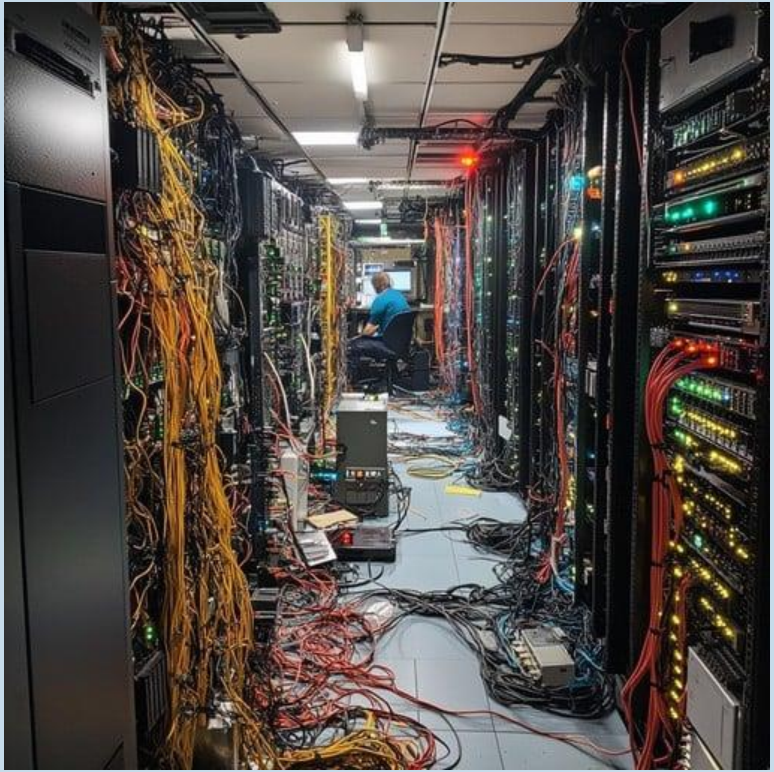
Hardware Security 1: Physical Security

Security Engineering (Spring 2026)

Lecturer: Jingjie Li & Daniel Woods



- Do you trust your code running on a messy server room open to the public?
- Physical access expose surfaces of security and privacy breach that can be more challenges than digital ones
- (1) Using physical mindset to evaluate digital attacks
- (2) Physical security as the root of trust for digital and cyber-physical system

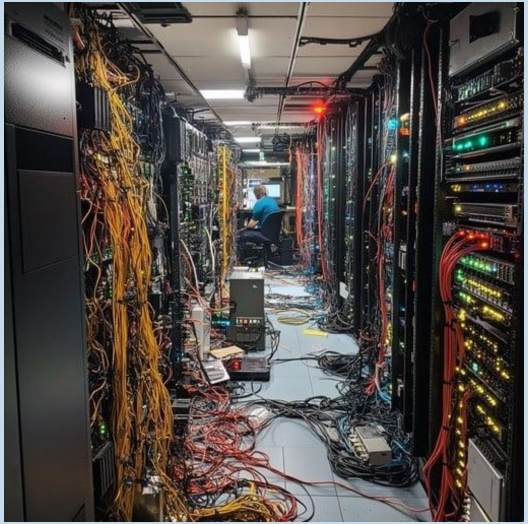
Physical Security Philosophy

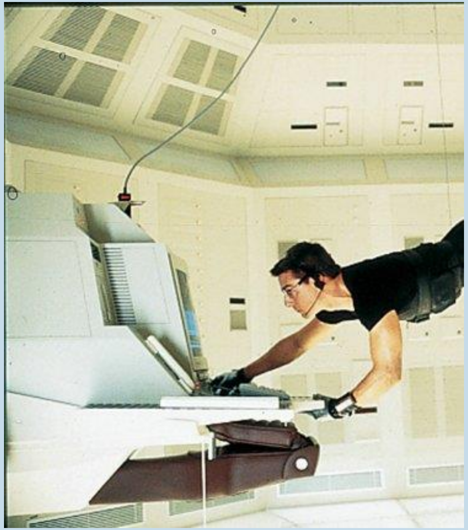
- Locks, and walls, will be some part of your infrastructure at some level
- While the techniques are simpler than digital security, the weaknesses are often as subtle.
- **Five stage of physical security: Deter-detect-alarm-delay-respond**
  - **Time matters!**

Attacker Capabilities in Threat and Risk Assessment

- Derek – 19-year old addict, **opportunistic** criminal looking for simple low-risk opportunities
- Charlie – 40-year old with 7 convictions, Not intelligent, **but cunning and experienced**, so knows the tools of the trade
- Bruno – "**gentleman criminal**" who steals art and takes pride in his work. Bruno is adept at lock and alarm hacking, and is interested in getting into computer hacking too.
- Abdurrahman – head of a dozen agents. He has access to **specialist weapons and PhD-grade technical support**
- Unskilled -> Skilled -> Highly Skilled with help -> Highly Skilled with resources

Who will be the most likely attacker?





