UC San Diego

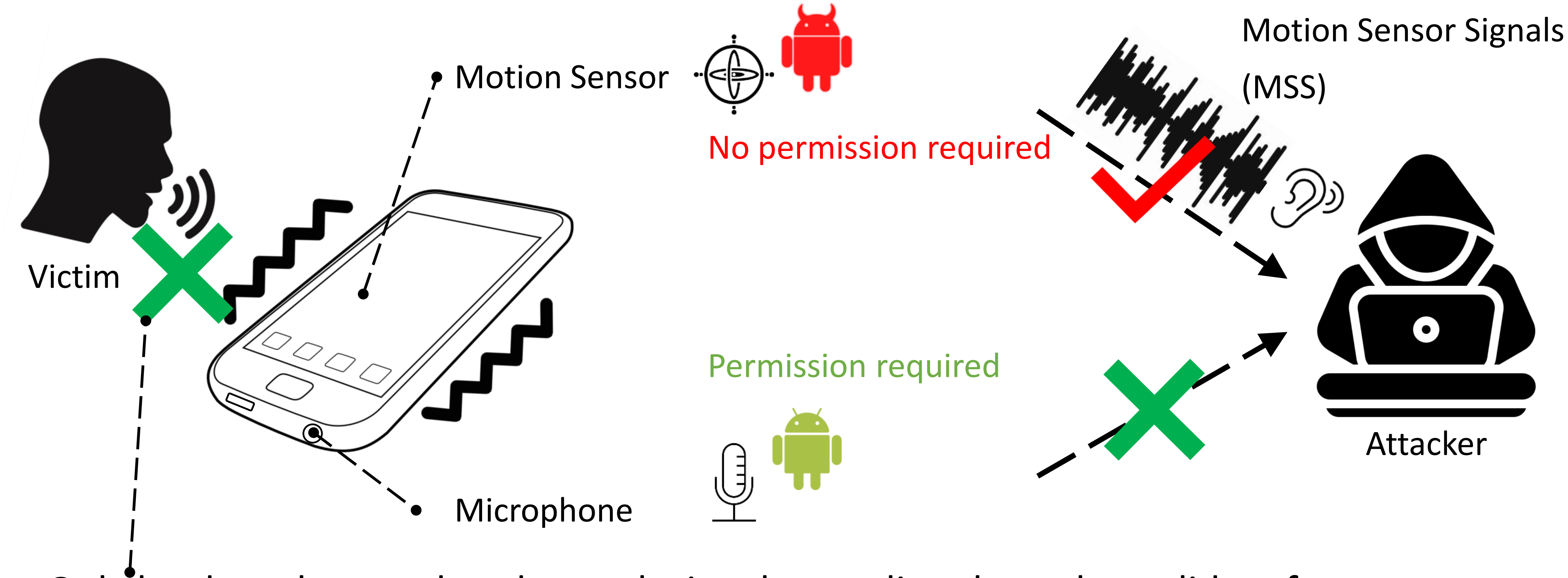
StealthyIMU: Extracting Permission-protected Private Information from Smartphone Voice Assistant using Zero-Permission Sensors

Ke Sun, Chunyu Xia, Songlin Xu, Xinyu Zhang

University of California, San Diego

NDSS’23, Mar 1st, 2023

Speech Eavesdropping on Smartphone



Only loudspeaker-rendered speech signals traveling through a solid surface can create noticeable impacts on motion sensors [1].

[1] Speechless: Analyzing the threat to speech privacy from smartphone motion sensors, S&P'18

Limitations of Prior works

Motion Sensor:

- Low sampling rate (<= 500 Hz)
- Low Signal-to-Noise Ratio (SNR)
- Additional interference

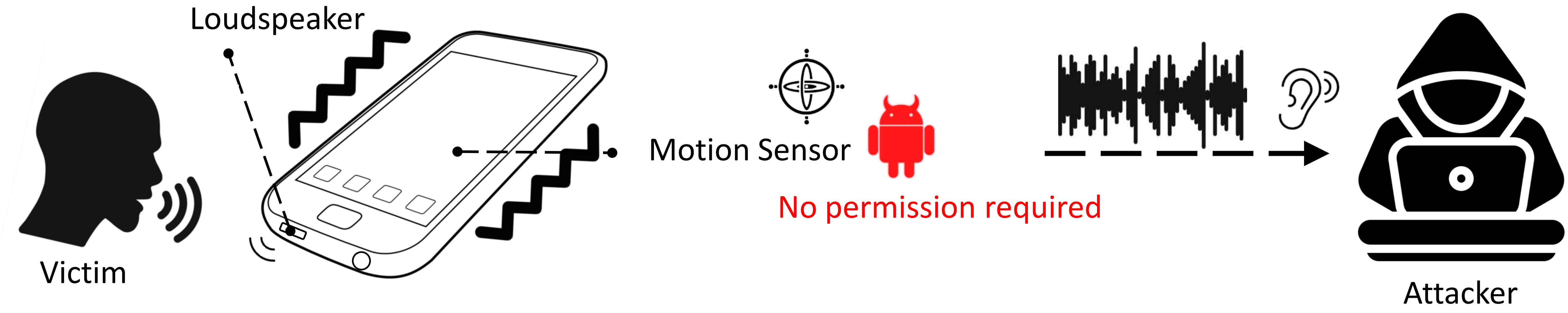
Achieve *low-risk* task

- Classifying a small set of digits and hot words [2,3]
- Partially recover the speech signals [4]

