Auditory Eyesight: Demystifying μs-Precision Keystroke Tracking Attacks on Unconstrained Keyboard Inputs

Yazhou Tu, **Liqun Shan**, Md Imran Hossen, Sara Rampazzi, Kevin Butler, Xiali Hei

UNIVERSITY of LOUISIANA LAFAYETTE

UF UNIVERSITY of FLORIDA

Auditory Devices and Applications

- Laptops, smart speakers, smart TVs, remote controllers
- Leakage in speech [1]

Office

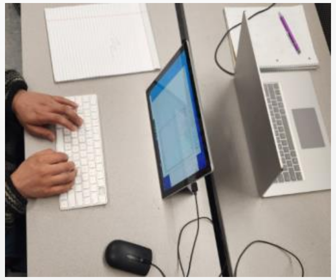
Meeting Room

Video Conference

Streaming

Home

Classroom



Library

[1] Lau et al. *Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers*. Proceedings of the ACM on human-computer interaction. 2018.

Privacy Perception

- **How about the leakage of sensitive information not communicated via speech?**

- **Users' natural, unconstrained keyboard inputs**
  - Such as account names, passwords, IDs, SSH credentials, real-world texts (with punctuation, numbers, capital letters, typos), and emails

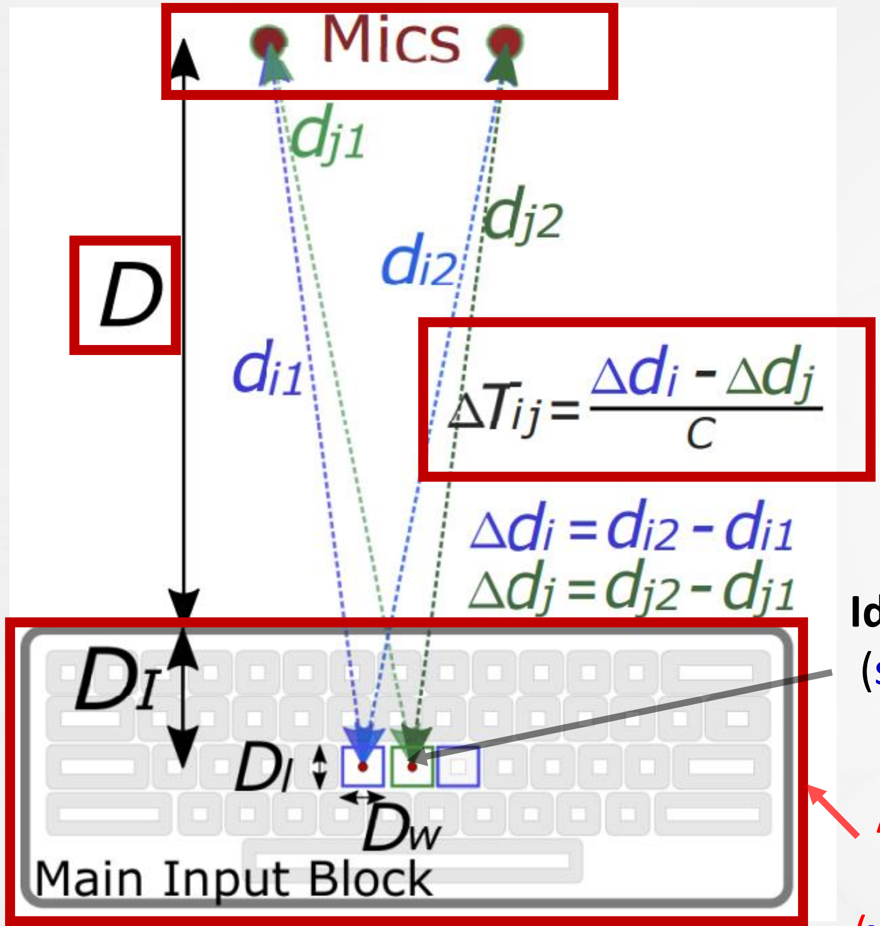
Challenges of Inferring Unconstrained Inputs

- **Expanded Solution Space**
  - From **single-letter-case alphabetic keys/words** and **known sequences** in a dictionary or training dataset
  - to **arbitrary keyboard inputs**

- Auditory devices are not designed for distinguishing a large number of **compactly spaced keys** from a distance
  - E.g., **Over 50 commonly used keys** in a **27.2×7.1 cm area**

- Complex keystroke sound physics
  - **Imperfect** sound source and measuring
  - **Interference** from vibrated keyboard, diffraction, reverberation

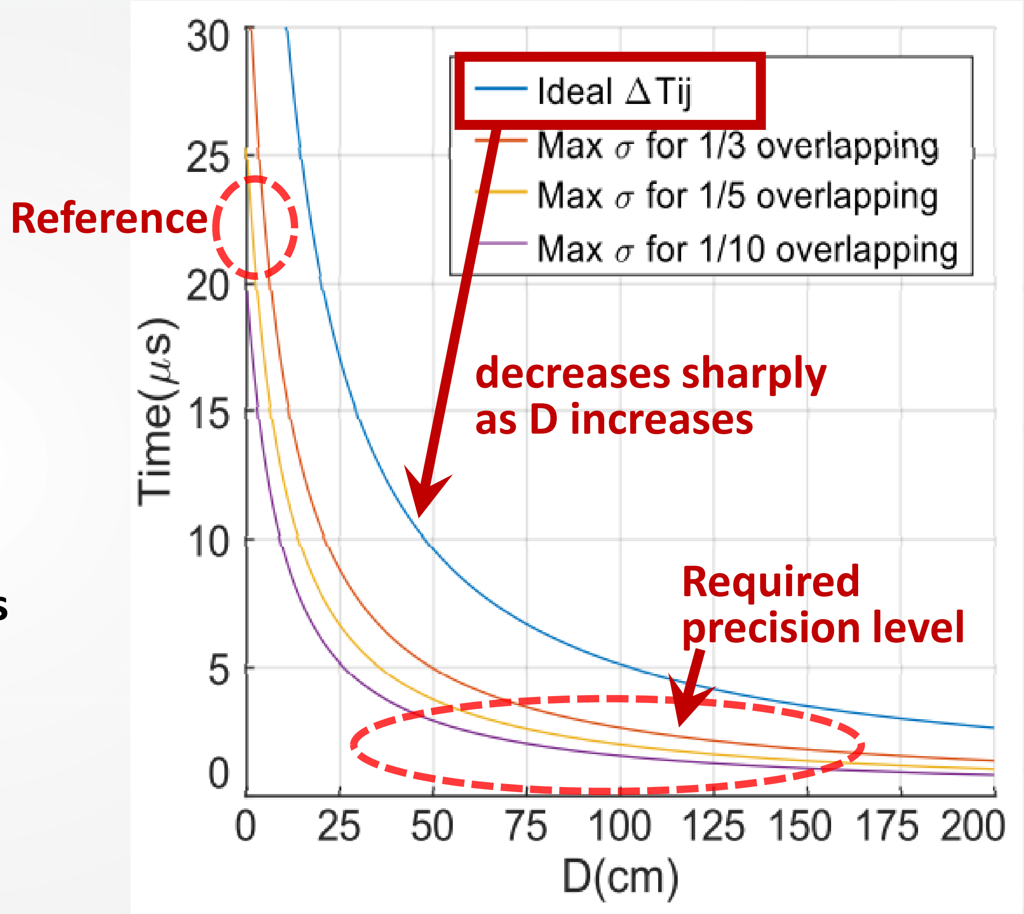
Keystroke Localization Precision Analysis



