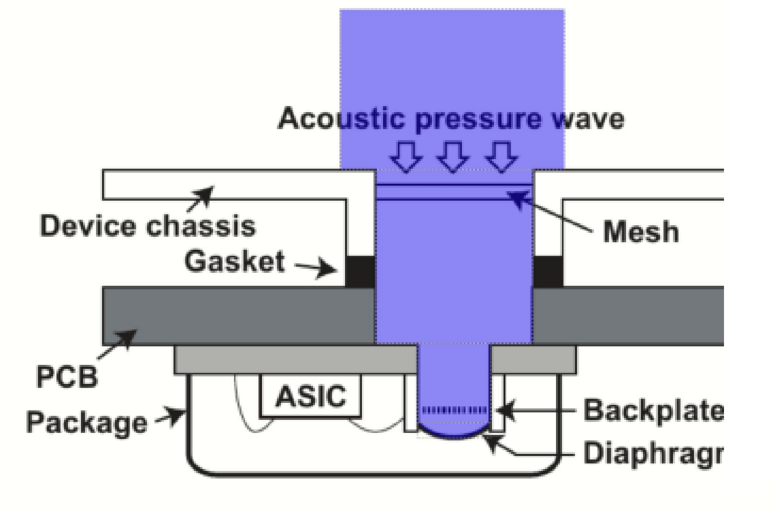


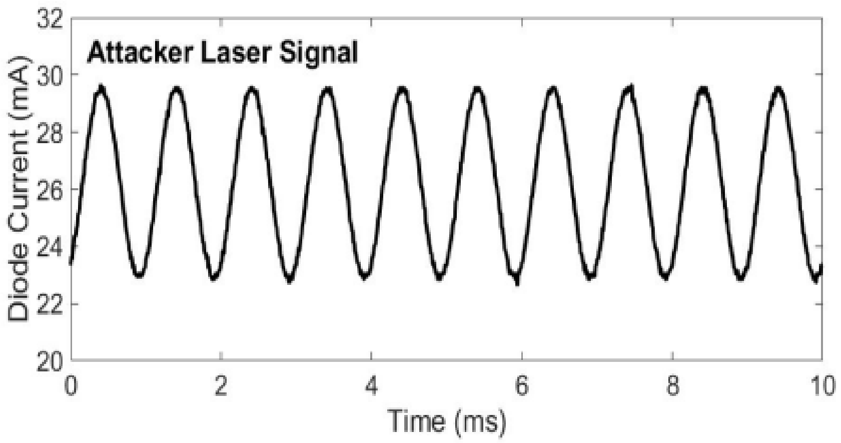
Irradiance:

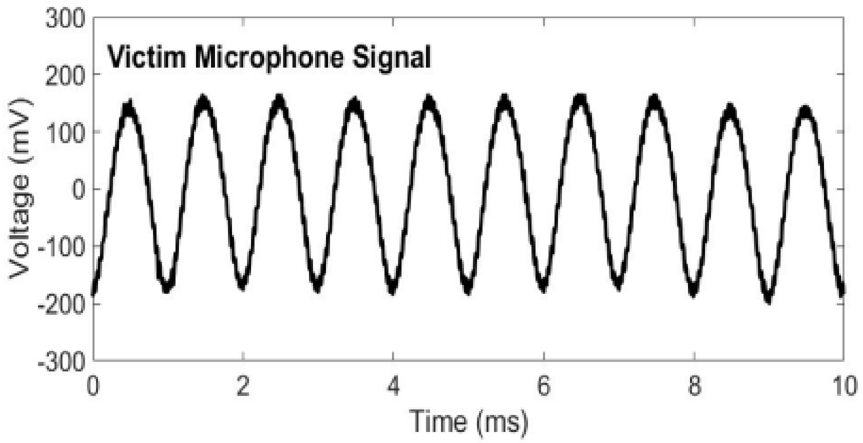
$$I = \frac{Optical\,Power(W))}{Beam\,Area(m^2)}$$

Key Ideas

1. Amplitude modulated light generates a voltage signal on microphone output
2. Higher amplitude light makes higher amplitude voltage
3. Very little distortion





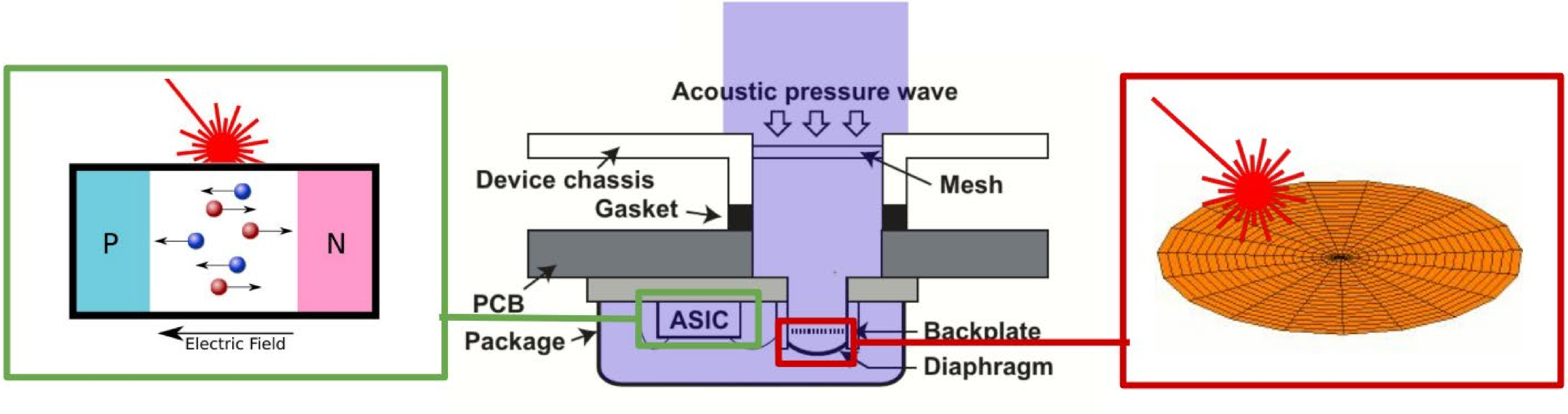


How is this Working?

