Auditable & **P**rovable **P**rivacy of Smart Speakers

Outline

- Background & Motivation of Privacy Violations
  - SEC filings
  - Patents
  - Paranoid Users
  - Consumer confidence matters
- Case Study of 4 Risks
  - New risk surface ‘Hot Mic’ revealed in hardware bus
- Methodology of Auditable & Provable Privacy Designs
  - Auditable: authority could use the help from manufacturer
  - Provable: better choice in design phase
  - Learn from those good designs

Background

- Smart devices are everywhere in our lives.
- Data become the crucial part in the competition for the vendors.
- The technology regarding massive voice/face/video processing is mature enough.

Motivation of Privacy Violations: SEC Filings

- *The goal of our advertising business is to deliver relevant ads at just the right time and to give people useful commercial information, regardless of the device they're using. -- Alphabet*
  - *We generated **88%** of total revenues from advertising in 2016.*
  - *We generated over **86%** of total revenues from advertising in 2017.*
  - We generate revenues primarily by delivering both **performance advertising** and **brand advertising**. Performance advertising creates and delivers relevant ads that users will click on, leading to direct engagement with advertisers. [...]
- ***We generate substantially all of our revenue from selling advertising placements to marketers.** Our ads enable marketers to reach people based on a variety of factors including **age, gender, location, interests, and behaviors.** Marketers purchase ads that can appear in multiple places including on Facebook, Instagram, Messenger, and third-party applications and websites. - Facebook*

#1

Correlating media consumption data with user profiles

US20170195435A1

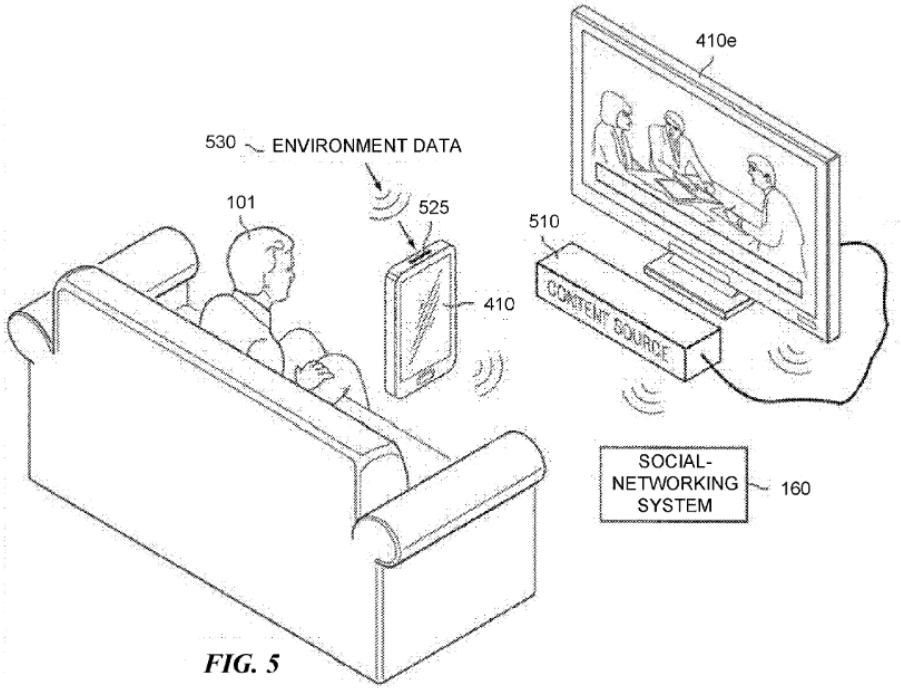


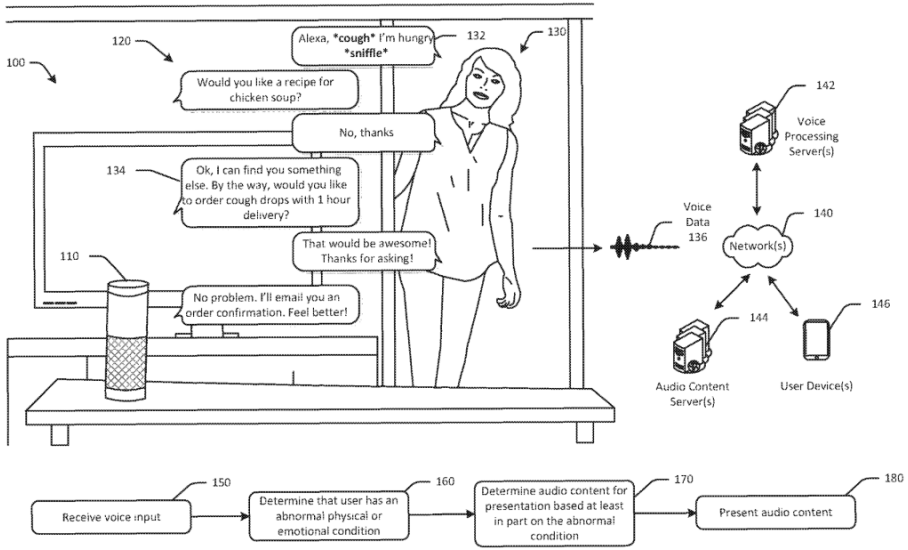
FIG. 5

[Ref: NYT]

#2

Voice-based determination of physical and emotional characteristics of users

US10096319B1

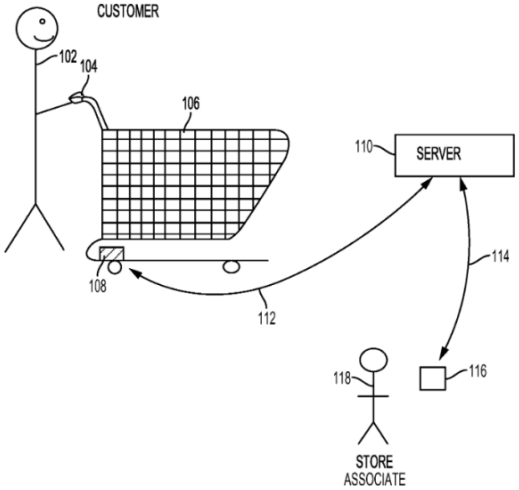


[Ref: slashdot]