Challenge: Large number of keys (>50) in compact keyboard area (27.2×7.1 cm) including non-alphabetic keys

Ideal sound sources (sounds from exact keycap centers)

Actual, imperfect sound sources (vibrated keyboard, diffraction, reverberation)



Challenge: Required precision(close to μs)
Reference: hardware sampling interval (22.7 μs) with standard audio sample rate (44.1 kHz)

Methodology

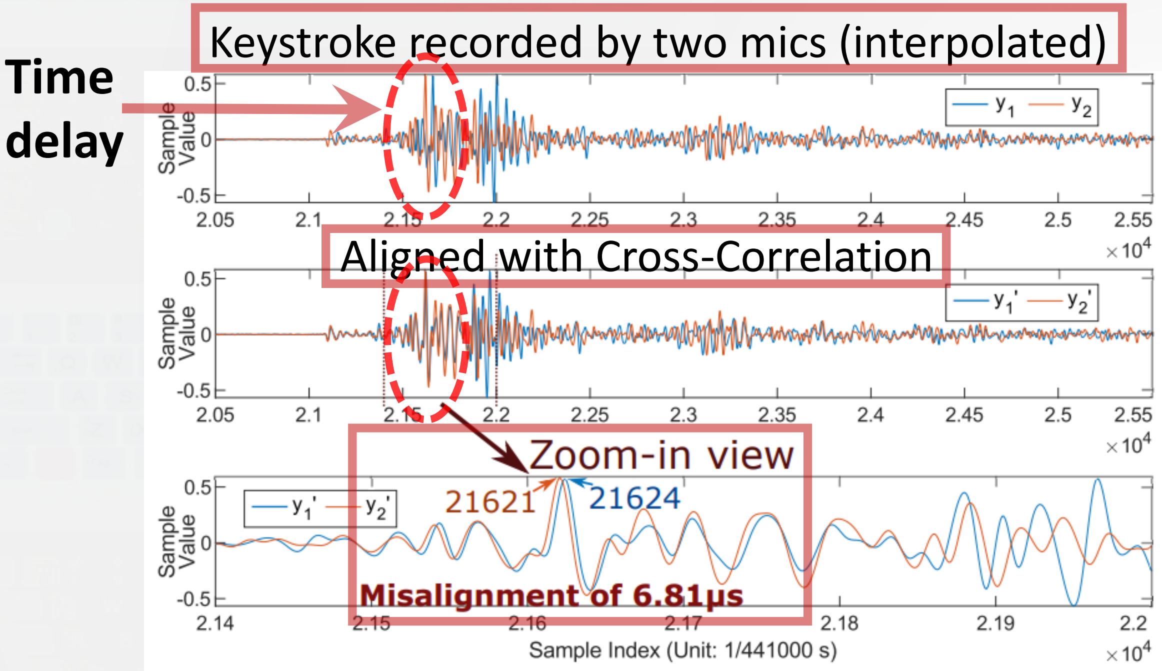
- Internal sound component and keystroke physics analysis
  - Temporal analysis, frequency-energy analysis (on **internal transient and noisy parts**)

- **Multi-round** keystroke localization with customized processing chains
  - Inspired by imperfect keystroke sound and measuring physics
  - Interpolation, align and recalculation (within keystrokes to μs-range)

- Unconstrained keyboard inputs (with unknown sequences and non-alphabetic keys)

Perceiving the (Im)precision

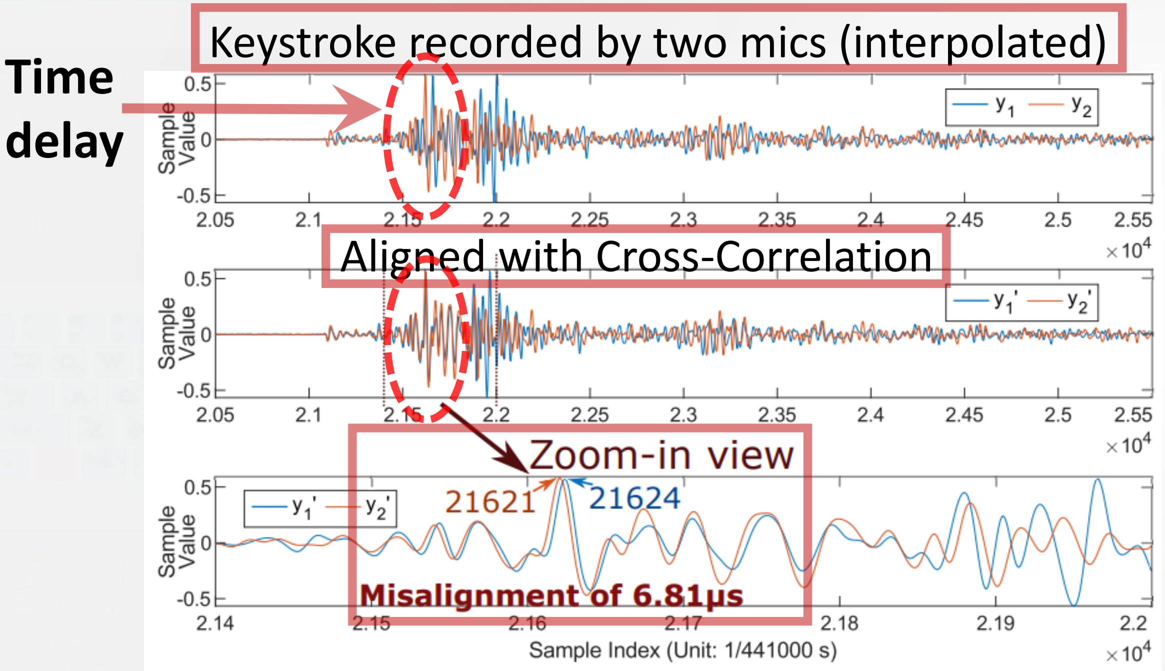
Time-domain Analysis



It is challenging to mitigate localization **errors in the range of several to tens of µs**