In *ideal scenario*

- Small number of users in the dataset (<20 users) [2][3][4]
- Training and testing with the same group of users [2][3][4]

[2] Learning-based practical smartphone eavesdropping with built-in accelerometer, NDSS’20
[3] Spearphone: a lightweight speech privacy exploit via accelerometer-sensed reverberations from Smartphone loudspeakers, WiSec’21
[4] AccEar: Accelerometer Acoustic Eavesdropping with Unconstrained Vocabulary, S&P’22



Will this threat model pose a real privacy threat to victims?

Extracting Permission-protected Private Information from Smartphone Voice-User Interface (VUI) responses.

StealthyIMU Threat Analysis

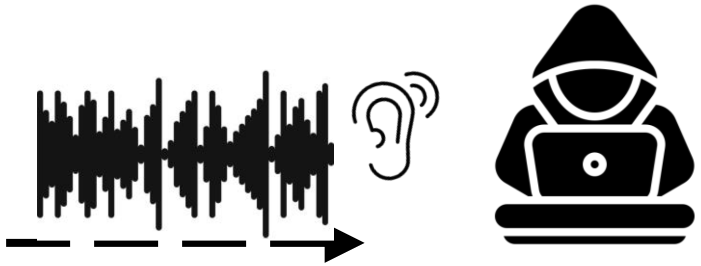
Reading permissions granted by VUI app

- Calendar
- Contacts
- Locations
- Search history
- Alarm
- Reminder
- Phone
- SMS
- Voicemail
- Billing
- etc

Voice Assistant responses contain the permission-protected private information.