How the security level aligns with your asset?

- E.g. wireless smart cards and card readers using challenge-response protocols
- Mifare Classic: Vulnerable but still widely deployed!
- All the usual crypto issues apply: weak ciphers, bad random number generators, short keys…

- Challenges in updating security (cryptography) protocol
- Easy to defeat (e.g., due to budget, form factor, export control, etc.)

How the security level aligns with your asset?

- E.g. wireless smart cards and card readers using challenge-response protocols
- Mifare Classic: Vulnerable but still widely deployed!
- All the usual crypto issues apply: weak ciphers, bad random number generators, short keys…

Don’t design for Charlie to keep about Bruno!

Alarm – it matters as locks and walls can be defeated

Fall of Constantinople (1453)

Maginot Line

Temper resistance – how do we know the physical system has been attacked

Temper resistance – Inspection of integrity

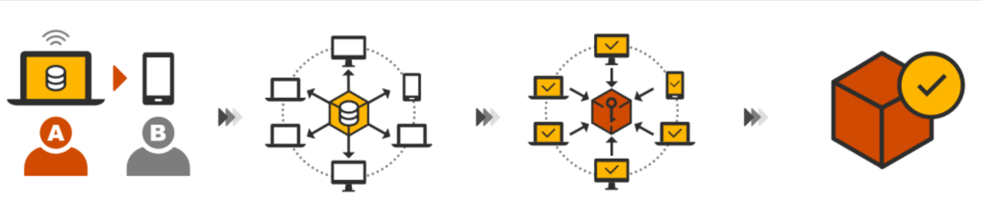
- Simultan presses, intaglio, letterpress, embossing, watermarks, microprinting, metal threads…
- Primary vs Secondary vs Tertiary inspectors
- Race against the forgers – add new features before your secondary inspectors get fooled

Temper resistance – Moving to the digital world

- What about digital currency?
- What about AI-generated content?

Temper resistance – Moving to the digital world

- What about digital currency?
- What about AI-generated content?



Blockchain

Goolge's SynthID watermark

What is the wall for digital devices?

Temper resistance – Hardware Security Modules

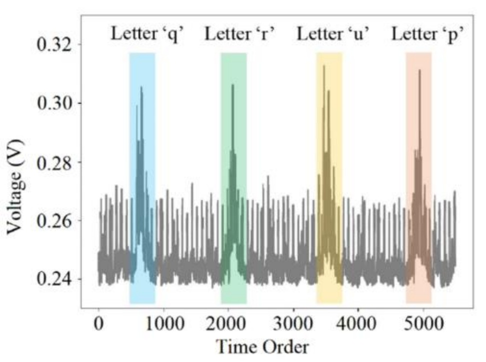
- Store confidential data and perform critical computation
- RAM set to 0 (destroying and refreshing encryption keys) when the physical case is open
- Meaning maintenance people can't get the key
- Early version vulnerable to cut through and people "seal" cores with epoxy resin
- Still leave information somewhere?

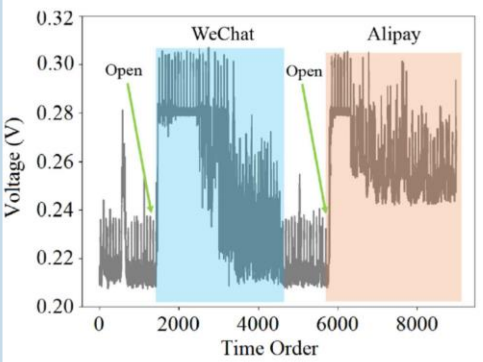
Temper resistance – Hardware Security Modules

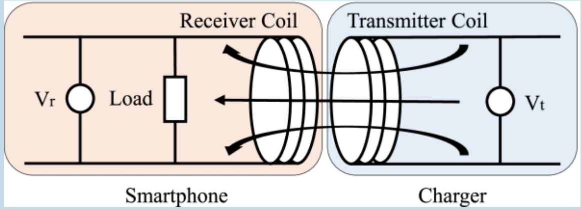
- Key bits get burnt into SRAM, so data / bit status remains even when it's refreshed!
- Similar phenomenon in DRAM/Flash