#3

System and method for a biometric feedback cart handle

US20180240554A1



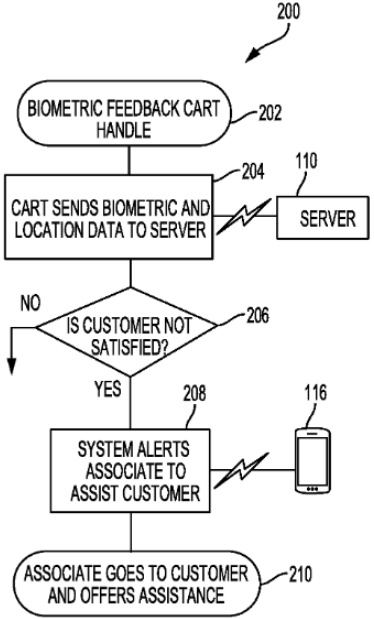


FIG. 2

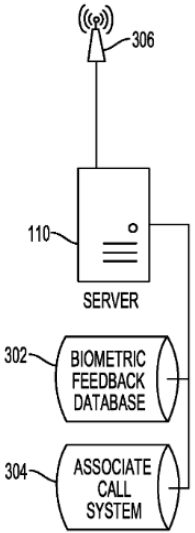


FIG. 3

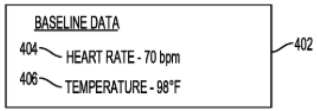


FIG. 4A

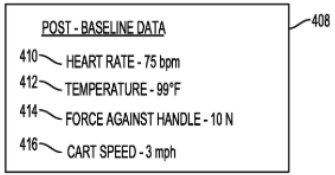


FIG. 4B

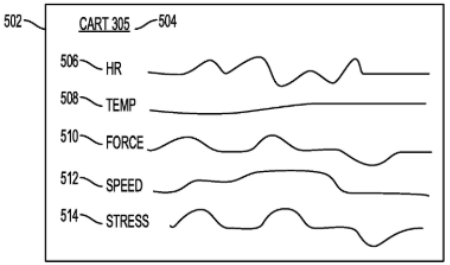


FIG. 5

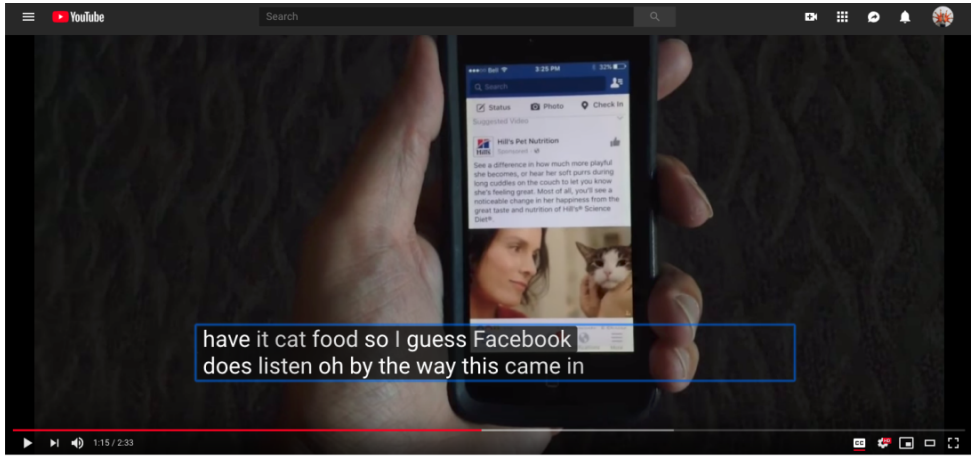
Judge orders Amazon to turn over Echo recordings in double murder case

2018/11/14

Docket No. 219-2017-CR-072

ORDER ON MOTION TO SEARCH IN LIEU OF SEARCH WARRANT

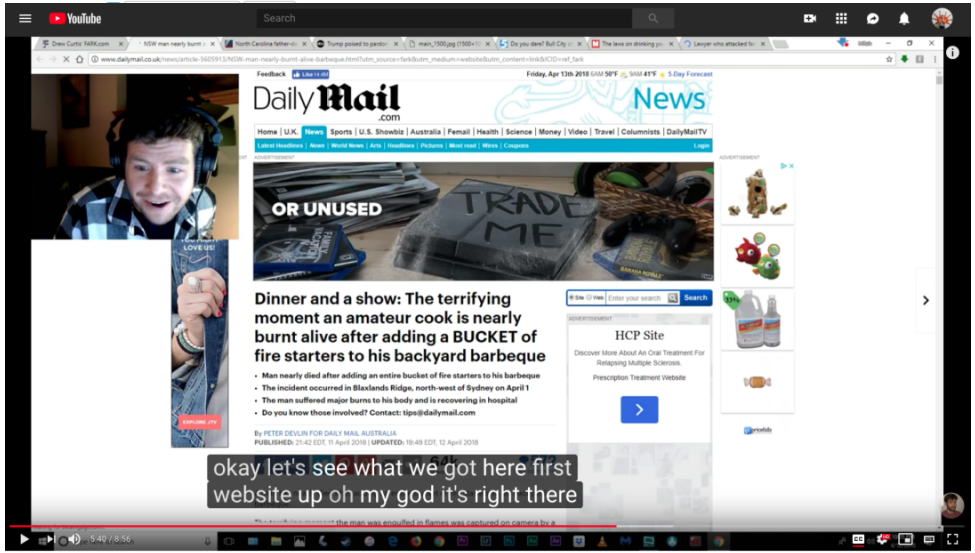
The State moves to allow the search of the server(s) and/or records maintained for or by Amazon.com for all recordings made by an Echo smart speaker with Alexa voice command capability, FCC ID number ZWJ-0823, from the period of January 27, 2017 to January 29, 2017, in addition to any information identifying cellular devices that were linked to that smart speaker during that time period, and to produce such information to the court. (Court index #204.) The State asserts there is probable cause to believe that evidence of crimes—audio recordings capturing the attack on victim Christine Sullivan that occurred in the kitchen of 979 Meaderboro Road and any events that preceded or succeeded the attack—may be found on the server(s) maintained by or for Amazon.com for all recordings made by the aforementioned Echo smart speaker. (Id.) The State asserts that the Echo smart speaker was among the items lawfully seized by investigators at 979 Meaderboro Road on February 3, 2017, and is currently in the possession of the New Hampshire State Police. (Id.)



Facebook iPhone Listening into our Conversations for Advertising TEST

Up next

AUTOPLAY



Is Google always listening: Live Test

Paranoid Users

Risks (Manufacturer considered to be **innocent**)

Rules Risk Case Study: Skill Squatting	**Firmware Risk** Case Study: Serial Port / Firmware Override
Software Risk Case Study: Eavesdropping	**Hardware Risk** Case Study: Hot Mic

Case Study: Skill Squatting Attacks

[1]D. Kumar et al., “Skill Squatting Attacks on Amazon Alexa,” presented at the 27th USENIX Security.
[2]N. Zhang, X. Mi, X. Feng, X. Wang, Y. Tian, and F. Qian, “Understanding and Mitigating the Security Risks of Voice-Controlled Third-Party Skills on Amazon Alexa and Google Home,” arXiv:1805.01525 [cs], May 2018.

Skill	Squatted Skill
Boil an Egg	Boyle an Egg
Main Site Workout	Maine Site Workout
Quick Calm	Quick Com
Bean Stock	Been Stock
Test Your Luck	Test Your Lock
Comic Con Dates	Comic Khan Dates
Mill Valley Guide	No Valley Guide
Full Moon	Four Moon
Way Loud	Way Louder
Upstate Outdoors	Upstate Out
Rip Ride Rockit	Rap Ride Rocket

Table 5: **Squattable Skills in the Alexa skills store** — We show 11 examples of squattable skills publicly available in the Alexa skill store, as well as squatted skill names an attacker could use to “squat” them.