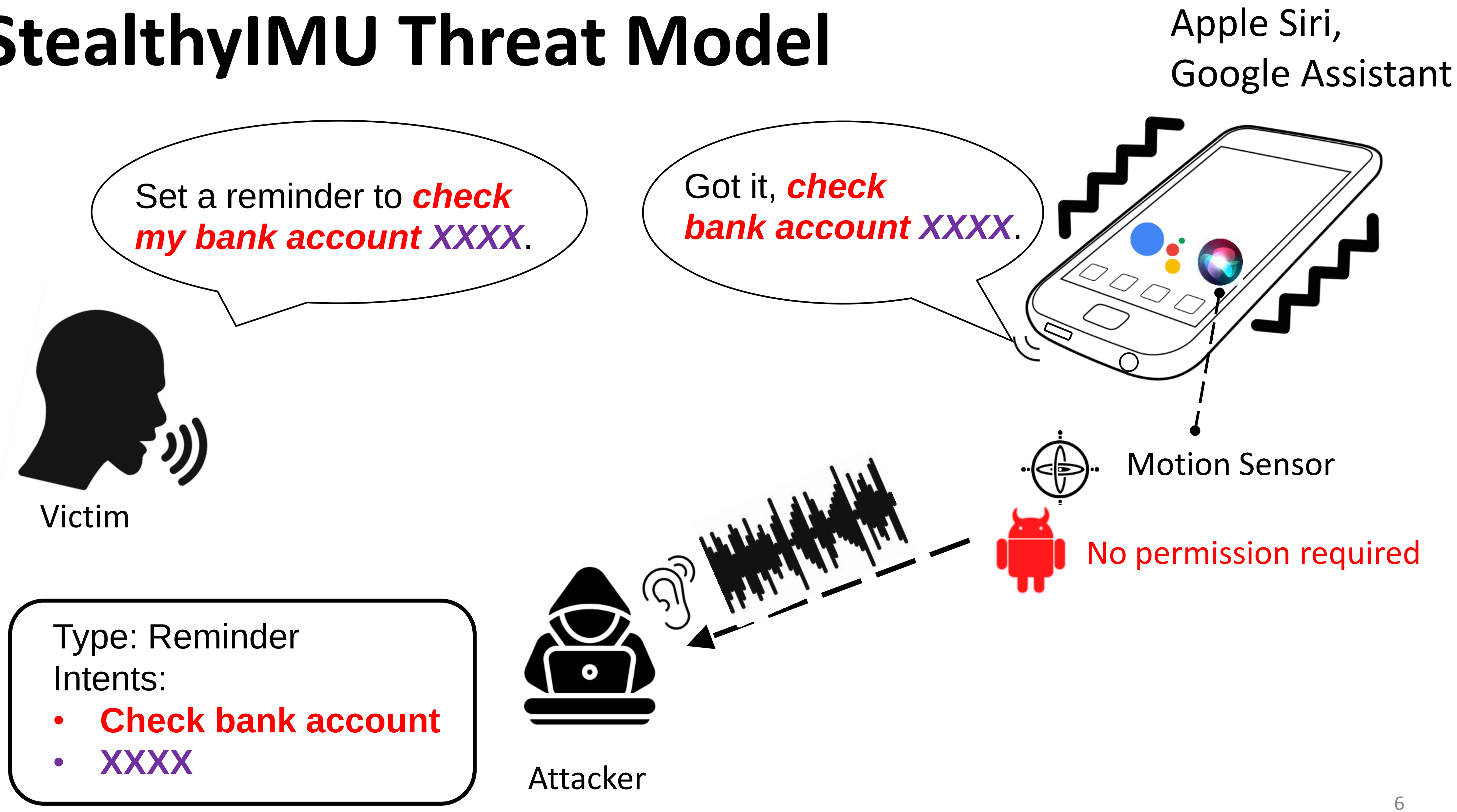
Motion Sensor

No permission required



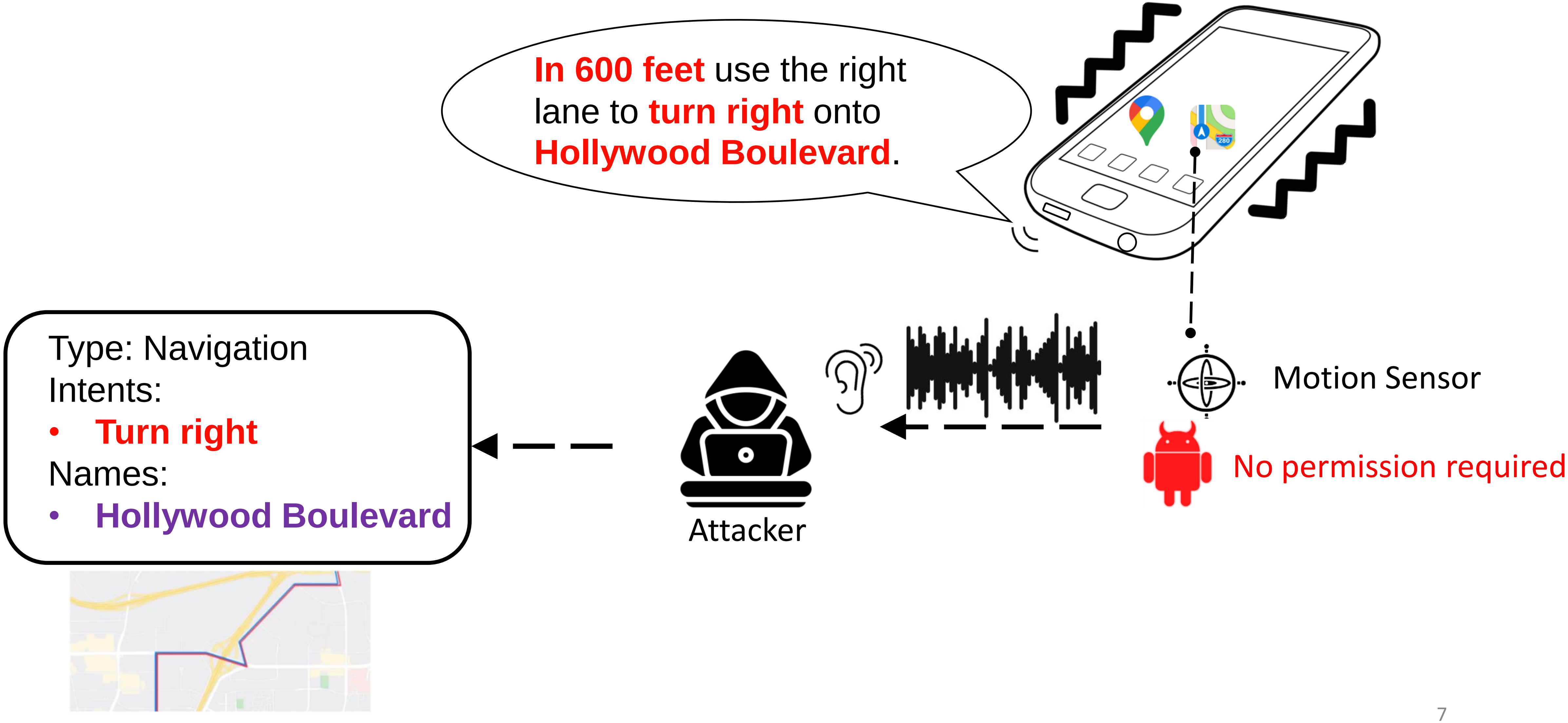
Attacker

StealthyIMU Threat Model



Type: Reminder
Intents:
- **Check bank account**
- XXXX

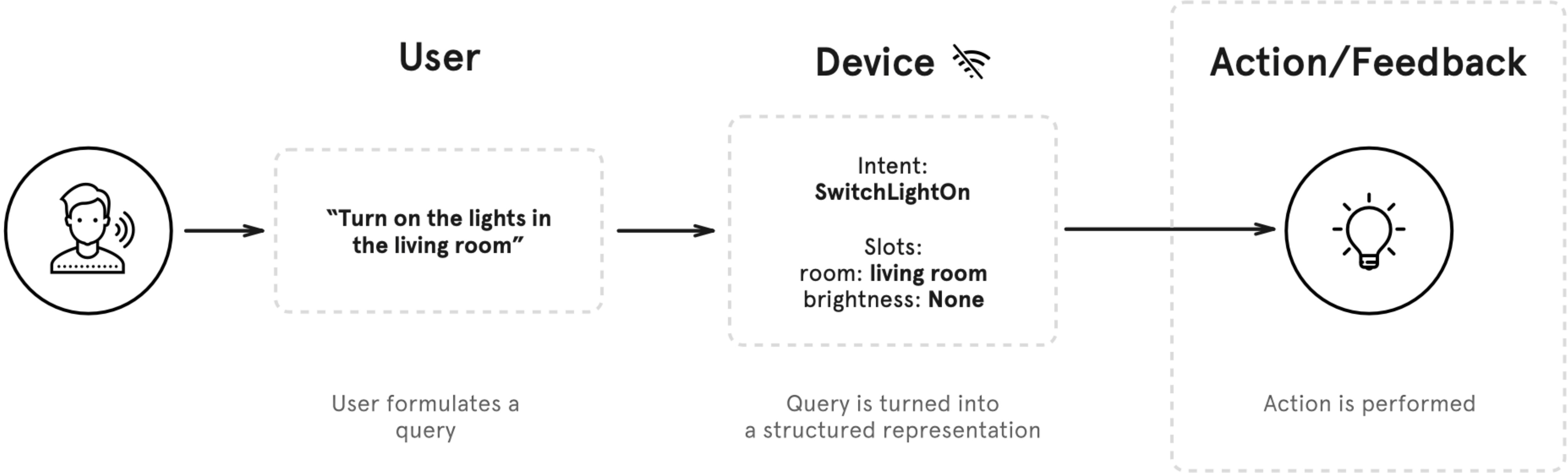
StealthyIMU Threat Model



GPS Trace

Inferring privacy from a single VUI response: problem statement

Traditional Spoken Language Understanding (SLU) system for VUI



[5] Snips Voice Platform: an embedded Spoken Language Understanding system for private-by-design voice interfaces

Comparison between SLU for VUI and StealthyIMU

	SLU for VUI	SLU for StealthyIMU
Goal	Extract the *request* from speech	Extract the *privacy* from VUI response induced MSS