Side channels: physical channels carry more information than you want

- Information breach from side channels



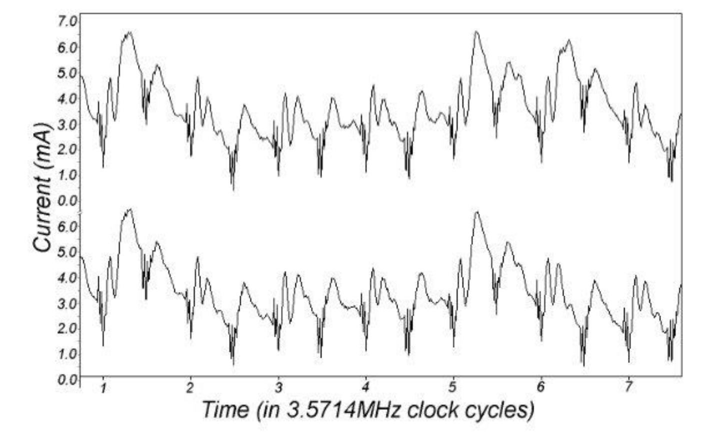




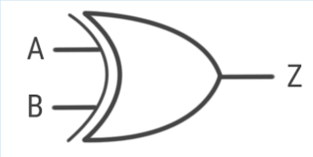
Liu, J., Zou, X., Zhao, L., Tao, Y., Hu, S., Han, J. and Ren, K., 2022. Privacy leakage in wireless charging. *IEEE Transactions on Dependable and Secure Computing*, *21*(2), pp.501-514.

Side channels: physical channels carry more information than you want

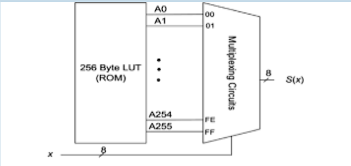
- Revealing encrypted data through power analysis



Differential Power Analysis, Kocher, Jaffe and Jun, CRYPTO ‘99



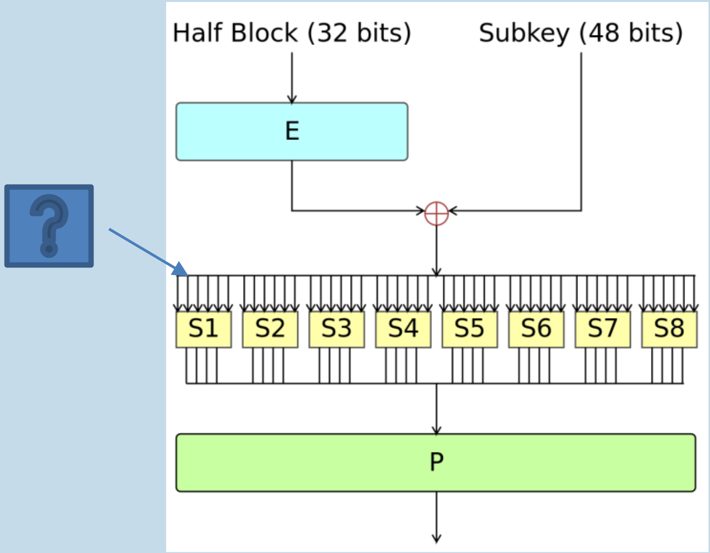
XOR

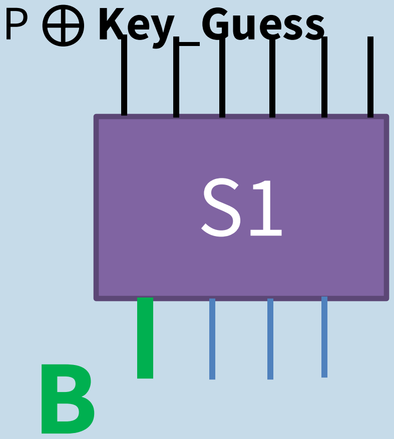


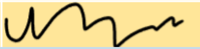
S BOX lookup table

- Power consumption is different for different bit operations, correlated with bit flips in CMOS
- Power side channel give adversary a way to measure and verify guess results

Side channels: physical channels carry more information than you want

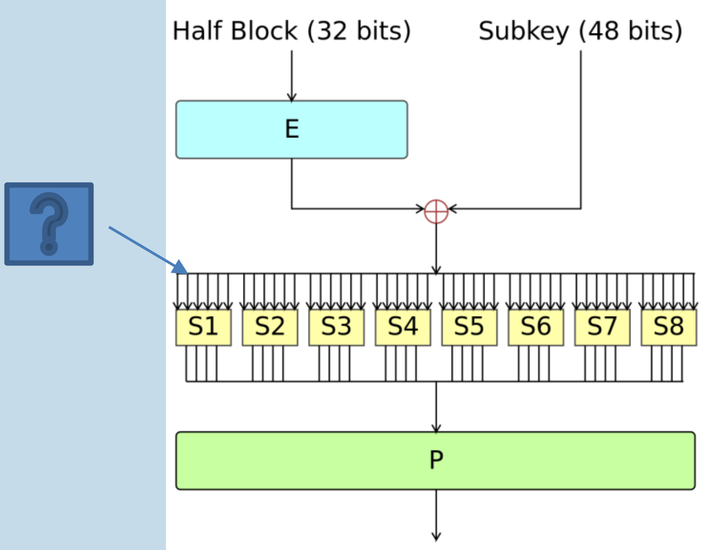


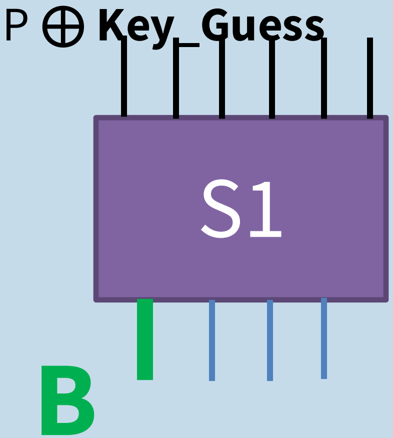


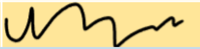
Plaintext	Trace
0x12345678…	
0x898979AB…	
0xDE424567…	
0XA0003341…	