Target Skill	Squatted Skill	Success Rate	Target Skill	Squatted Skill	Success Rate
Coal	Call	100.0%	Dime	Time	65.2%
Lung	Lang	100.0%	Wet	What	62.1%
Sell	Cell	100.0%	Sweeten	Sweden	57.4%
Heal	He’ll	96.4%	Earthy	Fi	53.3%
Sail	Sale	95.0%	Full	Four	26.8%
Accelerate	Xcelerate	93.7%	Outshine	Outshyne	21.2%
Rip	Rap	88.8%	Superhighway	Super Highway	19.7%
Mill	No	84.6%	Meal	Meow	18.3%
Con	Khan	84.2%	Bean	Been	17.8%
Luck	Lock	81.9%	Tube	Two	16.7%
Lull	Lol	81.9%	Main	Maine	3.1%
Dull	Doll	80.8%	Boil	Boyle	0.0%
Outdoors	Out Doors	71.0%	Loud	Louder	0.0%
Calm	Com	67.9%			

Table 4: **Skill Squatting Validation** — We show the results of testing 27 skill squatting attacks. The pairs of target and squatted skills are built using the squattable words of our training set. The success rates are computed by querying the speech samples of our test set. We are able to successfully squat 25 (92.6%) of the skills at least one time, demonstrating the feasibility of the attack.

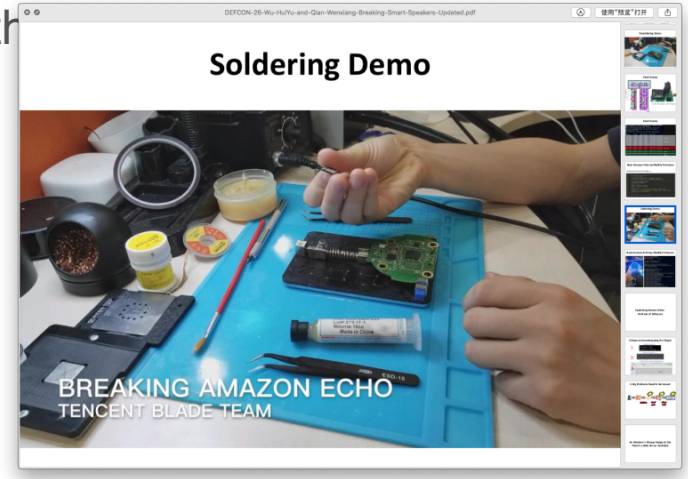
** D. Kumar et al., “Skill Squatting Attacks on Amazon Alexa,”*

Case Study: Eavesdropping Skills

- Amazon Eavesdropping Skills
- Researchers at cybersecurity firm Checkmarx were able to create an Alexa skill — an application for the voice-activated assistant — which was able to eavesdrop on users.
  - https://info.checkmarx.com/hubfs/Amazon_Echo_Research.pdf
  - https://www.youtube.com/watch?v=xfx90UJ4qGU
    - shouldEndSession
    - Reprompt

Case Study : Physical Contact

- **Serial Port:** British hacker Mark Barnes last year published a technique that uses physical access to a first-generation Echo to install malware on it via metal contacts accessible under its rubber base.
- **Firmware Override:** Security researchers from Tencent made a demo indicating that the firmware can be overridden with
  - Countermeasure: Secure boot.
  - WIRED



[1] Y. Gong and C. Poellabauer, "Protecting Voice Controlled Systems Using Sound Source Identification Based on Acoustic Cues," *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, Jul. 2018.

TABLE I
REPRESENTATIVE VOICE ATTACK TECHNIQUES

Attack Name	Attack Type	Implementation
GVS Attack [2]	Operating System	Continuously analyze the environment and conduct voice replay attack using the built-in microphone when the opportunity arises.
A11y Attack [3]	Operating System	Collect the voice of a user and perform a self-replay attack as a background service.
Monkey Attack [13]	Operating System	Bypass authority management of the OS and perform an interactive voice replay attack to execute more advanced commands.
Dolphin Attack [4]	Hardware	Emit ultrasound signal that can be converted into a legitimate speech digital signal by the MEMS microphone.
IEMI Attack [5]	Hardware	Emit AM-modulated signal that can be converted into a legitimate speech digital signal by the wired microphone-capable headphone.
Cocaine Noodles [14]	Machine Learning	Similar to the hidden voice command.
Hidden Voice Command [6]	Machine Learning	Mangle malicious voice commands so that they retain enough acoustic features for the ASR system, but become unintelligible to humans.
Houdini [15]	Machine Learning	Produce sound that is almost no different to normal speech, but fails to be recognized by both known or unknown ASR systems.
Speech Adversarial Example [7]	Machine Learning	Produce sound that is over 98% similar to any given speech, but makes the DNN model fail to recognize the gender, identity, and emotion.
Targeted Speech Adversarial Example [8]	Machine Learning	Produce sound that is over 99.9% similar to any given speech, but transcribes as any desired malicious command by the ASR.

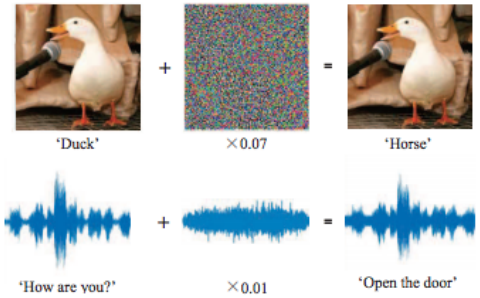


Fig. 3. An illustration of machine learning adversarial examples. Studies have shown that by adding an imperceptibly small, but carefully designed perturbation, an attack can successfully lead the machine learning model to making a wrong prediction. Such attacks have been used in computer vision (upper graphs) [16] and speech recognition (lower graphs) [7], [8], [15].

Risks (Manufacturer considered to be **innocent**)

Rules Risk Case Study: Skill Squatting	**Firmware Risk** Case Study: Serial Port / Firmware Override
Software Risk Case Study: Eavesdropping	**Hardware Risk** Case Study: Hot Mic