Captured Signals	Microphone ***High sampling rate*** ***High speech quality***	Motion Sensor **Low sampling rate** **Low speech quality**
Type of Voices	Human subjects: ***Arbitrary* # of voices profiles**	Machine-rendered: ***Limited* # of voices profiles**
Type of Speech	VUI requests: ***Arbitrary* format**	VUI responses: **More *deterministic* format**

Attacking Requirements and Targets

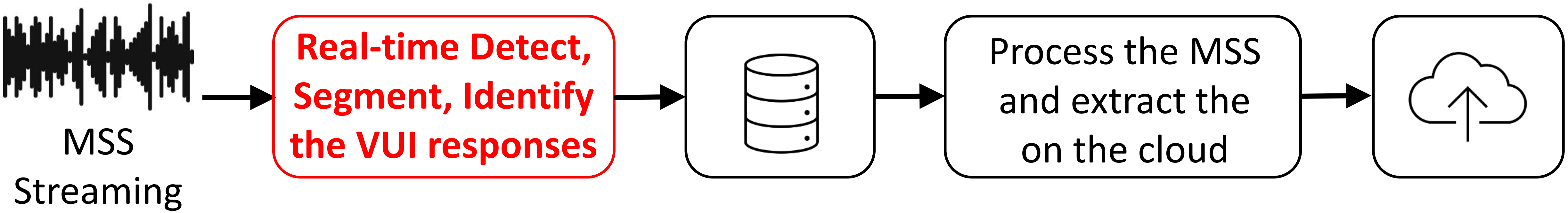
- ***Affordable*** Attack
- **High** attacking success rate
- **Explicit** permission-protected private information extraction

Targeted permissions:

- Read calendar
- Read contacts, SMS
- Read search history
- Access coarse location
- Access GPS trace
- etc.