Goal: guess the encryption key (e.g., in DES, a symmetric block cipher)

Side channels: physical channels carry more information than you want

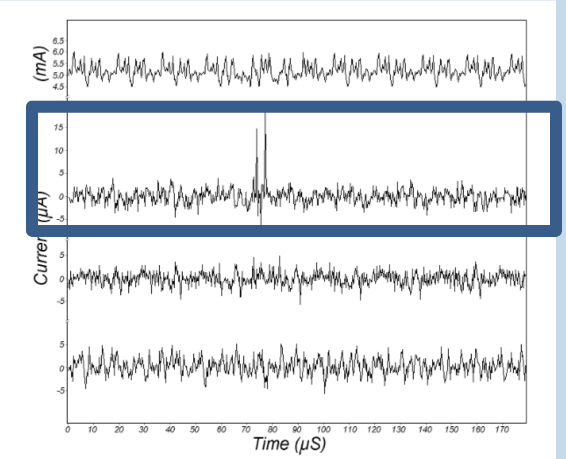




Plaintext	Trace
0x12345678...	
0x898979AB...	
0xDE424567...	
0XA0003341...	

- With an input text and a current guessed key
- Run DES algorithm and take power measurement
- Group power traces by predicted output 0 vs 1

Side channels: physical channels carry more information than you want



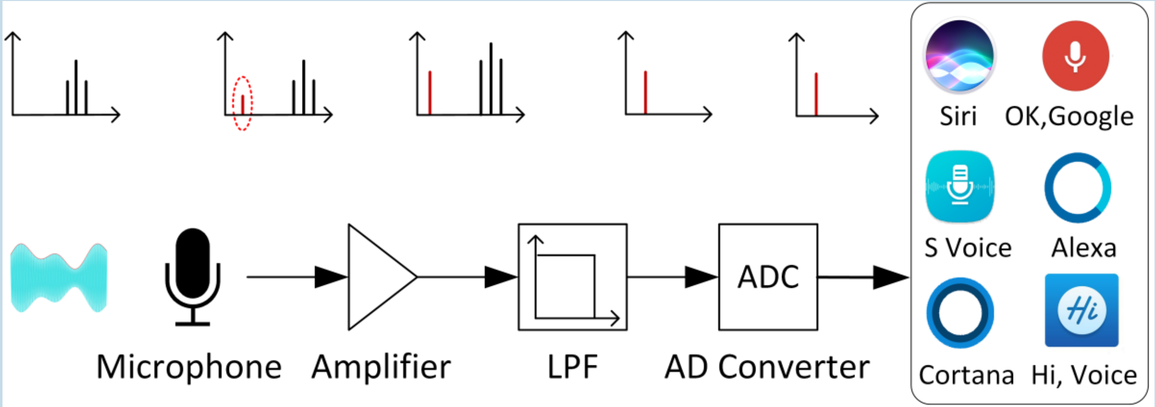
- Avg(Group 1 – Group 0)
- If we see a spike? Guess is correct – due to the power difference of right bit flips
- If we don't? Move on to the next guess and try again!
- Still, if we can guess it a byte at a time, and get a notification as to whether that byte is correct, **it turns an exponential search into an easy linear one**.

What causes smart speaker ghost activation?

Side channels: leveraging physical "fault" to inject malicious input

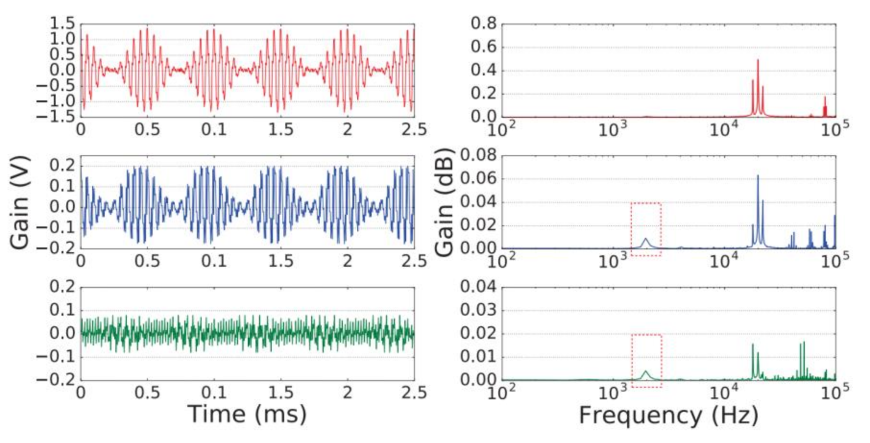
Zhang, G., Yan, C., Ji, X., Zhang, T., Zhang, T. and Xu, W., 2017, October. Dolphinattack: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security* (pp. 103-117).