Combination of two physical effects:

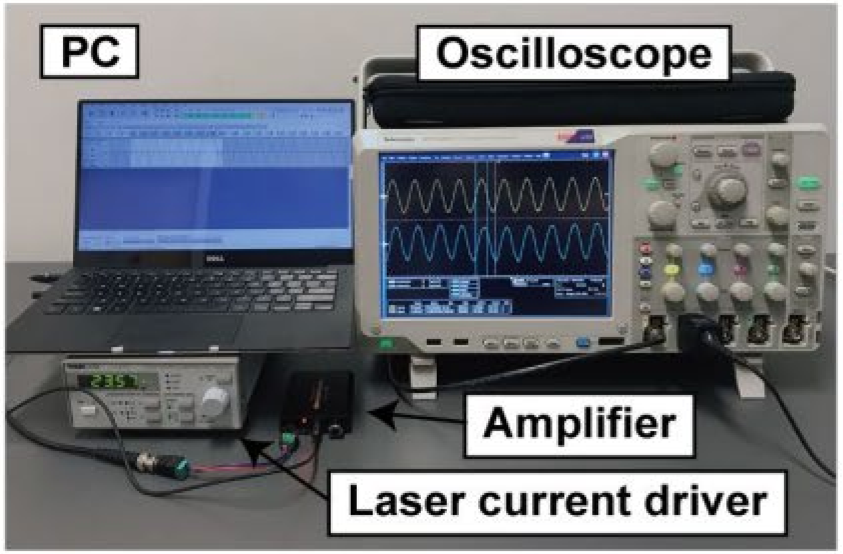
1. **Photoelectric** Effects

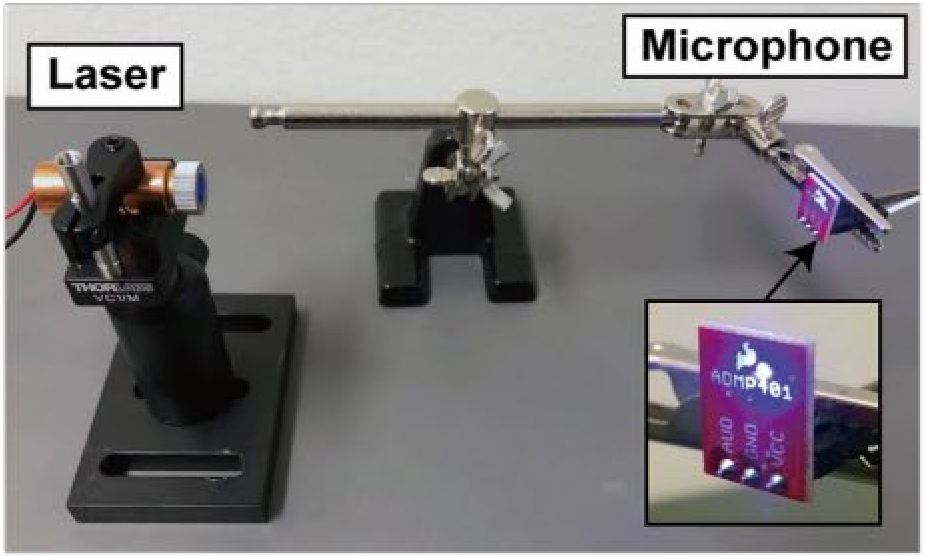
2. **Photoacoustic** Effects

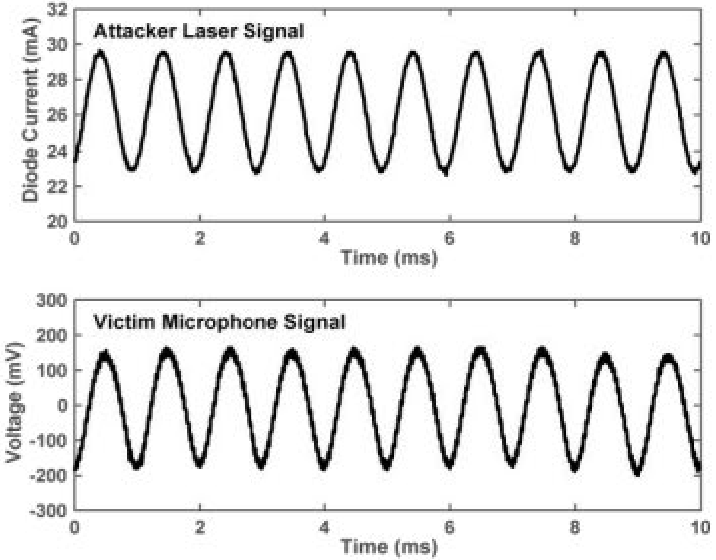


Signal Injection via Laser

- Audio voltage signal from laptop
- Laser current driver converts to current signal with DC bias
- Laser output power is proportional to current