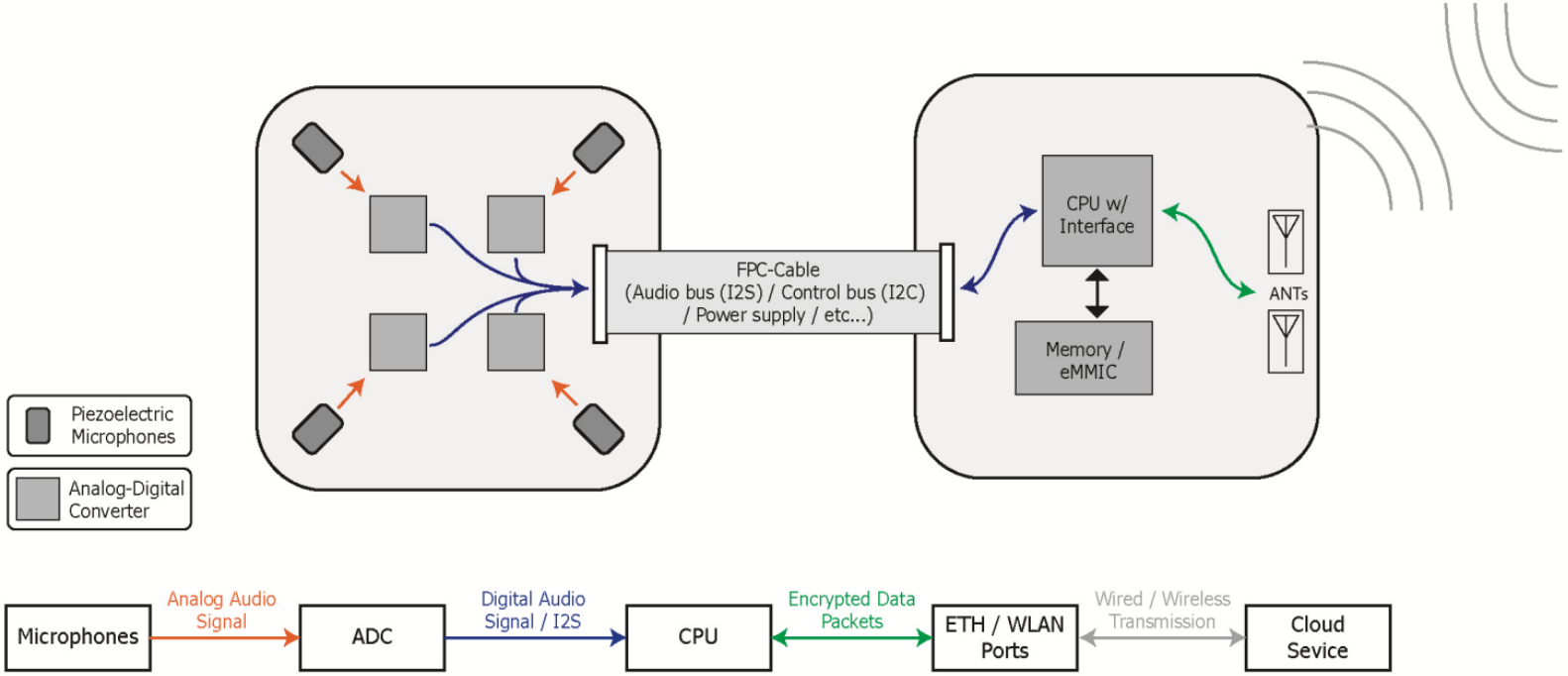
Case Study: Hot Mic

- *The first time we reveal a new attack surface caused by flaws in the hardware design*

- *Mute button*
  - *Last fence protecting the privacy*
  - *Scram Stop*

  - Does it eliminate all threats?

Basic Structure of Smart Speaker



I2S

- *Inter-IC sound*
  - *Synchronous serial digital audio transmission*
  - *Signals*
    - *Word clock*
    - *Bit clock*
    - *Data*
  - *N bit sample transferred sequencially*
  - *Using offset to transfer multiple channels through one data bus*

For I²S mode, the number of bit clocks per channel must be greater than or equal to the programmed word length of the data. Also the programmed offset value must be less than the number of bit clocks per frame by at least the programmed word length of the data.

10.3.6.4 DSP Mode

In DSP mode, the rising edge of the word clock starts the data transfer with the left-channel data first and is immediately followed by the right-channel data. Each data bit is valid on the falling edge of the bit clock. Figure 23 shows the standard timing for the DSP mode.

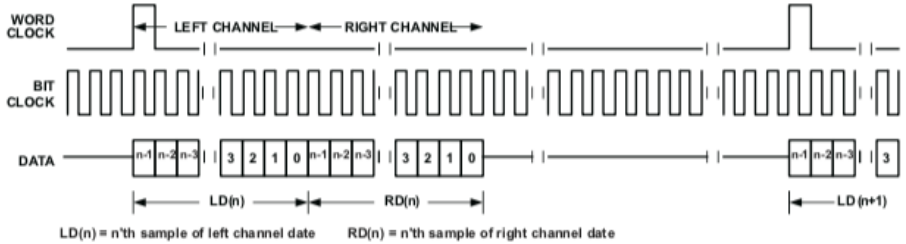


Figure 23. DSP Mode (Standard Timing)

Figure 24 shows the DSP mode timing with Ch_Offset_1 = 1.