Side channels: leveraging physical "fault" to inject malicious input



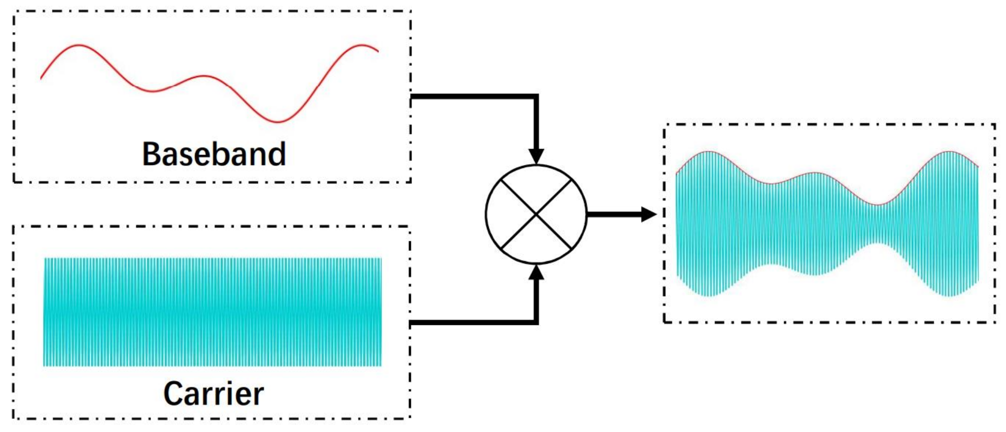
- How to inject malicious comment without human notice?

Side channels: leveraging physical "fault" to inject malicious input



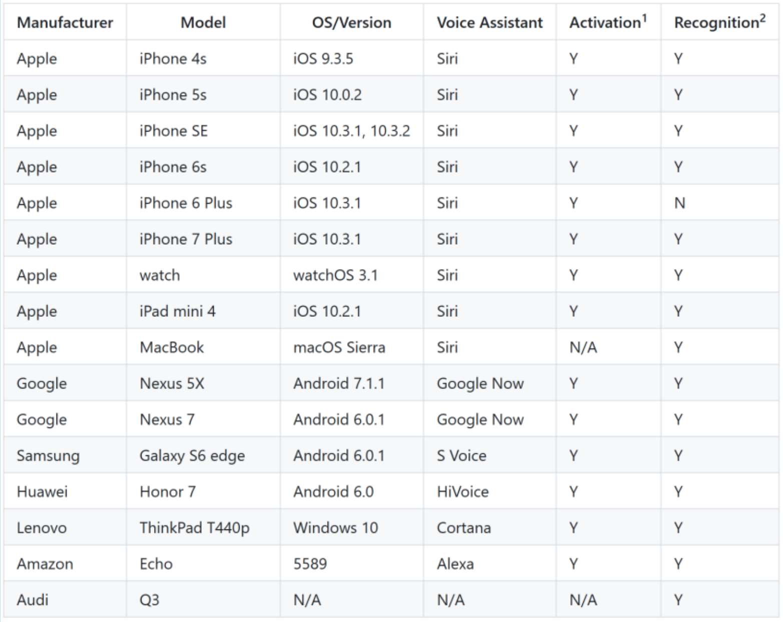
- High frequency sound that human can't hear will leave a low frequency "shadow" through MEMS microphone demodulation