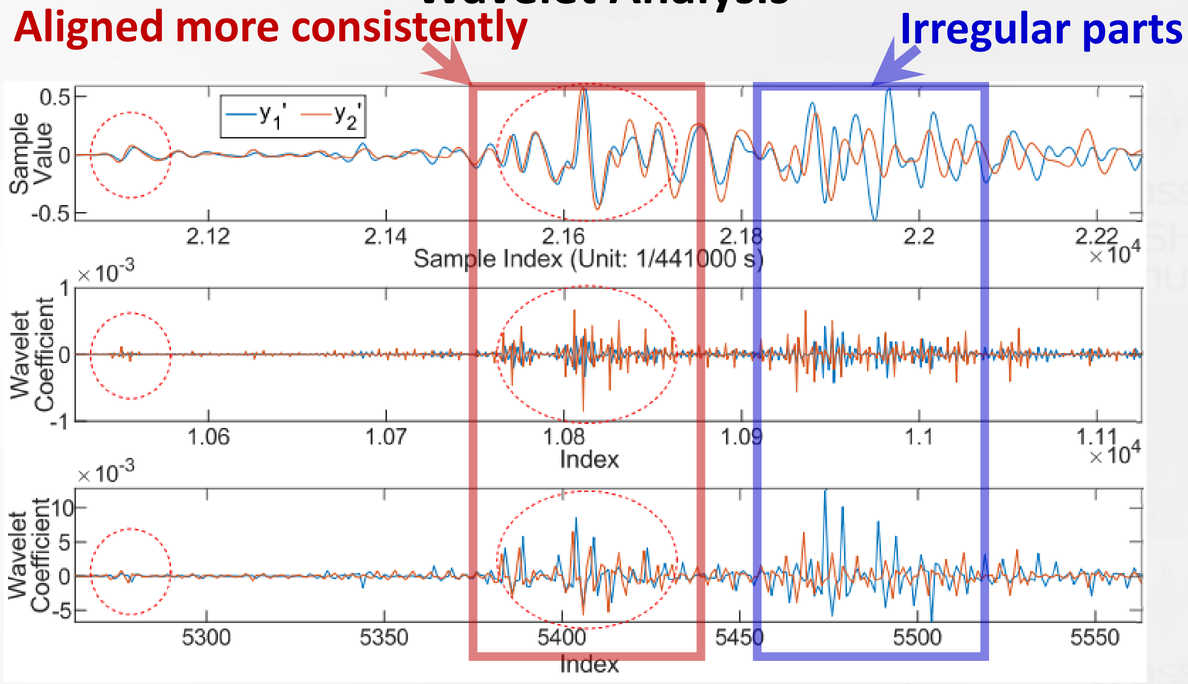
Perceiving the (Im)precision

Time-domain Analysis



It is challenging to mitigate localization **errors in the range of several to tens of μs**

Wavelet Analysis



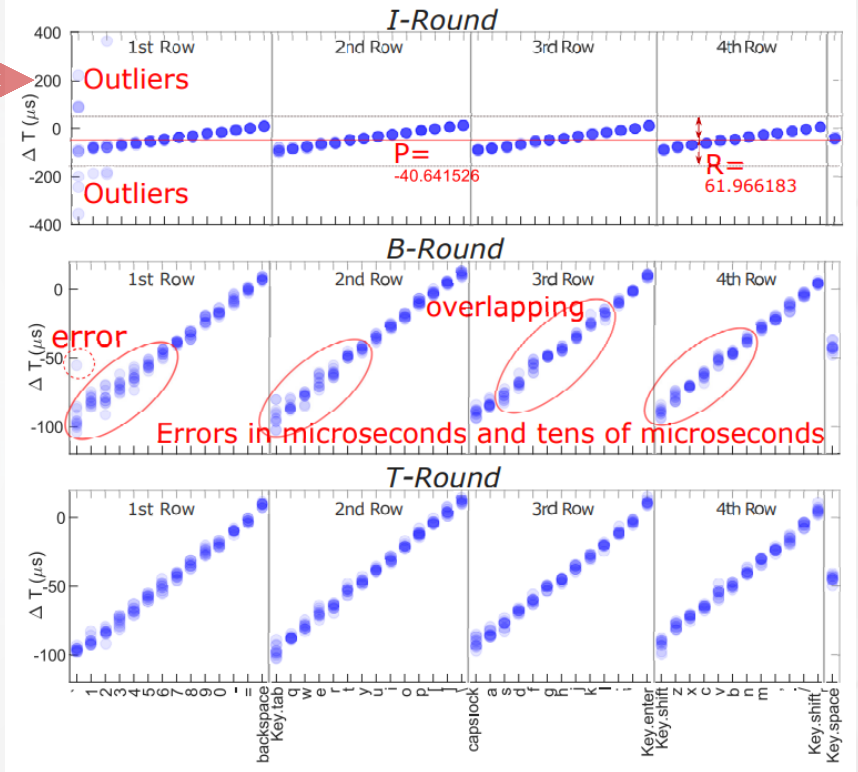
Observation: Signals in irregular parts provide coarse-grained information but can mask high-precision localization data (**self-masking**)

Multi-Round Processing

- Initial Round (I-Round)
  - Zero-phase Butterworth filter
  - Interpolation to μs range
    - 44.1 kHz recording sample rate
    - 1,761 kHz interpolation (Unit: 0.5686 µs)
  - Cross-Correlation

I-Round results include large-scale errors (outliers)