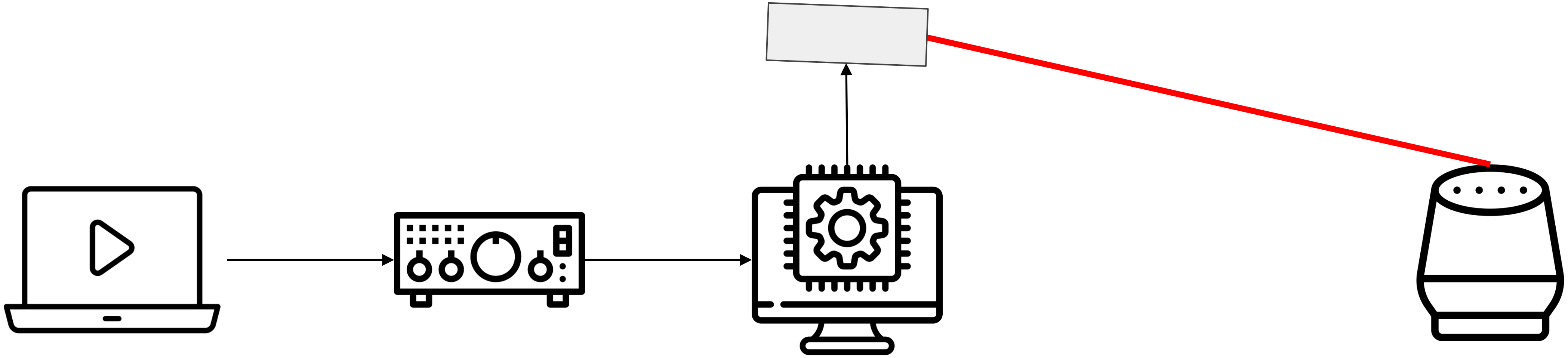




VCS Command Injection via Light

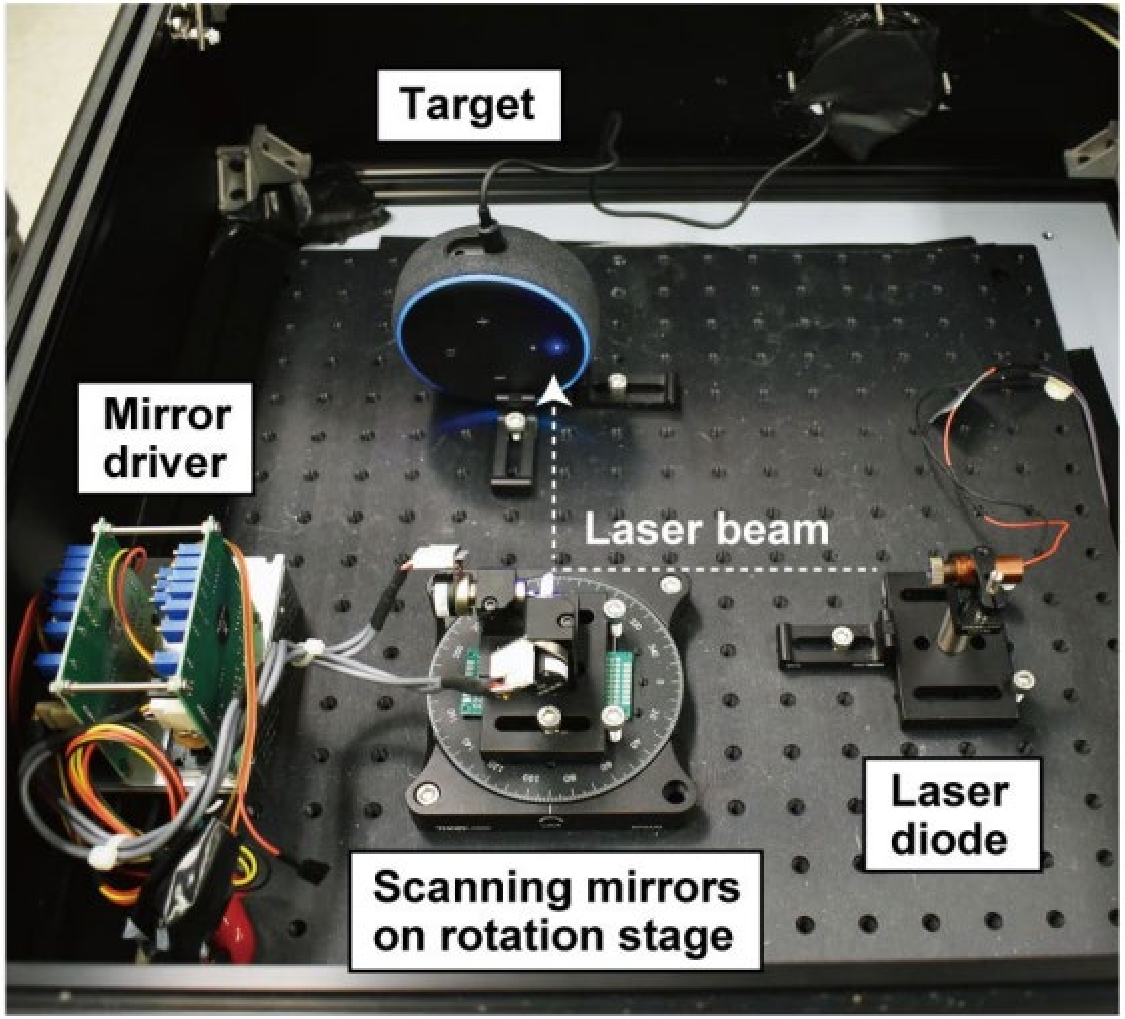
Digital Signal → Voltage Signal → Current Signal → Light Signal

‘OK Google, Open the garage door’



Evaluation - Power

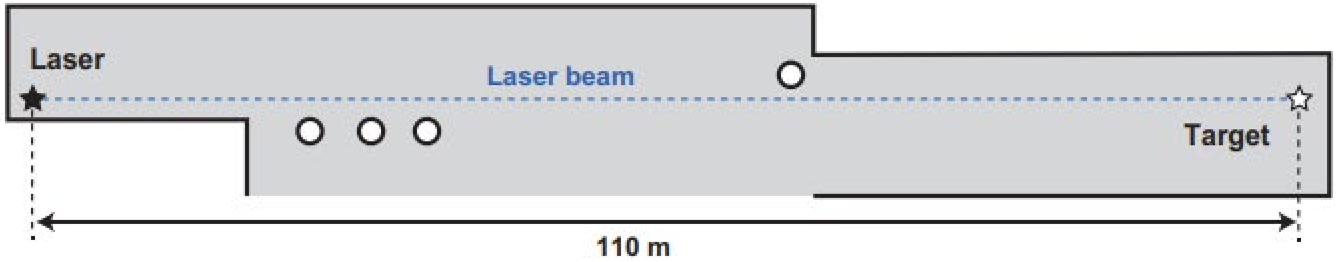
- Investigated 17 devices
- Used scanning mirrors
- Measured minimum optical power to recognize commands

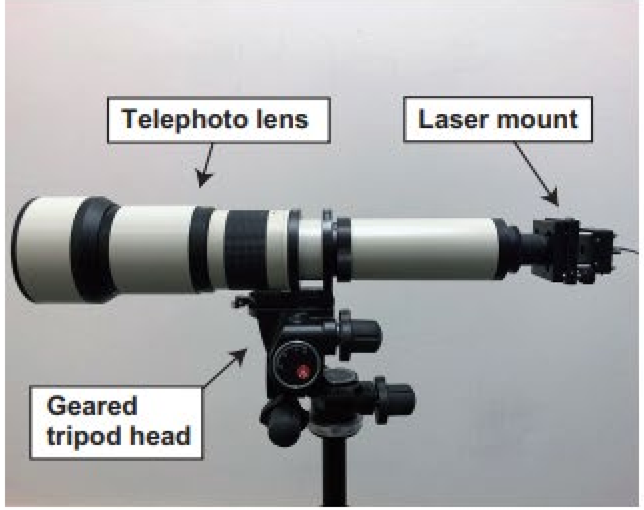


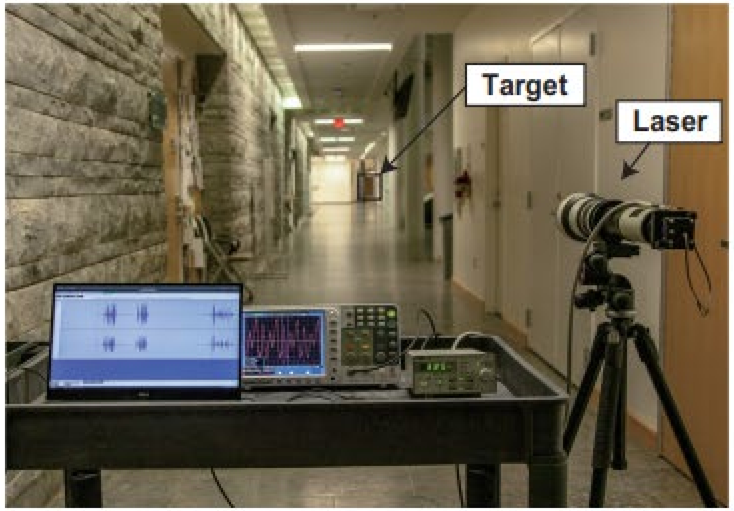
Evaluation - Range

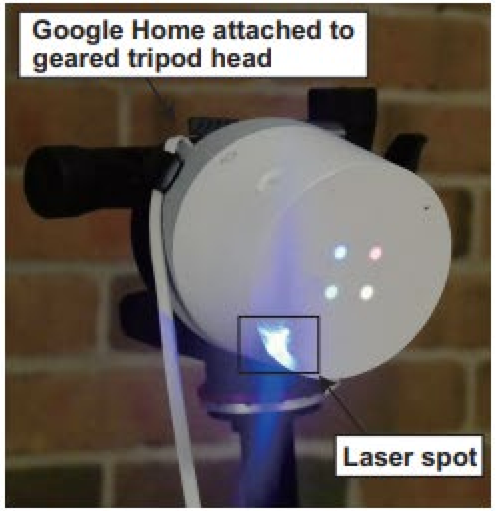
Measuring the maximum range of the attack