Side channels: leveraging physical "fault" to inject malicious input



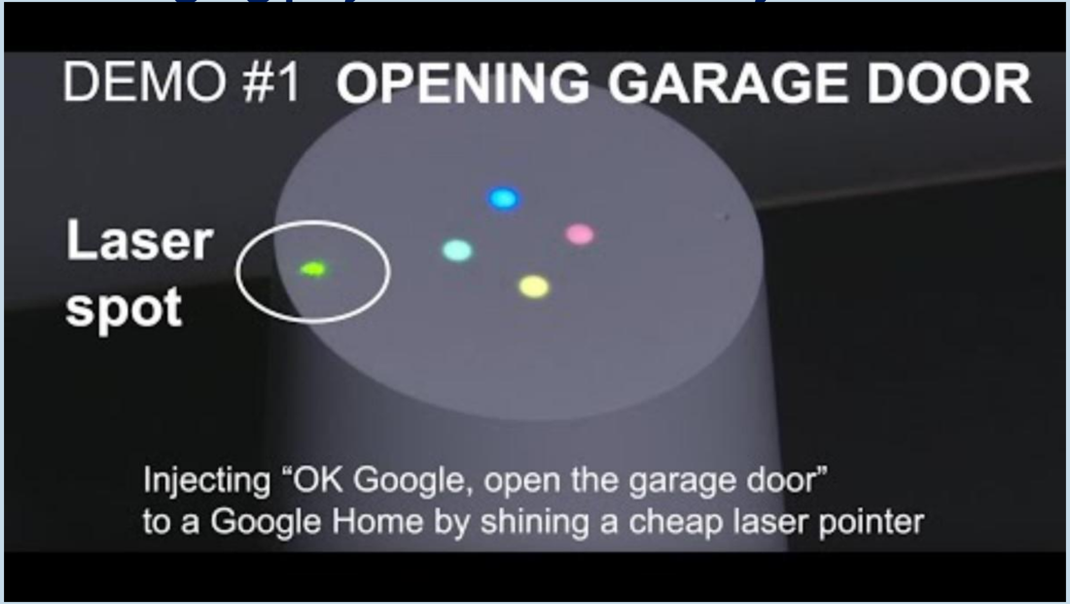
- Idea: using high frequency sound as the carrier band to modulate voice command

Side channels: leveraging physical "fault" to inject malicious input



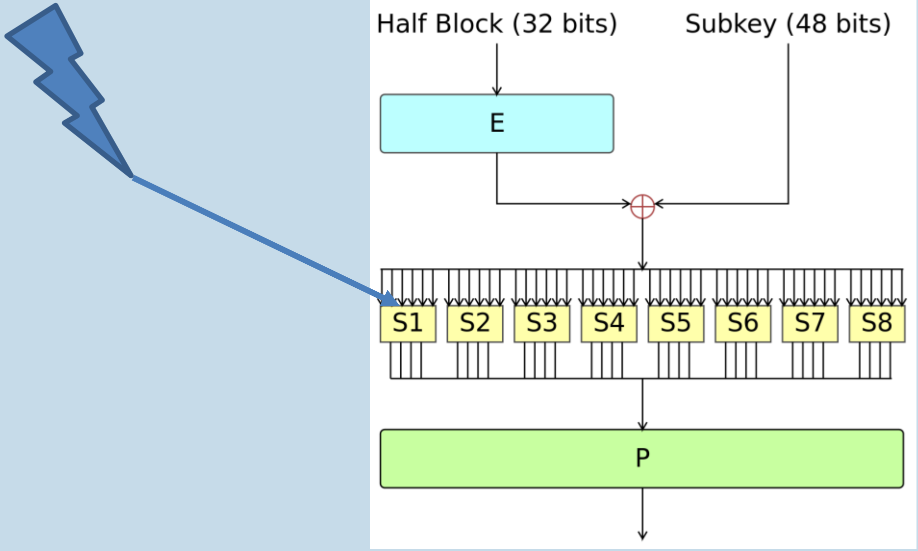
Manufacturer	Model	OS/Version	Voice Assistant	Activation[1]	Recognition[2]
Apple	iPhone 4s	iOS 9.3.5	Siri	Y	Y
Apple	iPhone 5s	iOS 10.0.2	Siri	Y	Y
Apple	iPhone SE	iOS 10.3.1, 10.3.2	Siri	Y	Y
Apple	iPhone 6s	iOS 10.2.1	Siri	Y	Y
Apple	iPhone 6 Plus	iOS 10.3.1	Siri	Y	N
Apple	iPhone 7 Plus	iOS 10.3.1	Siri	Y	Y
Apple	watch	watchOS 3.1	Siri	Y	Y
Apple	iPad mini 4	iOS 10.2.1	Siri	Y	Y
Apple	MacBook	macOS Sierra	Siri	N/A	Y
Google	Nexus 5X	Android 7.1.1	Google Now	Y	Y
Google	Nexus 7	Android 6.0.1	Google Now	Y	Y
Samsung	Galaxy S6 edge	Android 6.0.1	S Voice	Y	Y
Huawei	Honor 7	Android 6.0	HiVoice	Y	Y
Lenovo	ThinkPad T440p	Windows 10	Cortana	Y	Y
Amazon	Echo	5589	Alexa	Y	Y
Audi	Q3	N/A	N/A	N/A	Y

Side channels: leveraging physical "fault" to inject malicious input


DEMO #1 OPENING GARAGE DOOR
Laser spot
Injecting “OK Google, open the garage door” to a Google Home by shining a cheap laser pointer

Sugawara, T., Cyr, B., Rampazzi, S., Genkin, D. and Fu, K., 2020. Light commands:{Laser-Based} audio injection attacks on {Voice-Controllable} systems. In *29th USENIX Security Symposium (USENIX Security 20)* (pp. 2631-2648).