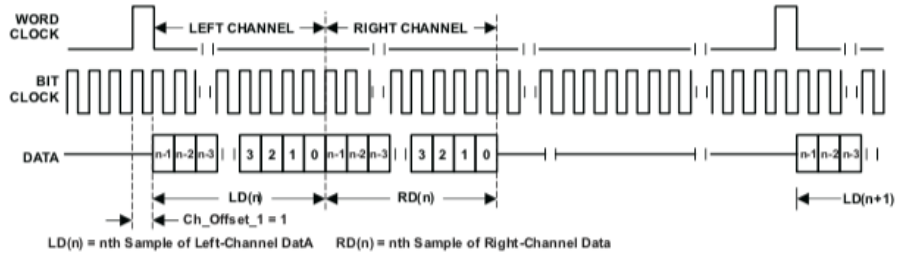
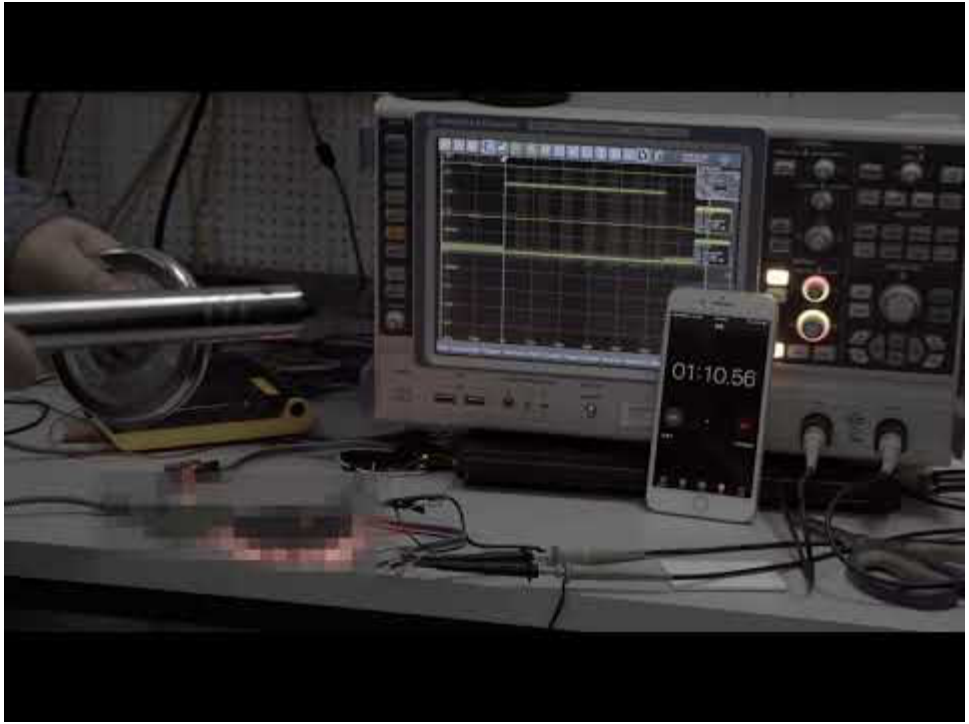
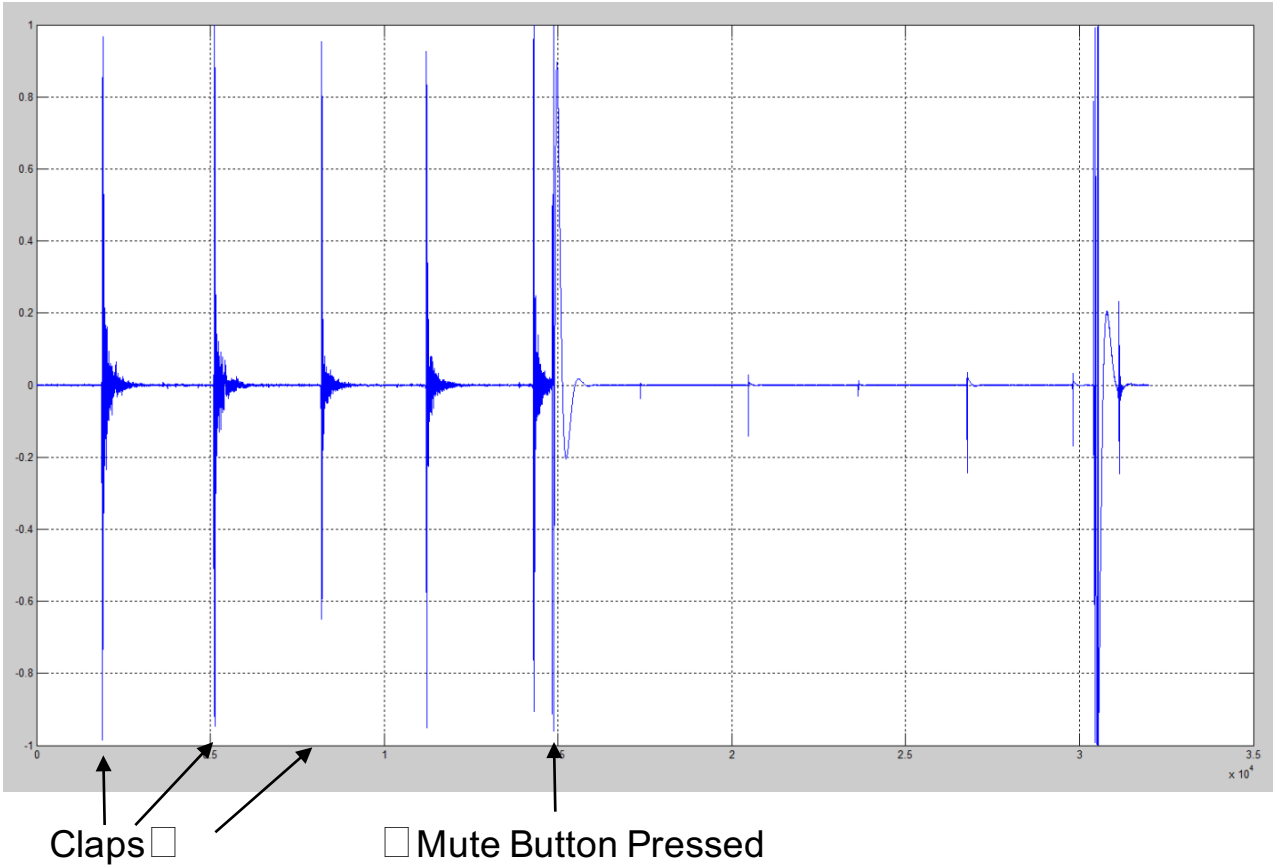
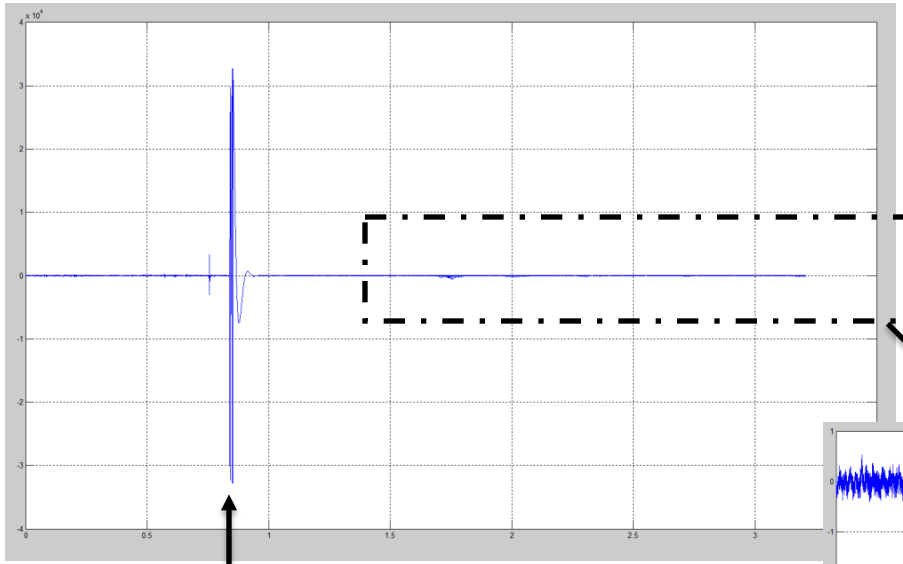


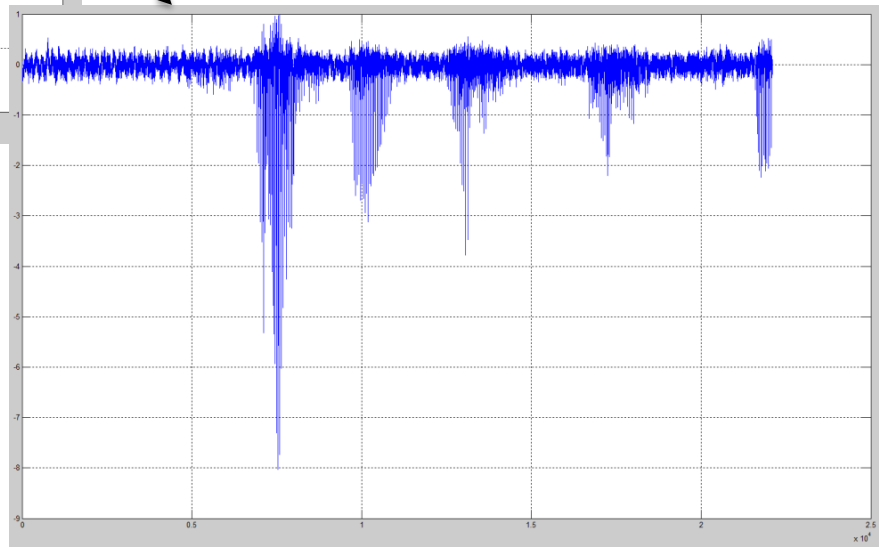
Figure 24. DSP Mode With Ch_Offset_1 = 1