$$I = \frac{Optical Power(W))}{Beam Area(m^2)}$$









Attack Result

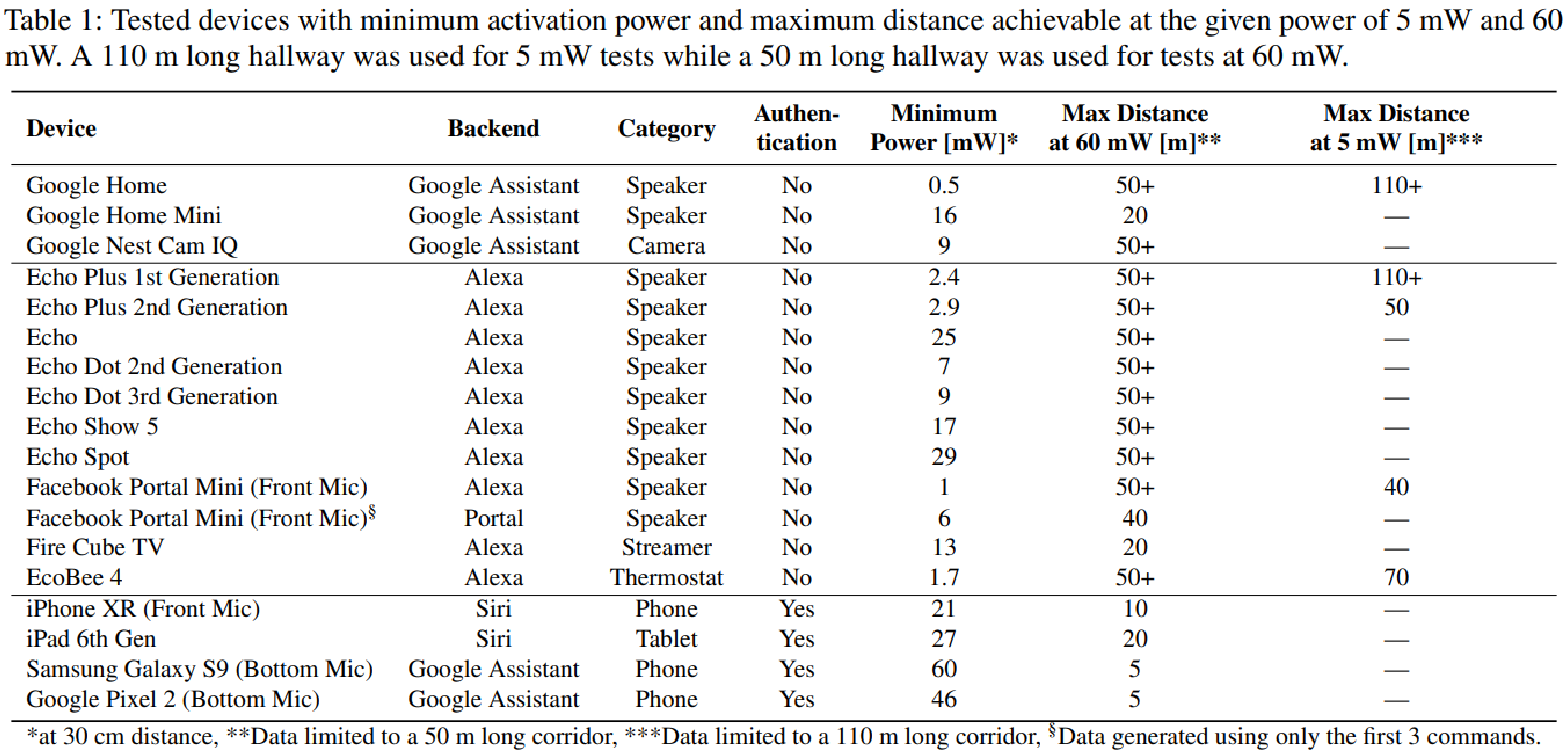


Table 1: Tested devices with minimum activation power and maximum distance achievable at the given power of 5 mW and 60 mW. A 110 m long hallway was used for 5 mW tests while a 50 m long hallway was used for tests at 60 mW.

Device	Backend	Category	Authentication	Minimum Power [mW]*	Max Distance at 60 mW [m]**	Max Distance at 5 mW [m]***
Google Home	Google Assistant	Speaker	No	0.5	50+	110+
Google Home Mini	Google Assistant	Speaker	No	16	20	—
Google Nest Cam IQ	Google Assistant	Camera	No	9	50+	—
Echo Plus 1st Generation	Alexa	Speaker	No	2.4	50+	110+
Echo Plus 2nd Generation	Alexa	Speaker	No	2.9	50+	50
Echo	Alexa	Speaker	No	25	50+	—
Echo Dot 2nd Generation	Alexa	Speaker	No	7	50+	—
Echo Dot 3rd Generation	Alexa	Speaker	No	9	50+	—
Echo Show 5	Alexa	Speaker	No	17	50+	—
Echo Spot	Alexa	Speaker	No	29	50+	—
Facebook Portal Mini (Front Mic)	Alexa	Speaker	No	1	50+	40
Facebook Portal Mini (Front Mic)§	Portal	Speaker	No	6	40	—
Fire Cube TV	Alexa	Streamer	No	13	20	—
EcoBee 4	Alexa	Thermostat	No	1.7	50+	70
iPhone XR (Front Mic)	Siri	Phone	Yes	21	10	—
iPad 6th Gen	Siri	Tablet	Yes	27	20	—
Samsung Galaxy S9 (Bottom Mic)	Google Assistant	Phone	Yes	60	5	—
Google Pixel 2 (Bottom Mic)	Google Assistant	Phone	Yes	46	5	—

*at 30 cm distance, **Data limited to a 50 m long corridor, ***Data limited to a 110 m long corridor, §Data generated using only the first 3 commands.