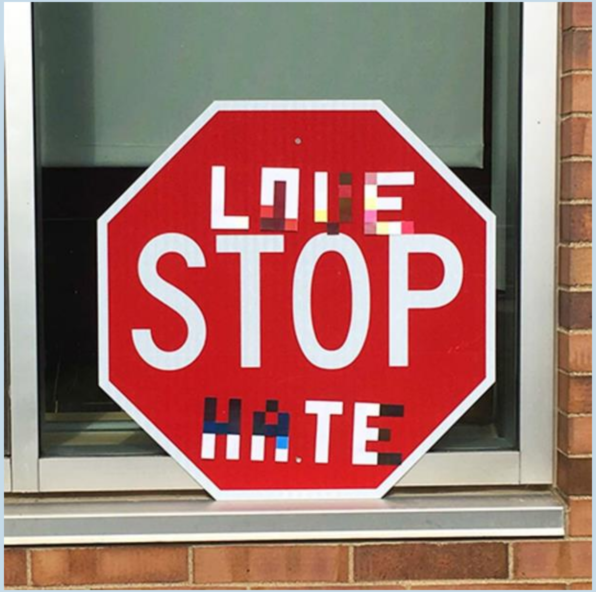
Side channels: leveraging physical "fault" to inject malicious input



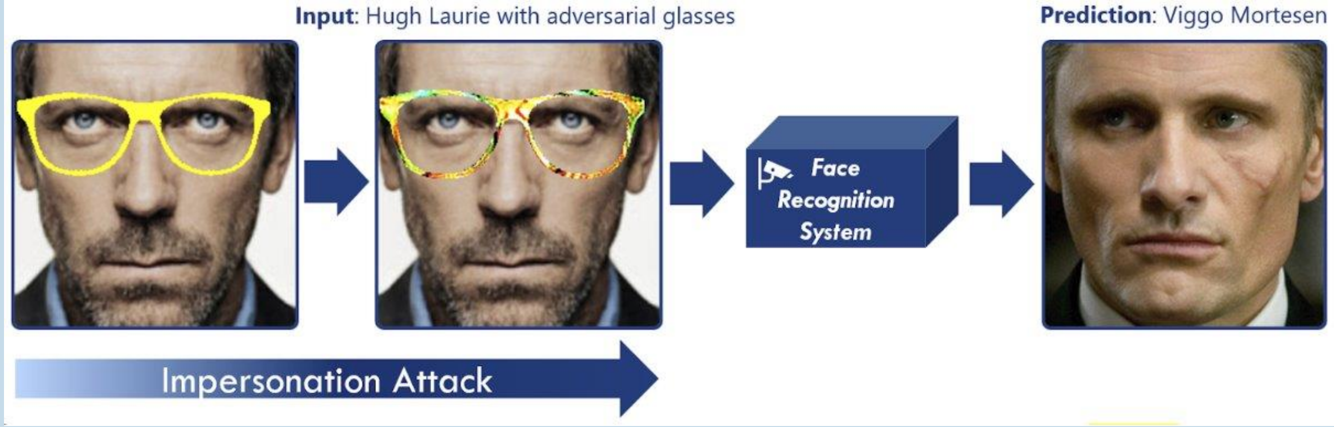
Of course, hacking encryption key again...

How likely are side channel attacks?

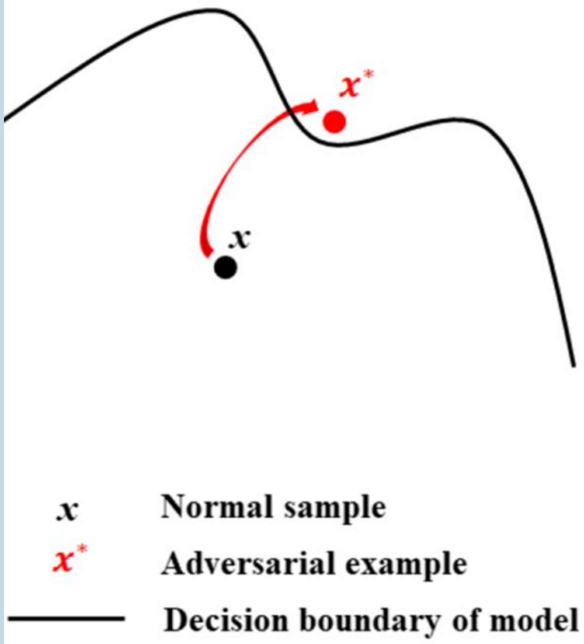
Computer and digital systems see a different physical world than us….
When digital system is at fault…