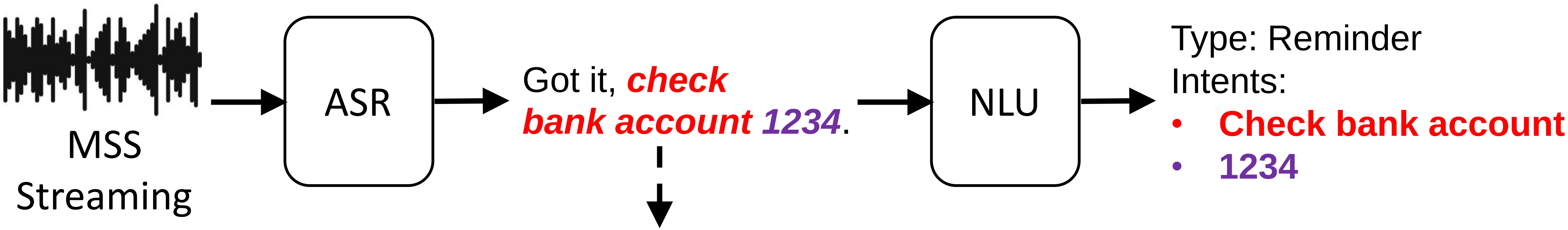
Challenge 1: how to achieve *affordable* attack?

Neither the **smartphone OS** nor the **user** can notice the attack.



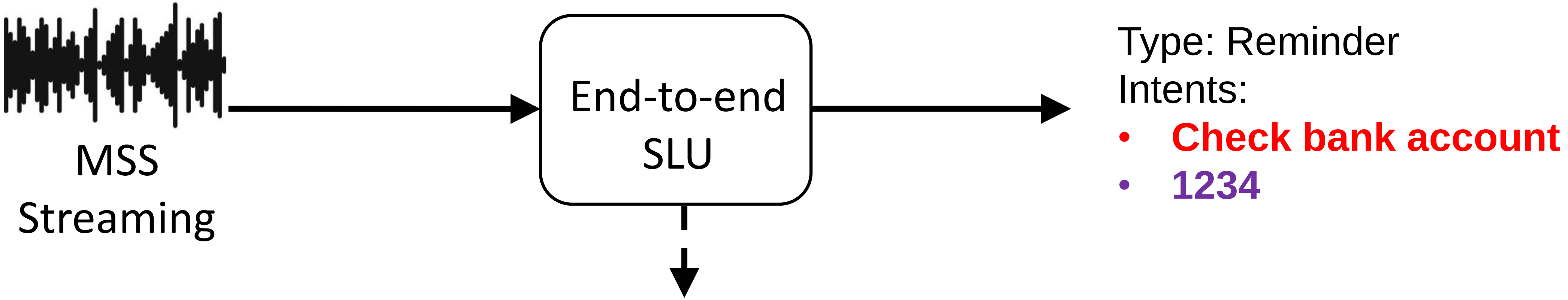
- Low on-device computational cost
  - Real-time two-stage ***detection*** algorithm → < 5% Peak CPU Usage
  - Lightweight ***voice identification*** DNN models. → 79.6 KB model size
- Low on-device storage cost
  - Only save the potential MSS → 16 kB/ response
- Low communication requirement
  - Only upload the private information → less than 50 bytes/ response

Challenge 2: Inferring Privacy from a Single VUI response → SLU model design



However, due to the *low sampling rate* and *low speech quality*, the first stage, i.e., ASR, can only achieve low accuracy.

End-to-end SLU Model



- The format of VUI response is deterministic.
- Only need to extract the private information while ignoring other information.

VUI Commands Generation

44,691 different VUI commands
23 types of VUI commands

Analysis the format and structures of VUI response, and then manually label them (***1,000 hours to label 100 hours data***).

Type	Example Voice Command	Privacy	#
Weather	What's the weather today?	Location	12,527
Sun set&rise	What's the sunset in Chennai	Location	1,505
AirCheck	AQI for San Francisco	Location	1,601
Clock	What time is it in London	Location	1,595
Reminders	Set a reminder to check my account	Todo	2,950
	Set a reminder for tomorrow morning	Time	2,140
Media Alarms	Set an alarm to go to fedex	Todo	2,630
	Set a music alarm at 8 PM	Time	2,350
Stock Updates	Stock price for Apple	Search	1,318
Calling	Call Sam	Contacts	1,120
Navigation	Navigate to Los Angeles	Location	1,570
Navigation App	/	GPS	7,885
Fun Tricks	What movies are playing?	Others	500
Sports Facts	What's the news about the NFL?	Others	500
News	What's the news about the covid?	Others	500
Calculations	What calculation can you do?	Others	500
Google Search	How tall is the Eiffel Tower?	Others	500
Youtube Music	Play music on Youtube Music	Others	500
Voice Mail	Call voicemail	Others	500
Youtube	Open Youtube	Others	500
Chrome	Open the Google Trends website	Others	500
Youtube TV	Play FS1 on Youtube TV	Others	500
Broadcast	Broadcast a message	Others	500
Overall			*44,691*