Attack Result

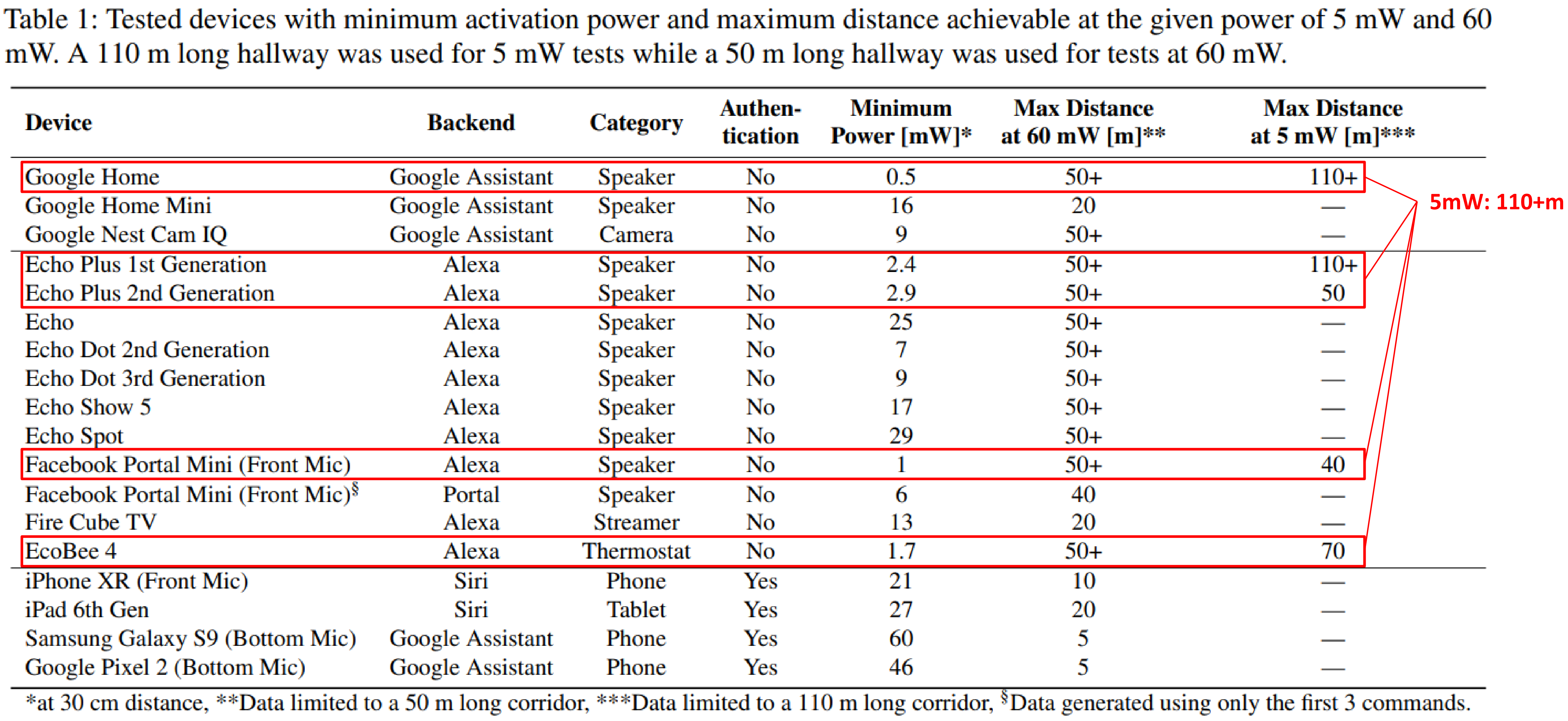


Table 1: Tested devices with minimum activation power and maximum distance achievable at the given power of 5 mW and 60 mW. A 110 m long hallway was used for 5 mW tests while a 50 m long hallway was used for tests at 60 mW.

Device	Backend	Category	Authentication	Minimum Power [mW]*	Max Distance at 60 mW [m]**	Max Distance at 5 mW [m]***
Google Home	Google Assistant	Speaker	No	0.5	50+	110+
Google Home Mini	Google Assistant	Speaker	No	16	20	—
Google Nest Cam IQ	Google Assistant	Camera	No	9	50+	—
Echo Plus 1st Generation	Alexa	Speaker	No	2.4	50+	110+
Echo Plus 2nd Generation	Alexa	Speaker	No	2.9	50+	50
Echo	Alexa	Speaker	No	25	50+	—
Echo Dot 2nd Generation	Alexa	Speaker	No	7	50+	—
Echo Dot 3rd Generation	Alexa	Speaker	No	9	50+	—
Echo Show 5	Alexa	Speaker	No	17	50+	—
Echo Spot	Alexa	Speaker	No	29	50+	—
Facebook Portal Mini (Front Mic)	Alexa	Speaker	No	1	50+	40
Facebook Portal Mini (Front Mic)§	Portal	Speaker	No	6	40	—
Fire Cube TV	Alexa	Streamer	No	13	20	—
EcoBee 4	Alexa	Thermostat	No	1.7	50+	70
iPhone XR (Front Mic)	Siri	Phone	Yes	21	10	—
iPad 6th Gen	Siri	Tablet	Yes	27	20	—
Samsung Galaxy S9 (Bottom Mic)	Google Assistant	Phone	Yes	60	5	—
Google Pixel 2 (Bottom Mic)	Google Assistant	Phone	Yes	46	5	—

*at 30 cm distance, **Data limited to a 50 m long corridor, ***Data limited to a 110 m long corridor, §Data generated using only the first 3 commands.

5mW: 110+m

Attack Result

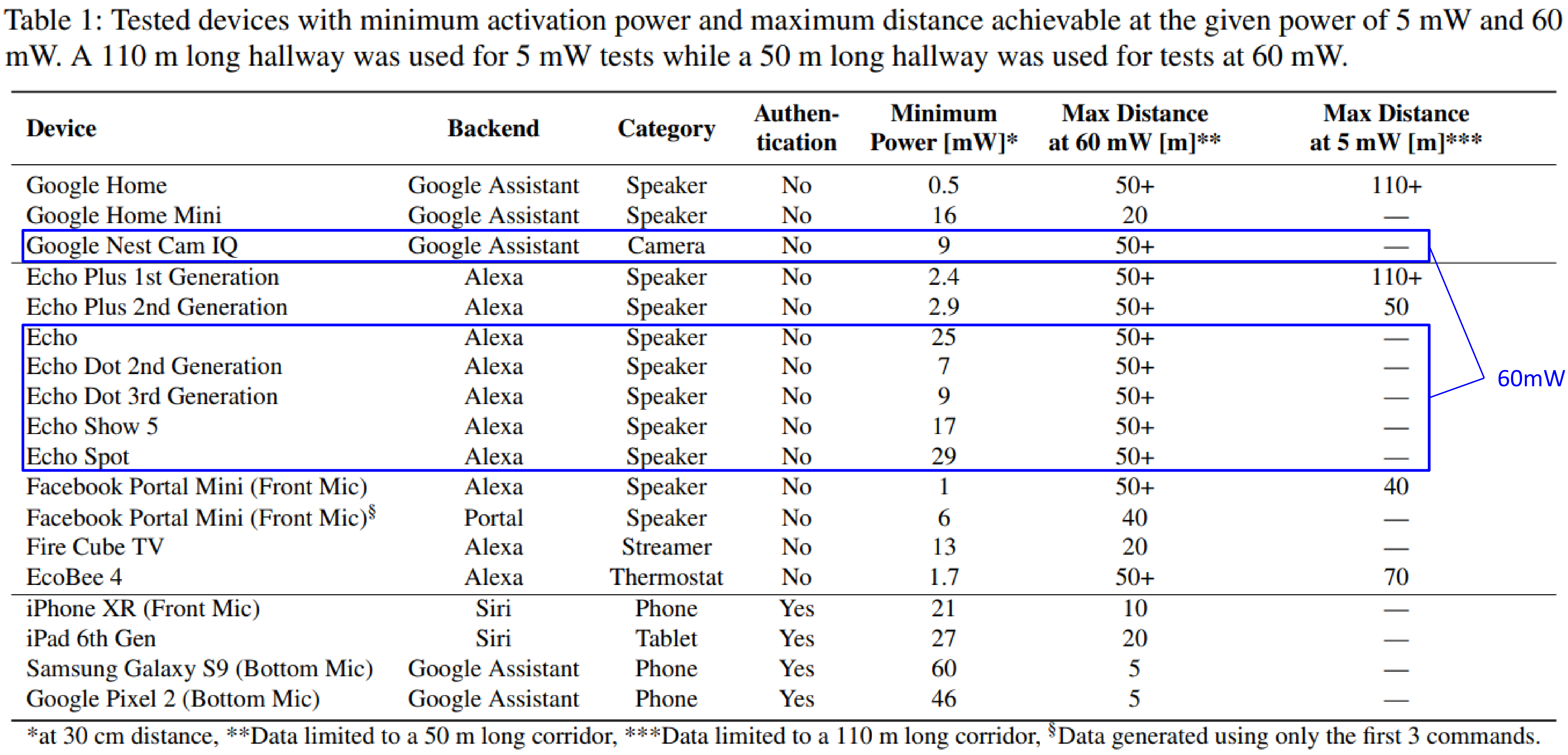


Table 1: Tested devices with minimum activation power and maximum distance achievable at the given power of 5 mW and 60 mW. A 110 m long hallway was used for 5 mW tests while a 50 m long hallway was used for tests at 60 mW.