1st row 2nd row 3rd row 4th row



Multi-Round µs-Precision Approach

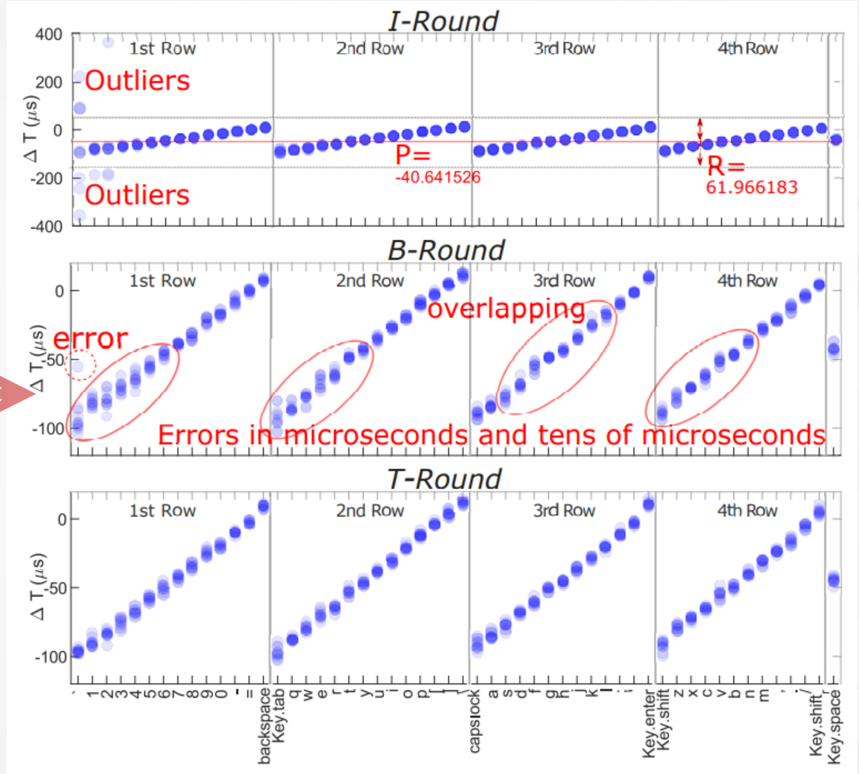
P: mean of all non-outlier measurements
R: half of the difference between max and min non-outlier measurements

Multi-Round Processing

- Initial Round (I-Round)
- Bounding the Range (B-Round)
  - Outlier identification
  - Align and recalculation
    - Align based on center point P
    - Bounding the time delay range with R

B-Round results still have errors and significant overlapping

1st row 2nd row 3rd row 4th row

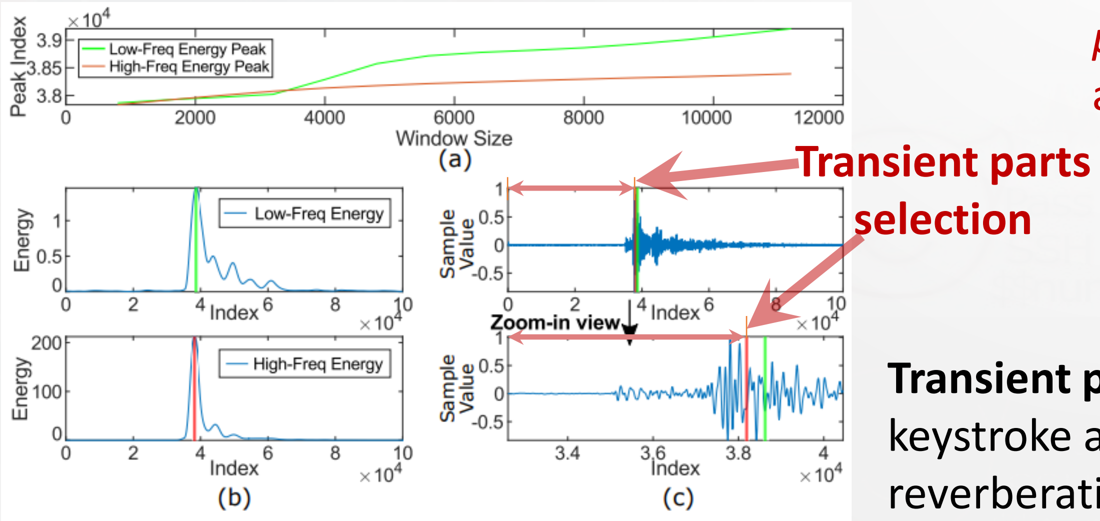


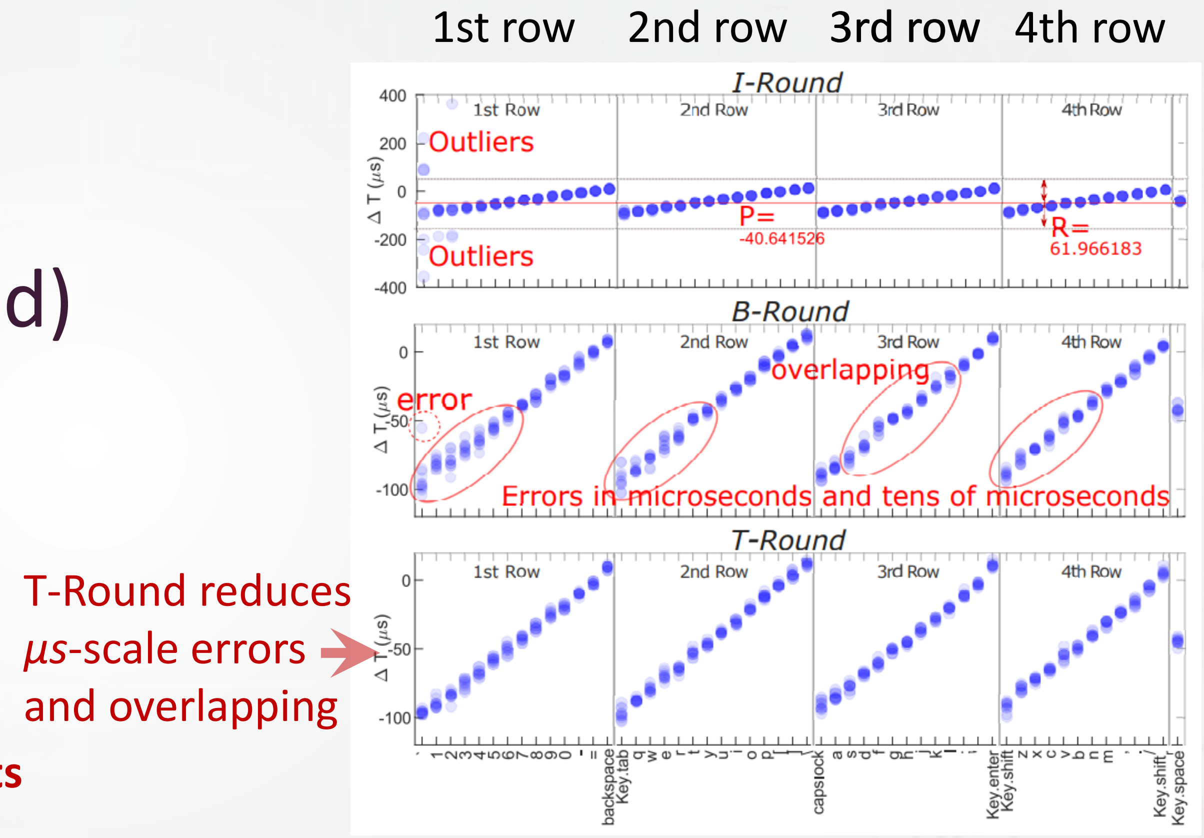
Multi-Round μs-Precision Approach

P: mean of all non-outlier measurements
R: half of the difference between max and min non-outlier measurements