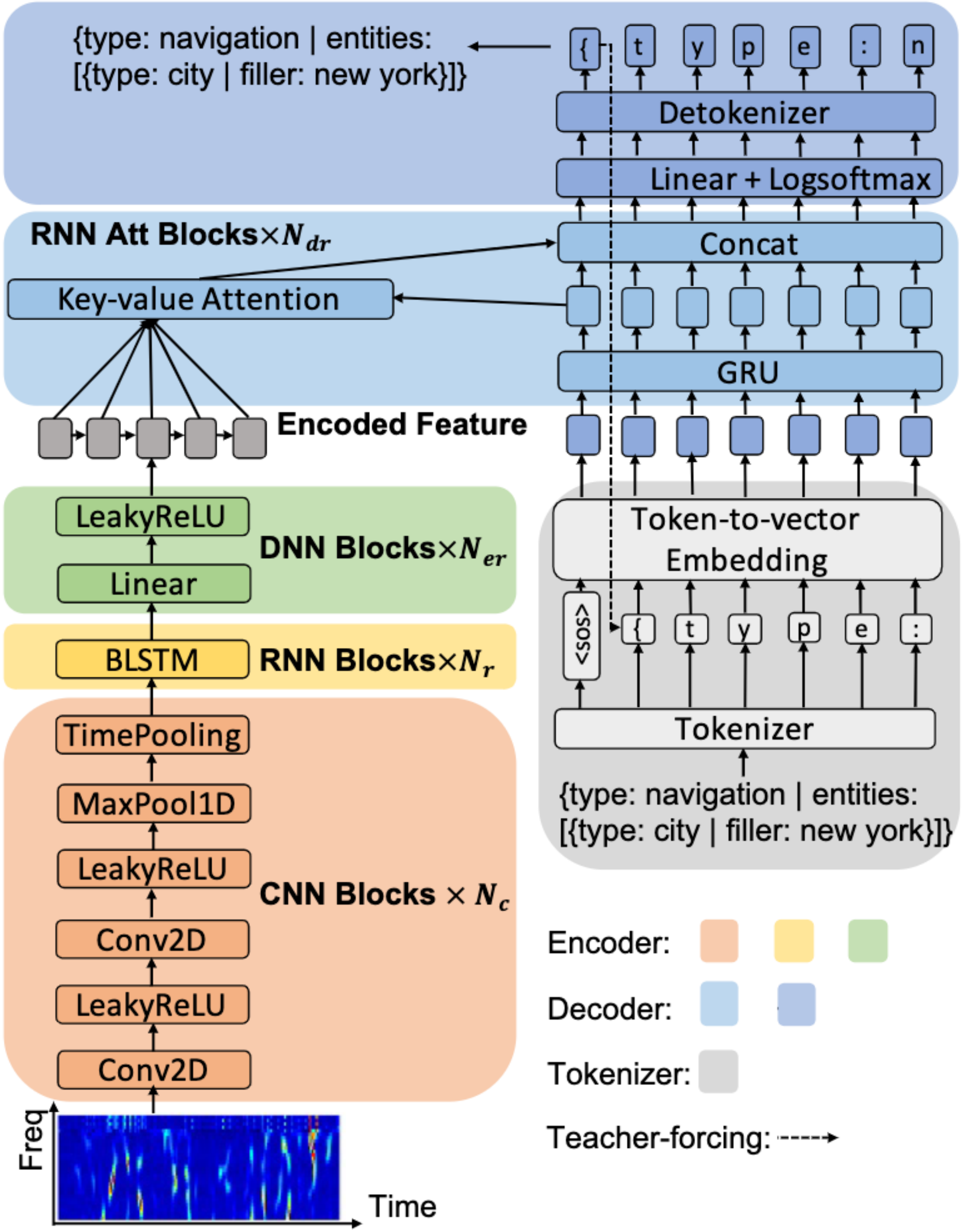
Open-sourced dataset and code: https://github.com/Samsonsjarkal/StealthyIMU

SLU Model Design

End-to-end SLU model

- Seq-to-seq model
- Encoder + Decoder design
- Sub-word level Tokenizer
- Teacher-forcing