Device	Backend	Category	Authentication	Minimum Power [mW]*	Max Distance at 60 mW [m]**	Max Distance at 5 mW [m]***
Google Home	Google Assistant	Speaker	No	0.5	50+	110+
Google Home Mini	Google Assistant	Speaker	No	16	20	—
Google Nest Cam IQ	Google Assistant	Camera	No	9	50+	—
Echo Plus 1st Generation	Alexa	Speaker	No	2.4	50+	110+
Echo Plus 2nd Generation	Alexa	Speaker	No	2.9	50+	50
Echo	Alexa	Speaker	No	25	50+	—
Echo Dot 2nd Generation	Alexa	Speaker	No	7	50+	—
Echo Dot 3rd Generation	Alexa	Speaker	No	9	50+	—
Echo Show 5	Alexa	Speaker	No	17	50+	—
Echo Spot	Alexa	Speaker	No	29	50+	—
Facebook Portal Mini (Front Mic)	Alexa	Speaker	No	1	50+	40
Facebook Portal Mini (Front Mic)§	Portal	Speaker	No	6	40	—
Fire Cube TV	Alexa	Streamer	No	13	20	—
EcoBee 4	Alexa	Thermostat	No	1.7	50+	70
iPhone XR (Front Mic)	Siri	Phone	Yes	21	10	—
iPad 6th Gen	Siri	Tablet	Yes	27	20	—
Samsung Galaxy S9 (Bottom Mic)	Google Assistant	Phone	Yes	60	5	—
Google Pixel 2 (Bottom Mic)	Google Assistant	Phone	Yes	46	5	—

*at 30 cm distance, **Data limited to a 50 m long corridor, ***Data limited to a 110 m long corridor, §Data generated using only the first 3 commands.

60mW: 50+m

Attack Result

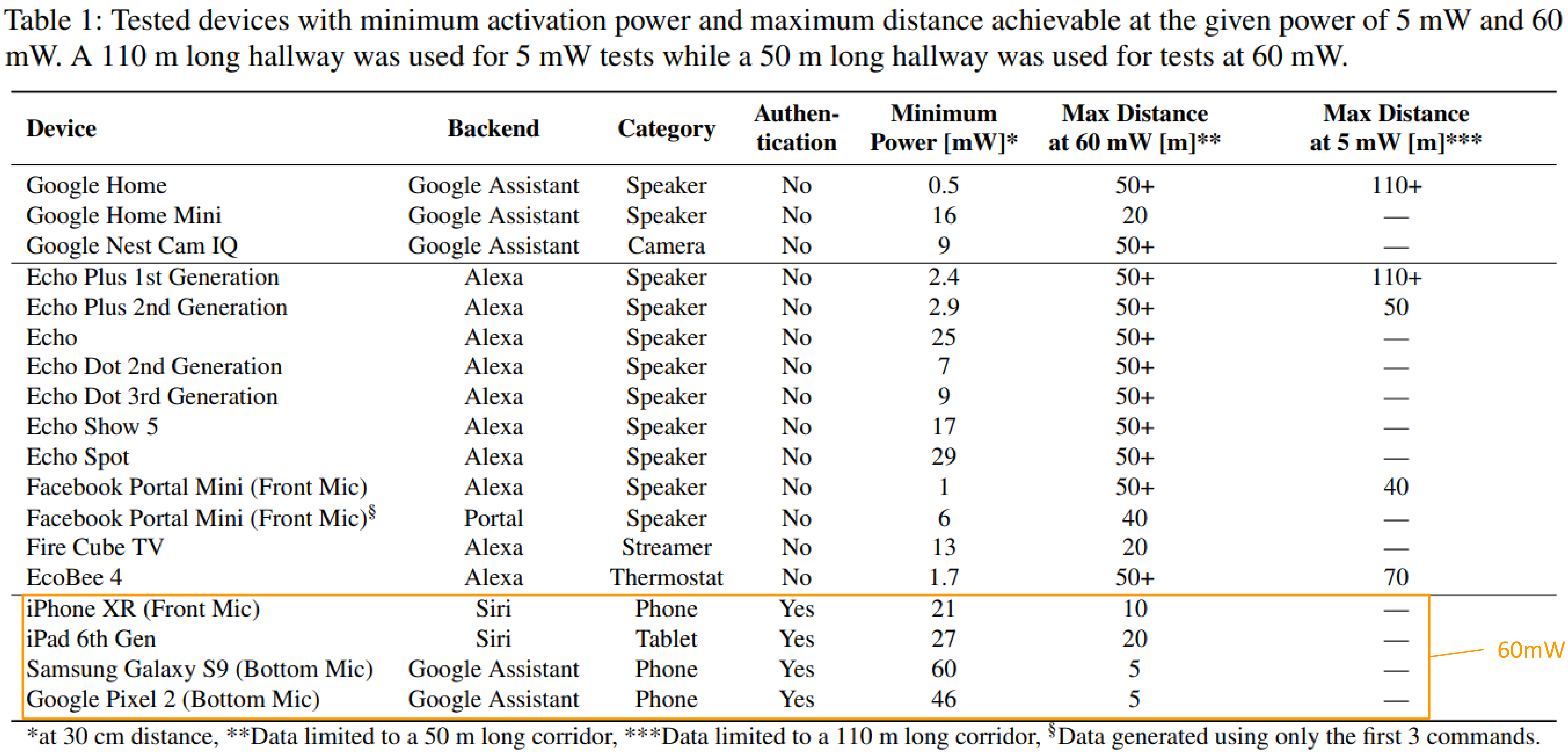


Table 1: Tested devices with minimum activation power and maximum distance achievable at the given power of 5 mW and 60 mW. A 110 m long hallway was used for 5 mW tests while a 50 m long hallway was used for tests at 60 mW.

Device	Backend	Category	Authen-tication	Minimum Power [mW]*	Max Distance at 60 mW [m]**	Max Distance at 5 mW [m]***
Google Home	Google Assistant	Speaker	No	0.5	50+	110+
Google Home Mini	Google Assistant	Speaker	No	16	20	—
Google Nest Cam IQ	Google Assistant	Camera	No	9	50+	—
Echo Plus 1st Generation	Alexa	Speaker	No	2.4	50+	110+
Echo Plus 2nd Generation	Alexa	Speaker	No	2.9	50+	50
Echo	Alexa	Speaker	No	25	50+	—
Echo Dot 2nd Generation	Alexa	Speaker	No	7	50+	—
Echo Dot 3rd Generation	Alexa	Speaker	No	9	50+	—
Echo Show 5	Alexa	Speaker	No	17	50+	—
Echo Spot	Alexa	Speaker	No	29	50+	—
Facebook Portal Mini (Front Mic)	Alexa	Speaker	No	1	50+	40
Facebook Portal Mini (Front Mic)[§]	Portal	Speaker	No	6	40	—
Fire Cube TV	Alexa	Streamer	No	13	20	—
EcoBee 4	Alexa	Thermostat	No	1.7	50+	70
iPhone XR (Front Mic)	Siri	Phone	Yes	21	10	—
iPad 6th Gen	Siri	Tablet	Yes	27	20	—
Samsung Galaxy S9 (Bottom Mic)	Google Assistant	Phone	Yes	60	5	—
Google Pixel 2 (Bottom Mic)	Google Assistant	Phone	Yes	46	5	—

*at 30 cm distance, **Data limited to a 50 m long corridor, ***Data limited to a 110 m long corridor, [§]Data generated using only the first 3 commands.