Hot Mic Duration

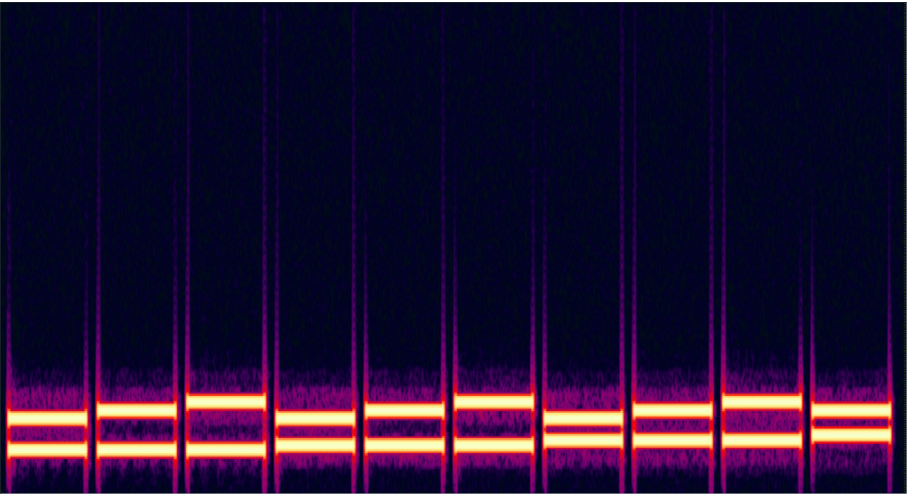


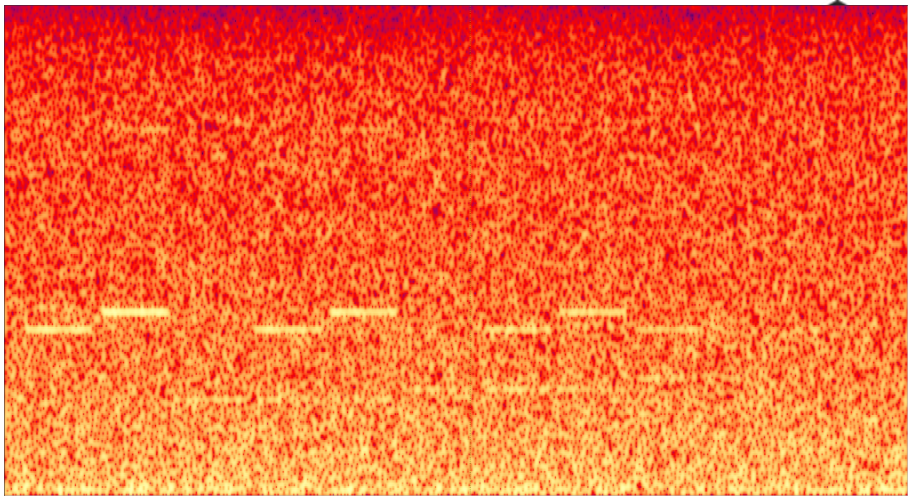

1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
1
2
2.5
3
3.5
x 10^4
Claps
Mute Button Pressed

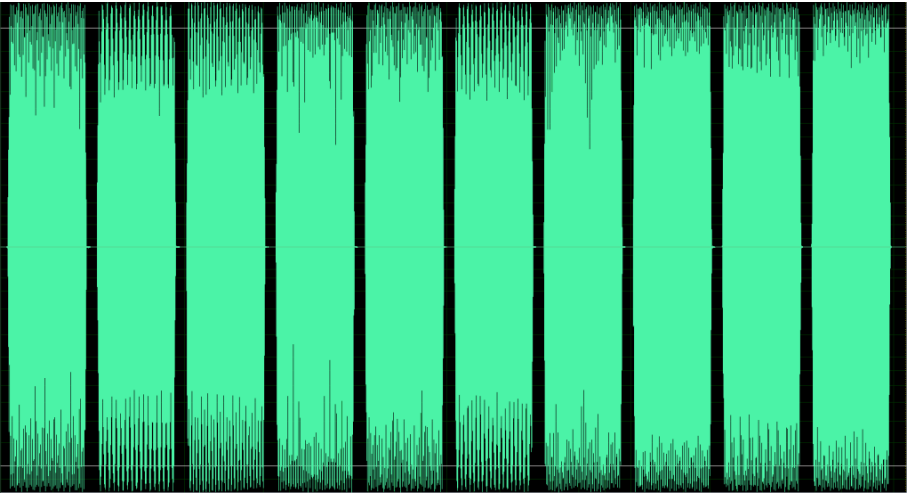


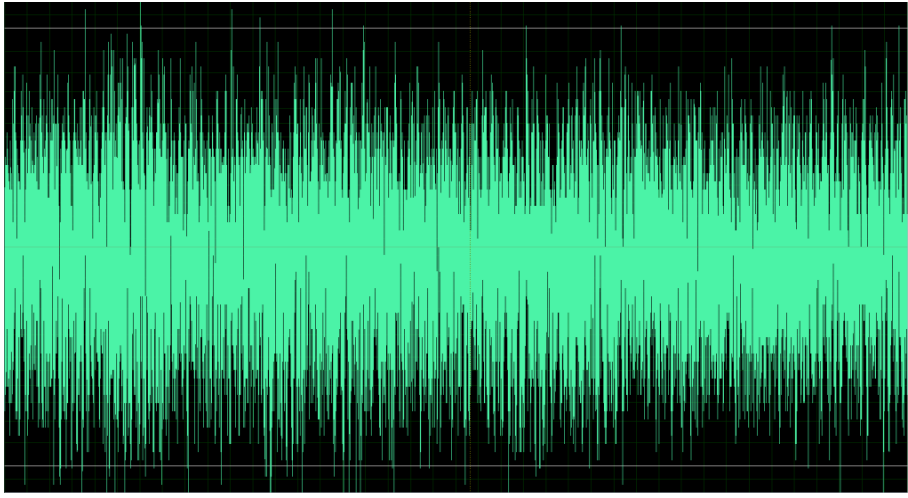


Mute Button Pressed









Summary for Hot Mic

- We don’t think vendors are on purpose.
  - No significant harm done …. Yet.
  - Still violating the expectation of the vendors
  - Users are not able to stop it or even aware of it

- Countermeasures for Hot Mic
  - Disable ADC immediately when muted

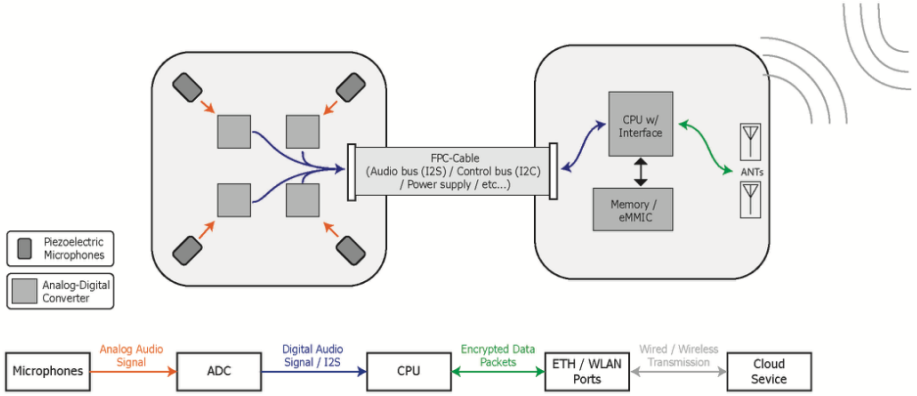
Confidence is vital

- To gain confidence from the market, innocent is required to be proven.
  - Sales promotion, no confidence no sales.