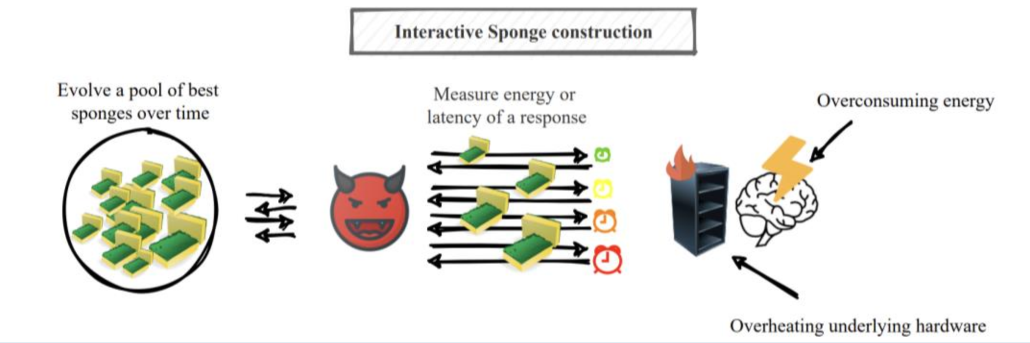
Computer and digital systems see a different physical world than us


Input: Hugh Laurie with adversarial glasses
Prediction: Viggo Mortesen
Face Recognition System
Impersonation Attack

Sharif, M., Bhagavatula, S., Bauer, L. and Reiter, M.K., 2019. A general framework for adversarial examples with objectives. *ACM Transactions on Privacy and Security (TOPS)*, *22*(3), pp.1-30.


x^*
x
x Normal sample
x^* Adversarial example
Decision boundary of model

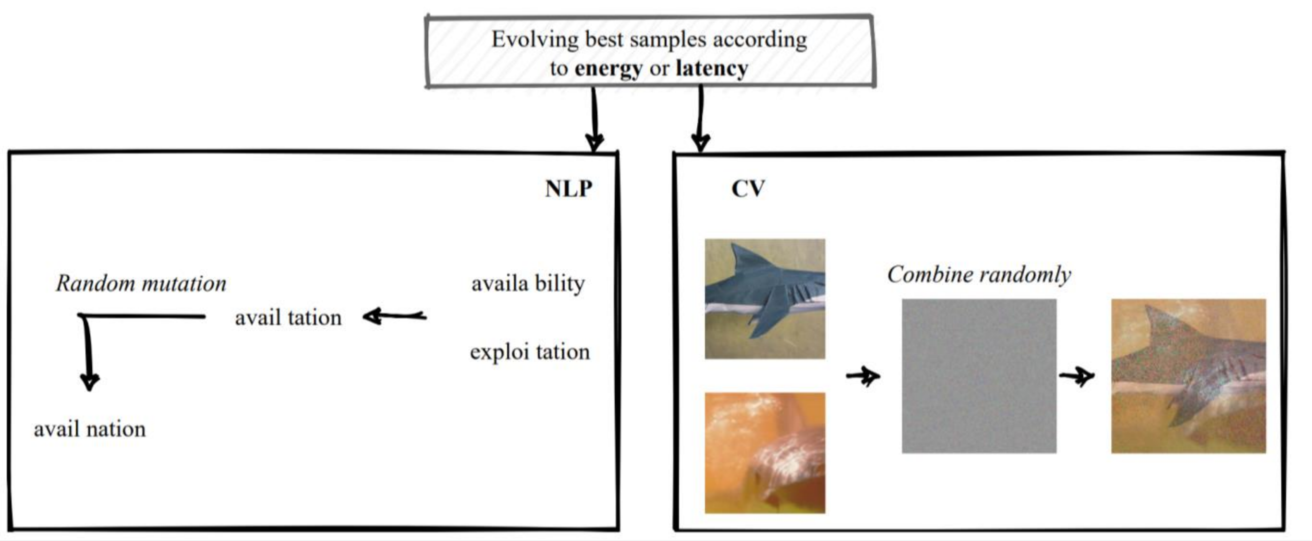
When digital attack causes physical damages


Interactive Sponge construction
Evolve a pool of best sponges over time
Measure energy or latency of a response
Overconsuming energy
Overheating underlying hardware

$$\begin{aligned} E =& (P_{\text{static}} + P_{\text{dynamic}}) \times t \\ =& \Big(\Big[\sum I_s \times \overbrace{(e^{\frac{qV_d}{kT}} - 1)}^{\text{overheat or increase overall consumption}} \times V_{\text{core}}\Big] \\ &+ \Big[\underbrace{\alpha}_{\text{more activity of the board}} \times C \times V_{\text{core}}^2 \times \overbrace{f}^{\text{throttle or exploit load predictor}}\Big]\Big) \times \underbrace{t}_{\text{run for longer or exploit the predictor}} . \end{aligned}$$

Shumailov, I., Zhao, Y., Bates, D., Papernot, N., Mullins, R. and Anderson, R., 2021, September. Sponge examples: Energy-latency attacks on neural networks. In *2021 IEEE European symposium on security and privacy (EuroS&P)* (pp. 212-231). IEEE.

When digital attack causes physical damages



Evolving best samples according to **energy** or **latency**

NLP

Random mutation

availa bility

exploi tation

avail tation

avail nation

CV

Combine randomly