Multi-Round Processing

- Initial Round (I-Round)
- Bounding the Range (B-Round)
- Focusing on Transients (T-Round)
  - Align based on B-Round results
  - Sum, Transient parts selection
  - Time delay recalculation

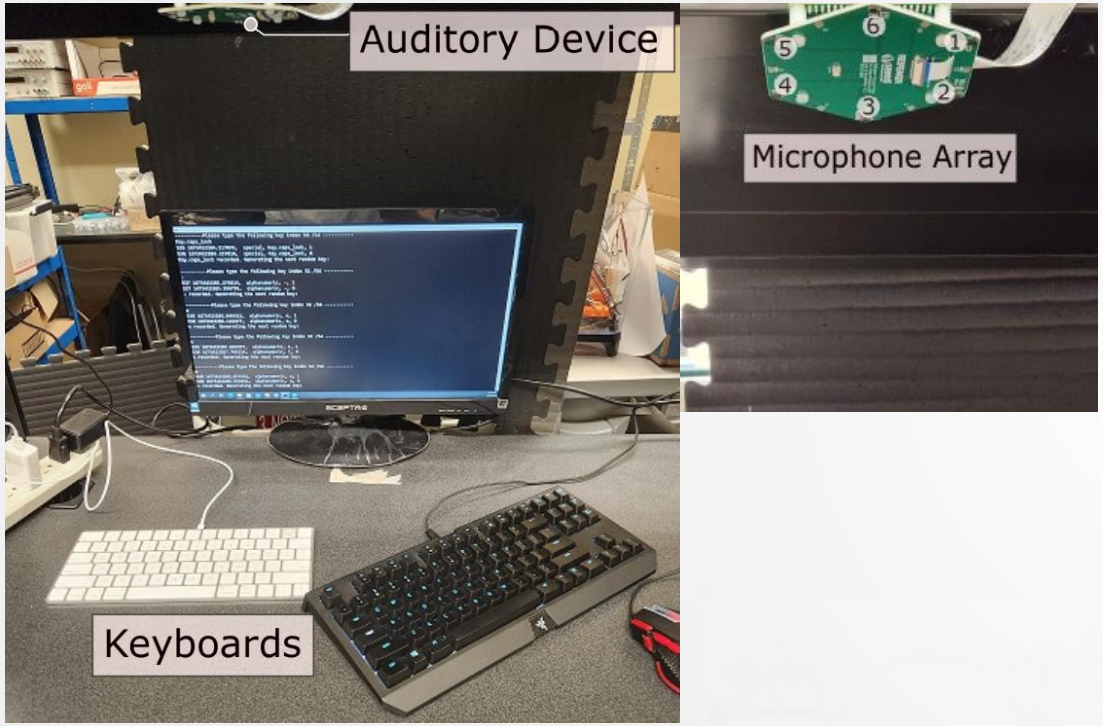




Multi-Round µs-Precision Approach

Transient parts include the short burst of energy (higher SNR) at start of keystroke and are also less susceptible to interference caused by reverberation and keyboard base vibration

Keystroke Sound Localization Results



Localization results of 595 keystrokes on Razor Blackwidow keyboard from 0.5 m

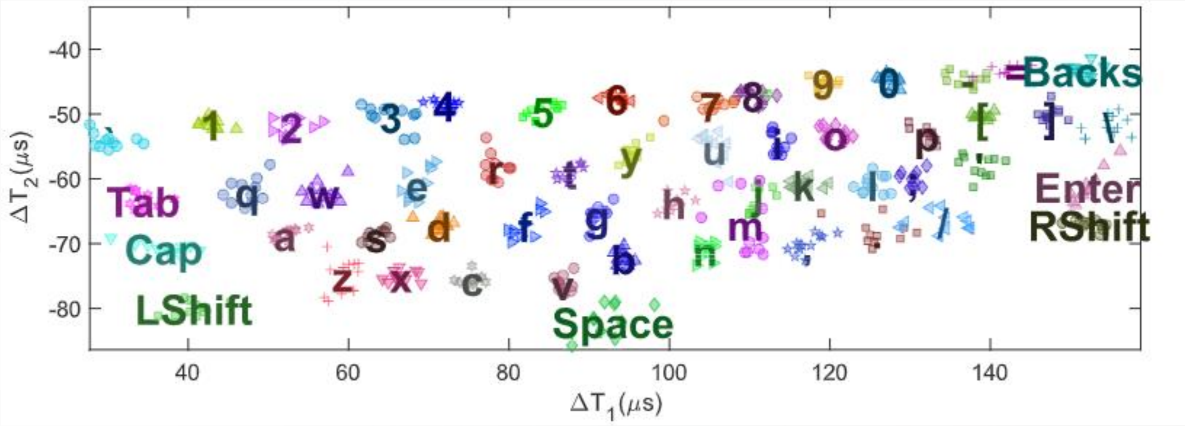


Table 2: Average standard deviation (Unit: μs)

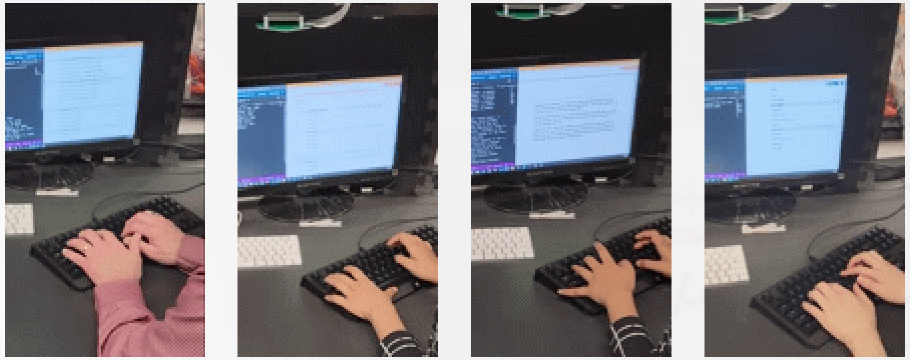
	Apple ΔT_1	Apple ΔT_2	Razer ΔT_1	Razer ΔT_2
σ	2.1339	2.0272	1.6274	1.3890

Table 1: *n*th-attempt accuracy of 594 keystrokes on an Apple keyboard and 595 keystrokes on a Razor keyboard.