60mW: 5-20m

Cross-Building Attack Scenario

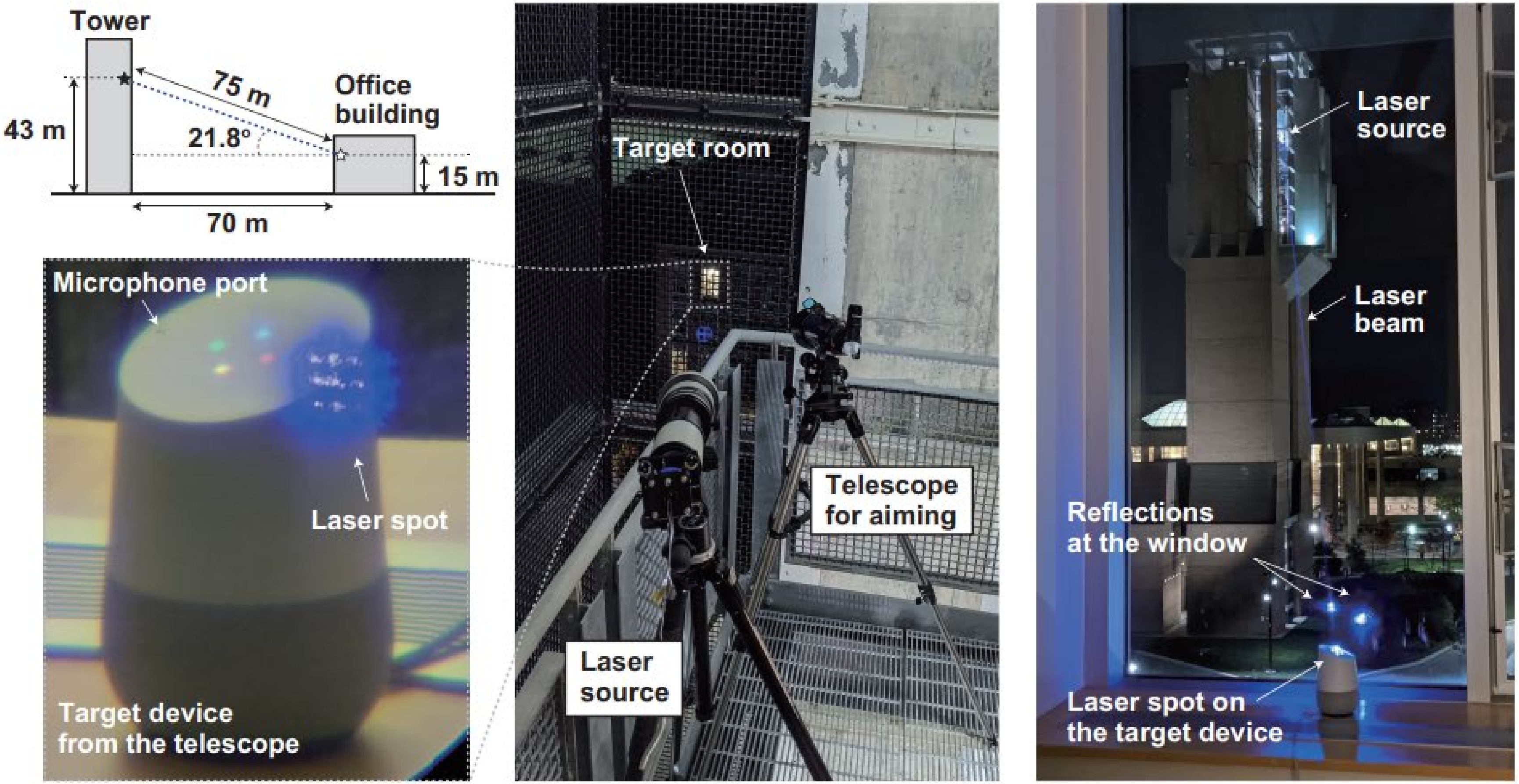


Figure 10: Setup for the low-power cross-building attack: (Top left) Laser and target arrangement. (Bottom left) Picture of the target device as visible through the telescope, with the microphone ports and laser spot clearly visible. (Middle) Picture from the tower: laser on telephoto lens aiming down to the target. (Right) Picture from the office building: laser spot on the target device.

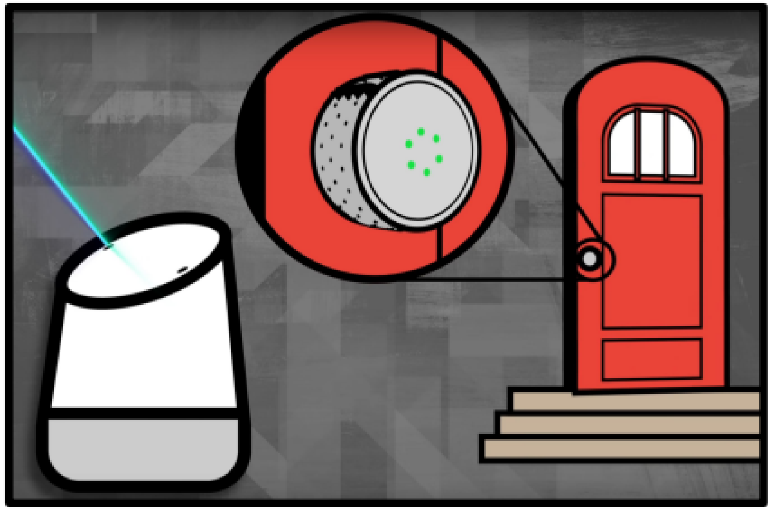
Attack Demonstration

Consequences

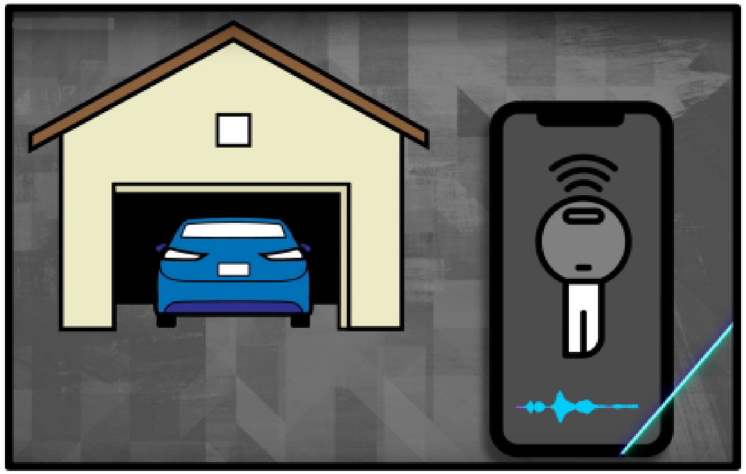
Hi, how can i help?
Ok, Google, turn on light
Light on