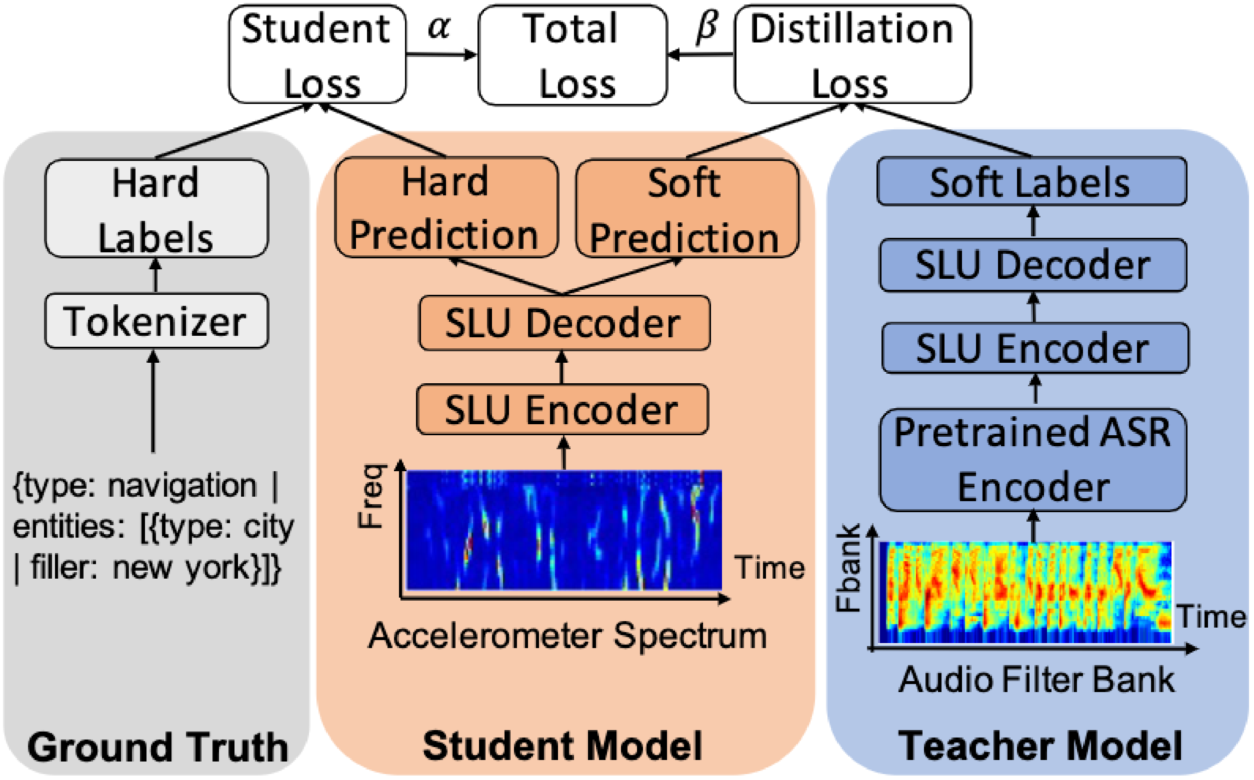
How to improve the performance through crossmodal data?



Knowledge Distillation from Speech-based SLU

Speech-based SLU achieves significantly better performance than the MSS-based.

Use the knowledge from the Speech-based SLU to help the training of the MSS-based SLU.



Evaluation: VUI Response Private Entity Recognition

Evaluation metrics:

- TER: Type Error Rate
- SEER: Single Entity Error Rate
- SER: Sentence Error Rate → only if all the entities in a single VUI response are recognized correctly

	Model Size	TER	SEER	SER
ASR+NLU	26.5 MB	0%	46.45%	77.91%
SLU	3.8 MB	0%	14.76%	25.16%
SLU+KD	3.8 MB	0%	8.46%	14.45%

Lower TER, SEER, SER means better performance.

- SLU model significantly outperforms the traditional ASR + NLU solution.
- Knowledge distillation from speech signals can help

Challenge 3: Extract the Explicit Permission-protected Privacy

Combining the private intents from single or multiple VUI responses.

- One-time Stealing
- Short-term Contextual Inference
- Long-term Monitoring

One-time Stealing

Take ***a single VUI response*** as the input to extract privacy information.

- An average 85.28% success rate
- Voice commands like
  - Reminders
  - Media Alarm
  - Hands Free Calling
  - Navigation App

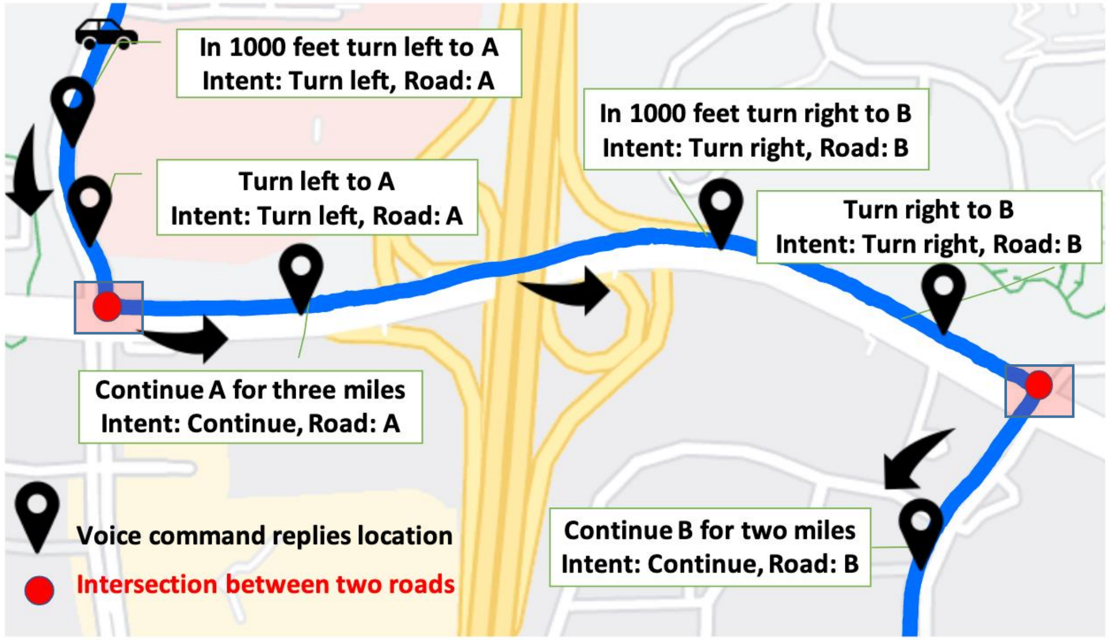
  has complicated response formats resulting in relatively low success rate (> 70%).