Keyboard	First	Second	Third	Forth
Apple	90.64%	98.16%	99.50%	100.00%
Razor	96.47%	99.16%	99.50%	99.83%

User Study

- Different users type differently
- Same user types differently when inputting different contents
  - ID numbers, dates, addresses, GPS coordinates
  - Real-world texts with punctuation, numbers, capital letters, typos
  - Usernames and passwords
  - Strong passwords, SSH credentials

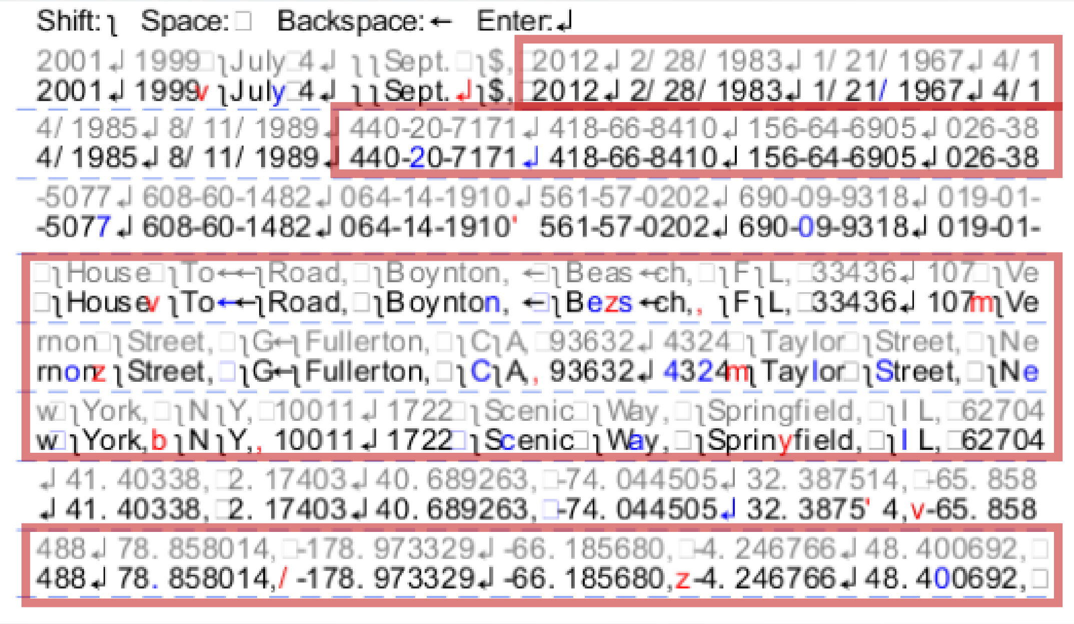


Natural typing styles (touch typing)
Can adjust typing styles/speeds

Attack Accuracy and Total Keystrokes

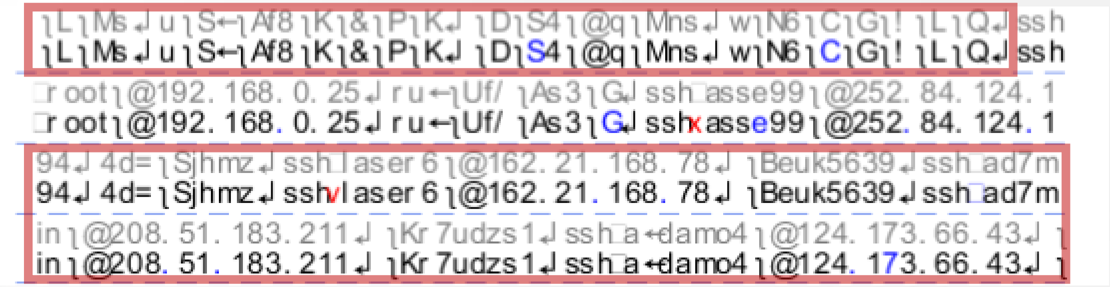
User	1st Attempt		2nd Attempt		Total
	Accuracy	Correct	Accuracy	Correct	Keystrokes
A	90.6%	2635	95.3%	2773	2909
B	83.8%	2018	92.5%	2228	2408
C	89.3%	2145	93.8%	2253	2402

Recovering Sensitive Information



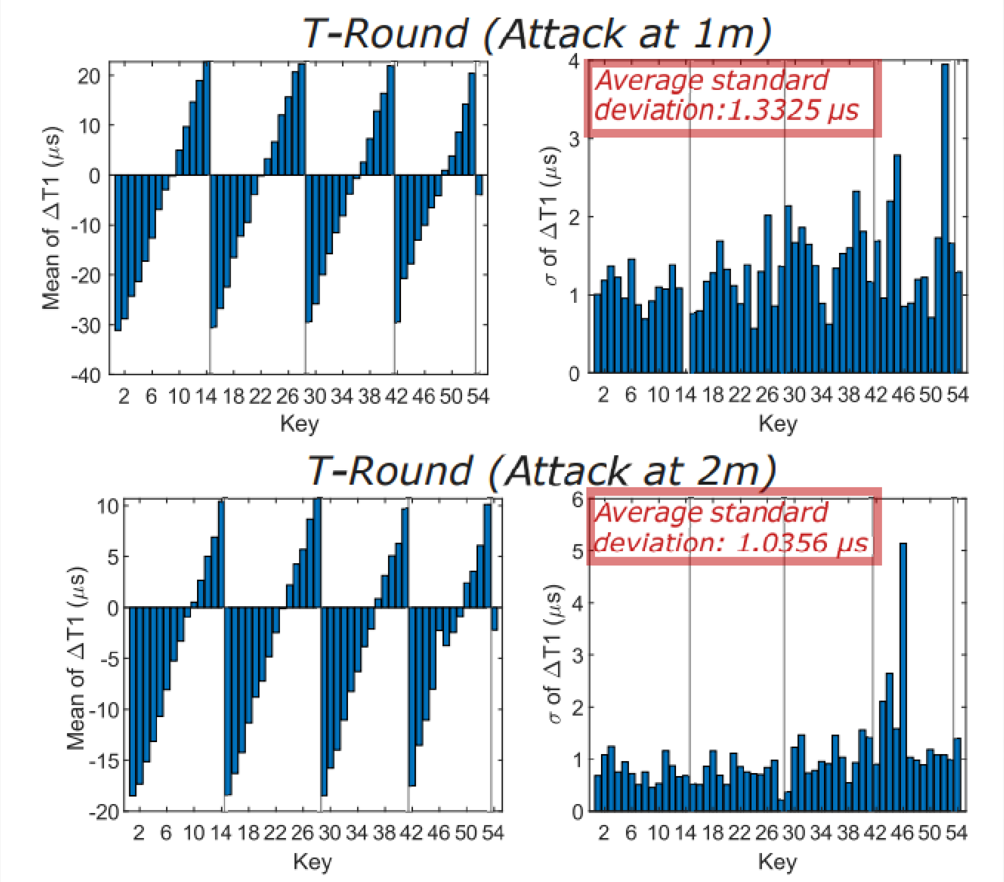
Recovered dates, SSN numbers, addresses, GPS coordinates, etc.