Turn on/off
Enable/Disable

Brute force unlock door



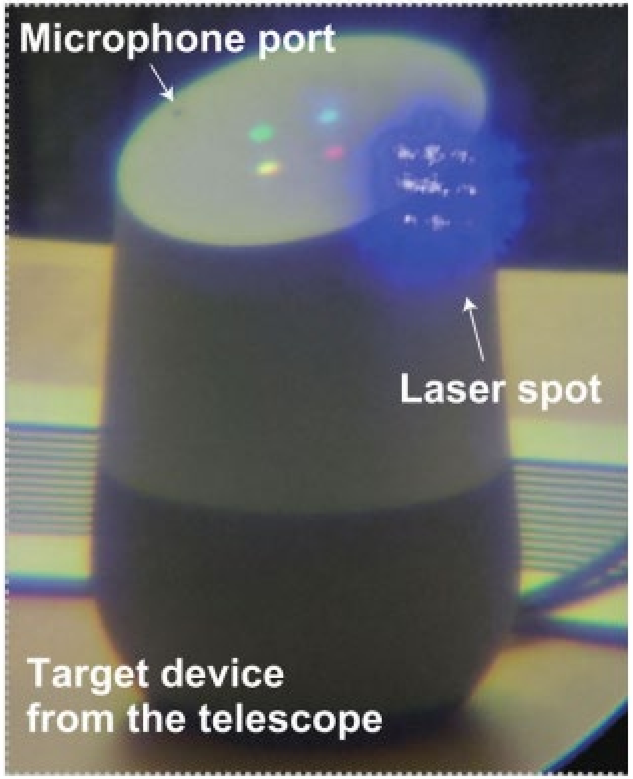
Unauthorized purchases



Open garage door
Unlock car
Start engine

Limitations

- Dependence on Focusing, Aiming, Acoustic Noise, and Audio Quality
- Requires Line of Sight
  - Very little diffraction
  - Difficult to target top microphones
- Limited Feedback



Countermeasures

Software Approaches

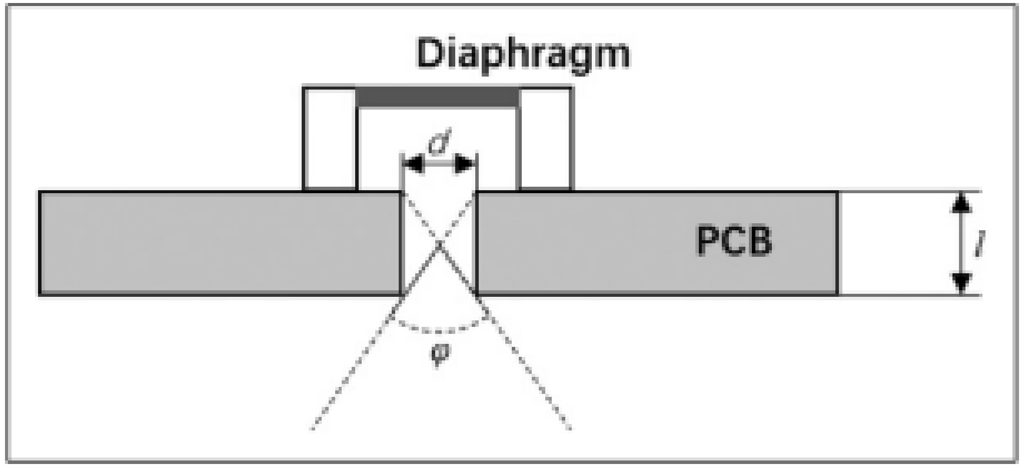
- Stronger authentication
- Liveness tests
- Sensors fusion: compare multiple microphones

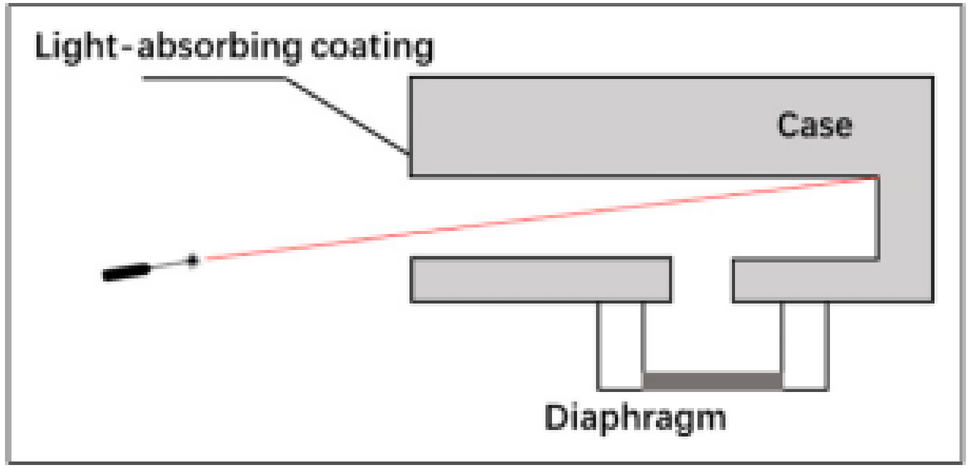
Hardware Approaches

- Light-blocking covers
  - On the VCS(fabric)
  - Inside the MEMS microphone

Future Work

- Evaluation and defense of light commands attacks against voice controllable systems in smart cars
  - Zhijian Xu, Guoming Zhang, Xiaoyu Ji and Wenyuan Xu





Related Work

- Attacks on VCS Speech Recognition
  - Vaidya et al., “Cocaine noodles: exploiting the gap between human and machine speech recognition,” USENIX WOOT, 2015
  - Carlini et al., “Hidden voice commands.” in USENIX 2016
  - Yuan et al., “CommanderSong: A systematic approach for practical adversarial voice recognition,” in USENIX 2018

- Acoustic Injection on VCS via Ultrasound
  - Roy et al., “Backdoor: Making microphones hear inaudible sounds,” in ACM MobiSys 2017.
  - Zhang et al., “DolphinAttack: Inaudible voice commands,” in ACM CCS 2017.
  - Roy et al., “Inaudible voice commands: The long-range attack and defense,” in USENIX NSDI 2018.

Conclusion

- Lasers can inject commands into VCSs
- Long range with low optical power
- Physical vulnerability in MEMS microphones
- Highlights security flaws in VCSs
- Blind trust of any input often points to vulnerabilities

Questions

Q. (오범석) how can we define overall criteria that sensors should satisfy to avoid sensor attacks?

- Block possible side channels
- There's nothing we can do about attacks that can't defend.

Questions

Q. (윤정한) For a countermeasure, what about using another sensor that only can sense the light, not sound?

- Blocking light is better

Questions

Q. (김한나) Is there any related work about laser heating?

- Yes
- e.g. fire alarm