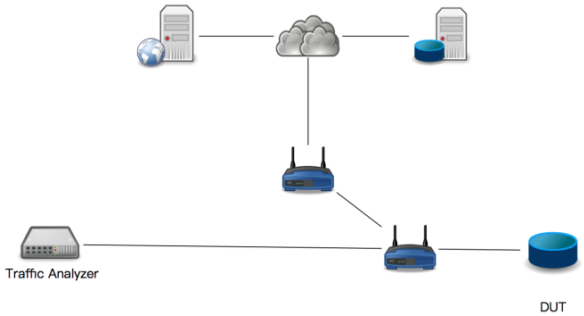
- How to do that?
  - Endorsement from relying authority after privacy and security audit.
    - However, there are increasing difficulties brought by:
      - System design diversity / ‘Fragmentation’
      - Protected source code
      - Encrypted network traffic
  - Self-proven Innocent:
    - Designs that can be self-proven or audited within the ability of **end-users**
      - **How?**

Insufficient Audit Surface

- Potential Audit Surface
  - Cloud Data Center
  - **Network Traffic**
  - **Hardware Bus**
  - Firmware
  - Mobile App
  - Source Code
    - Reproducible build



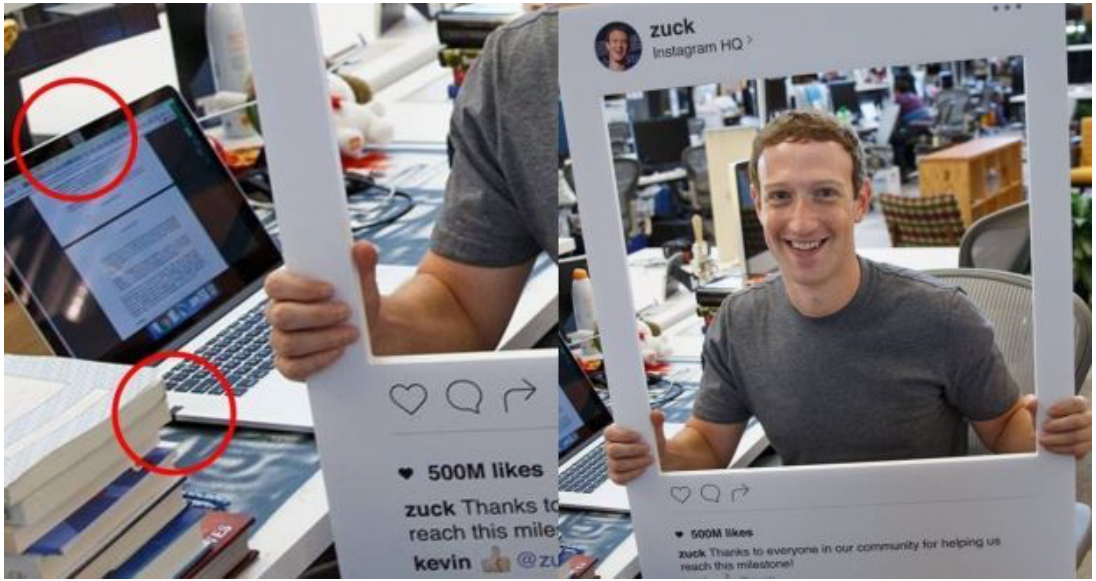
Network Traffic Monitoring



Hardware Bus

- Pros
  a. The signal between chips doesn't lie.
  b. No help from auditee needed.
  c. Enumerable types of hardware bus
- Limitations
  a. Reverse analysis in hardware, luck needed.
  b. Heterogeneity and function-integrated ASIC
  c. Black boxes still exists, hands tied
- Not a silver bullet. Still a promising point.

In tape we trust. ™



Privacy Indicator Coupling

- Privacy Sign
  - Indicator
  - Blocker

- The coupling between privacy data and the sign has to be trustworthy.
  - Physical Coupling *(Very hard to spoof, easy to understand & believe)*
    - Electrical contact / coupling:
      - LED
    - Mechanical design
      - Plug in / Eject

  - Virtual Coupling (Suspicious, audit needed)
    - Trusted Software: OS API
    - Trusted Firmware, TEE, Enclave, ...
    - Secure Chip
      - Apple T2 …

Physical Coupling

The Power of User Awareness

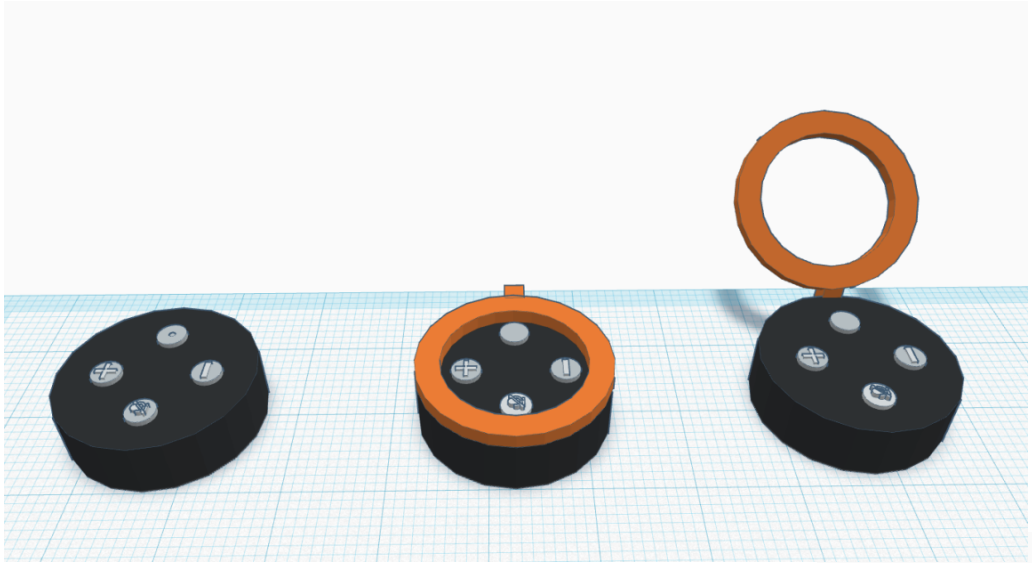
Vivo NEX Smartphone: “The Detector of Rogue Apps”

QQ browser opens a webpage and the vivo NEX camera rises. Tencent responds: confirms existence but does not shoot ***2018-6-28 17:28:42***

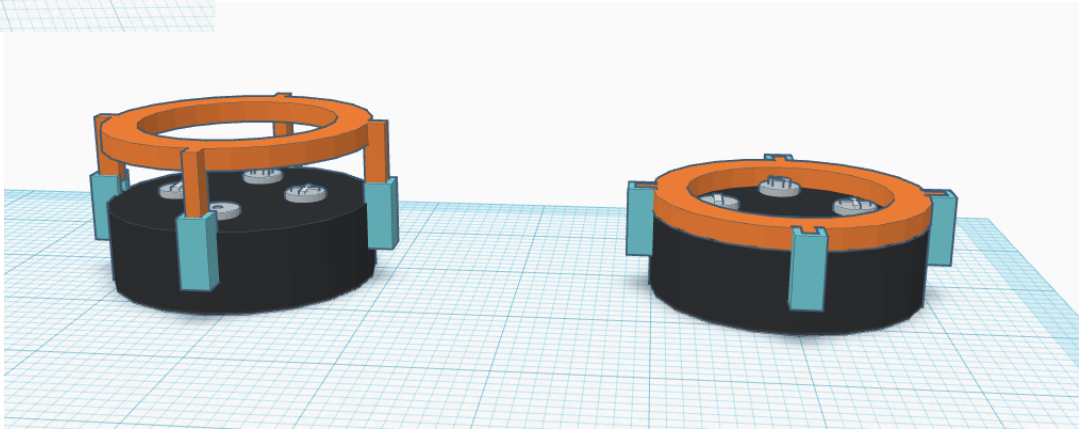
Recently, with the launch of the vivo NEX mobile phone, many users have already got the machine and are interested in the vivo camera of the vivo NEX. It is said that there is also the saying that the vivo NEX rogue App evaluator , but some netizens explained that the front camera of the vivo NEX suddenly rises. It may be that the app checks the availability of the front camera of the mobile phone, and not necessarily some apps appear. Rogue behavior. QQ browser opens a webpage and the vivo NEX camera rises. Tencent responds: confirms existence but does not shoot. Previously, some netizens found that when opening certain web pages on the Vivo NEX mobile phone device through the mobile QQ browser, the mobile phone camera will have a “lifting” action. In response, the QQ browser team responded by confirming that there is a camera action, but said that this action will not open the camera, and will not shoot or record. The mobile QQ browser does not collect any user privacy.