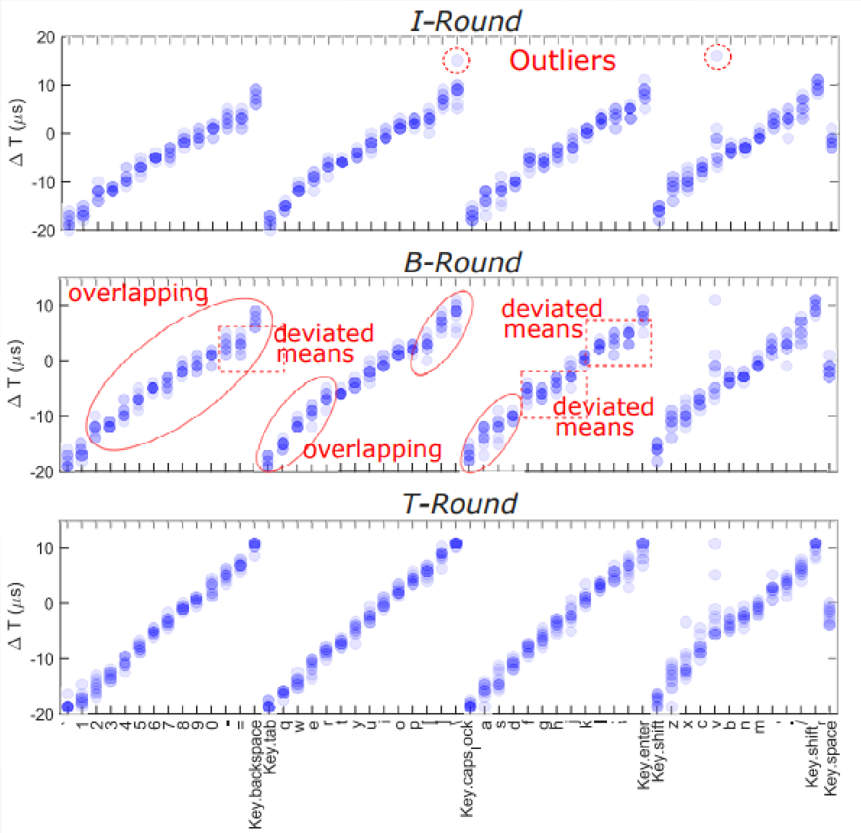
Gray: User Input (Ground Truth)
Black: Attack Result (1st Attempt)
Blue: Attack Result (2nd Attempt)
Red: Error



Recovered passwords and SSH credentials

Distance





The range of the time delay has become very small ([−19,11] μs) at **2-m attack distance**

NLOS Attacks: Covert Typing

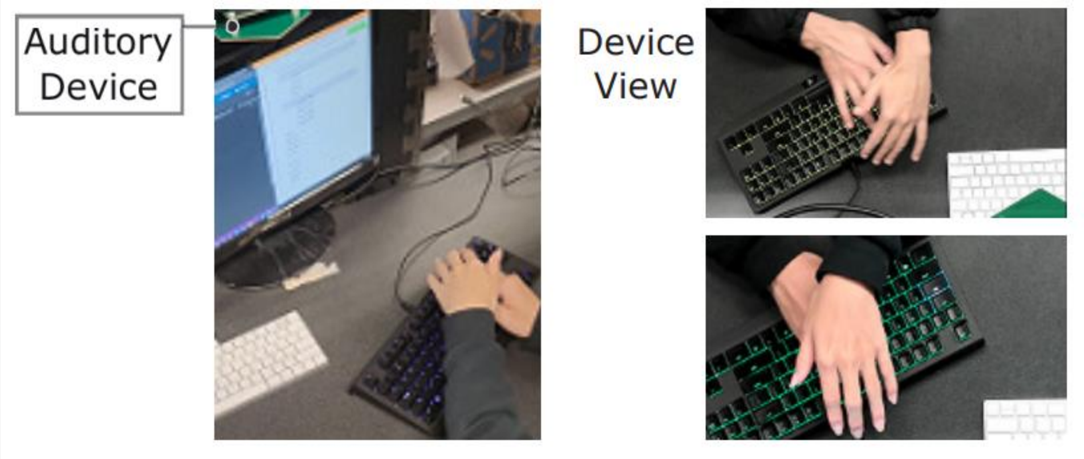
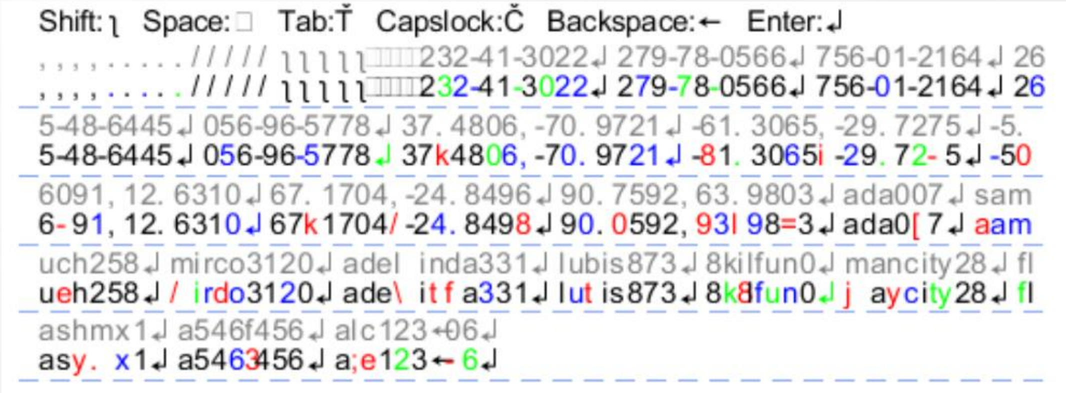


Table 4: *n*th-attempt accuracy, correctly identified keys, and total number of keystrokes of covert user inputs.

User	1st Attempt Accu.	1st Attempt Corr.	2nd Attempt Accu.	2nd Attempt Corr.	3rd Attempt Accu.	3rd Attempt Corr.	Total Keystrokes
N1	74.3%	378	88.4%	450	93.5%	476	509
N2	56.8%	269	75.3%	357	84.4%	400	474

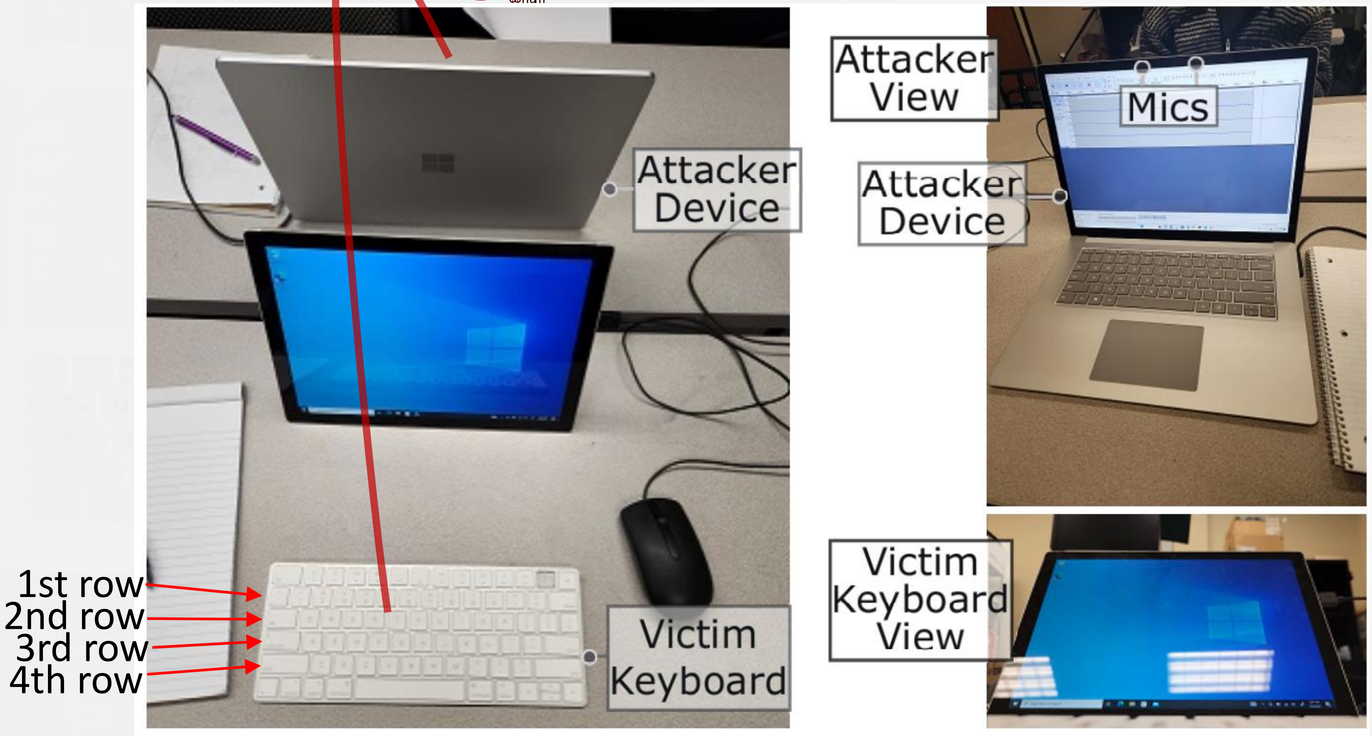
Localization information is not completely lost in refracted keystroke sounds after multi-path transmission in NLOS setting



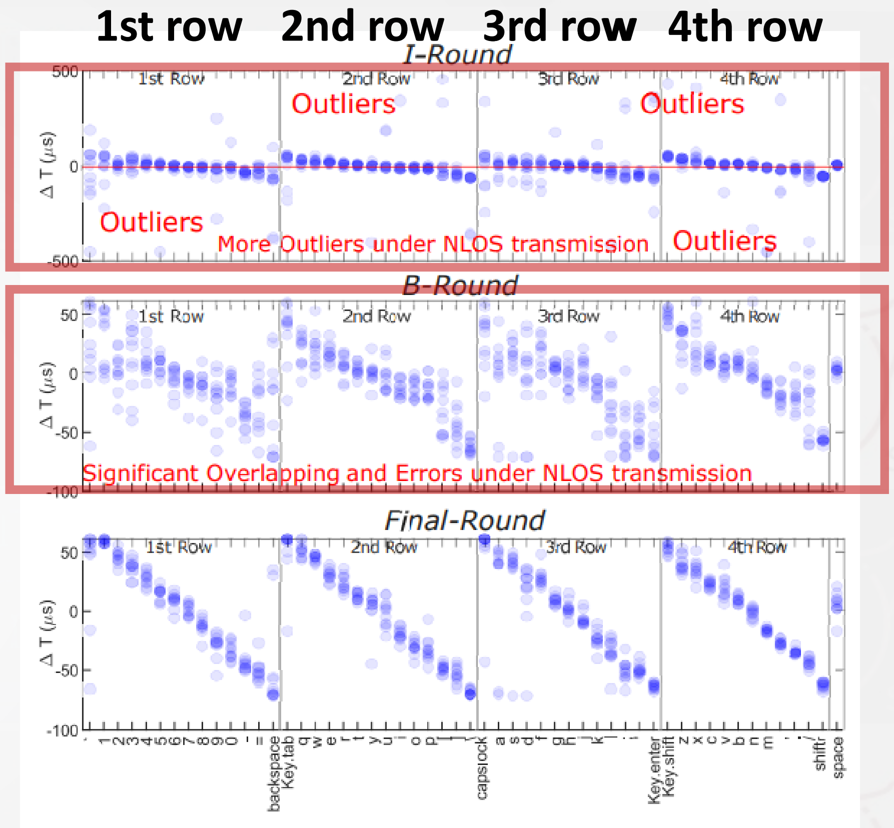
Gray: actual inputs; Black: 1st-attempt results; Blue: 2nd-attempt results; Green: 3rd-attempt results.

NLOS Laptop-Based Attacks

Bendable Eyesight



The attack can be launched without pointing any sensors toward the victim's keyboard



Our multi-round approach effectively reduces the excessive errors caused by NLOS keystroke sound transmissions

N-th Attempt Accuracy

1st	2nd	3rd	4th	5th	6th	7th	8th	9th
21.96	42.10	54.91	68.05	75.54	82.36	85.19	89.35	91.18

Conclusion

- Real-world user inputs are usually not purely alphabetic, single-letter-case keys/words
  - This work explored keyboard side-channel attack on unconstrained inputs
- Attacks using limited-resolution audio interfaces can reveal unconstrained keyboard inputs with a fairly sharp and bendable "auditory eyesight"
- Sound component and the underlying physics study allows extracting more targeted and accurate information

Conclusion

- Dataset
  - Benchmark
  - Future research and education to improve privacy awareness
- Artifact
  - https://github.com/auditoryeye/auditoryeye_artifact

GitHub Repository