The following is the full response of the QQ browser team:

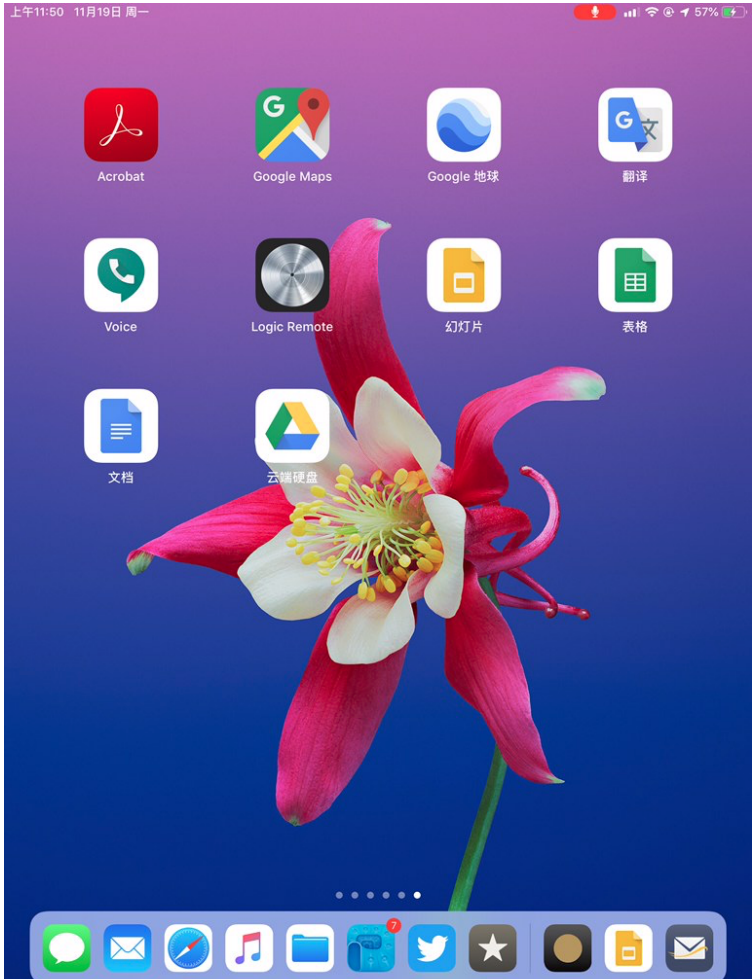
> *Instructions for using VQ to open some web pages will raise the Vivo NEX camera problem. The QQ browser team received user feedback. When the user opens certain web pages on the Vivo NEX mobile device through the mobile QQ browser, the mobile phone camera will have a “lifting and lowering action”. The QQ browser technical team has carried out the problem. Test the recurrence and confirm that there is a camera action, but this action will not turn on the camera, and will not shoot or record. Now the following is explained for this problem:*
>
> *1. Reasons for the problem and technical principles:*
>
> *In order to realize the user's use of some functions (such as scanning QR code), the W3C specification has a front-end standard interface navigator.mediaDevices.enumerateDevices() to traverse the media device and obtain camera parameters for subsequent use. Android has two sets of APIs to operate the camera. , camera1 and camera2, where camera2 can obtain the camera parameters without opening the camera, and camera1 needs to call Camera.Open() function to initialize to get the camera handle, and then get the camera parameters through the camera handle (https://developer.android. Com/reference/android/hardware/Camera). Considering that camera2 has many problems in performance and compatibility in applications such as AR cameras, the mobile QQ browser kernel uses the camera1 interface, which causes the camera to “lift” the VIVO NEX mobile phone user experience.*
>
> *2. The mobile QQ browser does not collect any privacy of the user.*
>
> *In fact, the user's camera of the VIVO NEX mobile phone does not pop up completely when the user needs to obtain the parameters of the camera during the process of opening some webpages with the mobile QQ browser (the user can open the webpage https://qiyaoyuan using the QQ browser. Github.io/source/webar.html test recurrence), and the camera did not do any shooting or collection behavior, mobile QQ browser does not collect any user privacy. This page only calls the mediaDevices.enumerateDevices interface and has no other operations.*
>
> *3. We will optimize the user experience and experience.*
>
> *Thanks again to the users for their attention and feedback on the QQ browser, and apologize to the user for misunderstanding and confusion, we will optimize and prompt this experience.*
>
> *QQ browser product team*



Detachable Microphone
Daughterboard



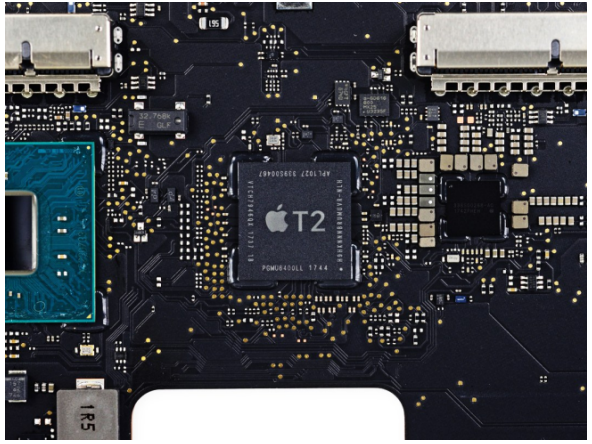
Virtual Coupling


上午11:50 11月19日 周一
57%
Acrobat
Google Maps
Google 地球
翻译
Voice
Logic Remote
幻灯片
表格
文档
云端硬盘

Macbook T2 Security Chip: Hardware mic disconnect

Trusted Indicator

- Provable: Not yet.
- Auditable: Make sense.



[Ref: WIRED]

Hardware microphone disconnect

All Mac notebooks with the Apple T2 Security Chip feature a hardware disconnect that ensures the microphone is disabled whenever the lid is closed. On 13-inch MacBook Pro and MacBook Air computers with the T2 chip, this disconnect is implemented in hardware alone, and prevents any software—even with root or kernel privileges in macOS, and even the software on the T2 chip—from engaging the microphone when the lid is closed. (The camera is not disconnected in hardware because its field of view is completely obstructed with the lid closed.)

-- Apple T2 Security Chip Security Overview - October 2018

Auditable $\supseteq$ Provable

Res ipsa loquitur

Provable is the better design.

Key Takeaways

1. Detection of potential privacy violation will be an everlasting problem.
2. Hot Mic: for the first time, we revealed a new threat in smart speakers.
3. The concept and methodology of auditable and provable designs.

Q&A

Thank you.