Type	Private Entity	TER	SEER	SER
Weather	Location	0.00%	3.05%	5.38%
Sunset & Sunrise	Location	0.00%	7.14%	13.97%
AirCheck	Location	0.00%	1.49%	3.33%
Clock	Location	0.00%	2.13%	3.18%
Reminders	Todo	0.00%	7.36%	13.21%
	Time	0.00%	15.25%	29.94%
Media Alarms	Todo	0.00%	8.24%	15.29%
	Time	0.00%	14.11%	26.50%
Stock Updates	Search	0.00%	7.33%	11.54%
Hands Free Calling	Contacts	0.00%	12.18%	22.64%
Navigation	Location	0.00%	2.19%	3.89%
Navigation App	GPS	0.00%	16.2%	26.79%
Others	/	0.31%	0.31%	0.31%
Overall		*0.00%*	*8.06%*	*14.45%*

Short-term Contextual Inference

Infer private information using ***multiple consecutive VUI responses***.

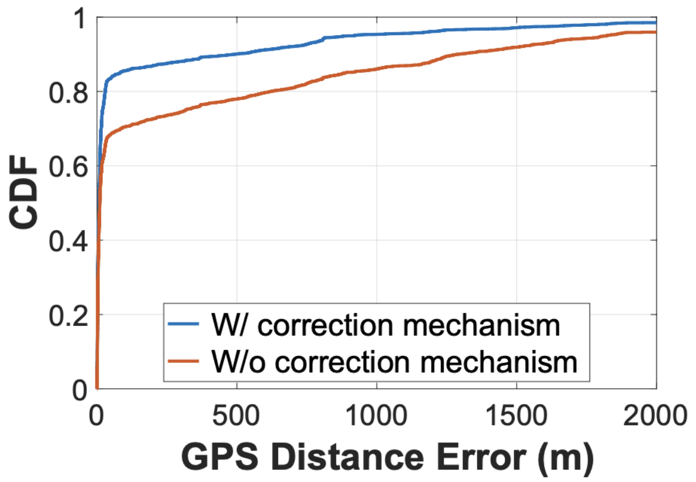
Example: Navigation app to recover GPS trace

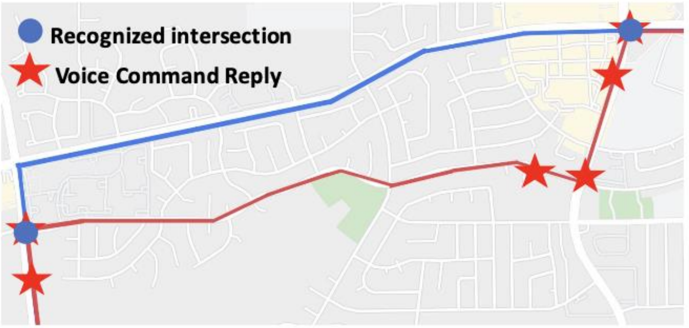


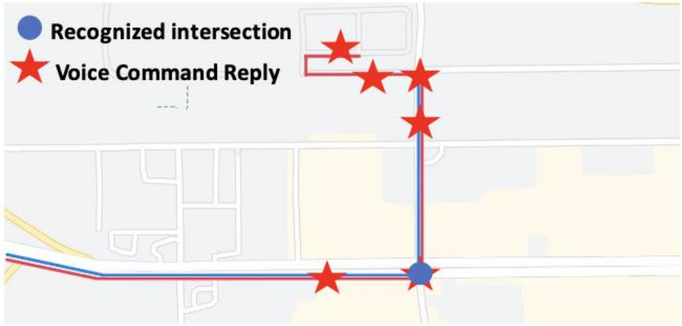
Short-term Contextual Inference

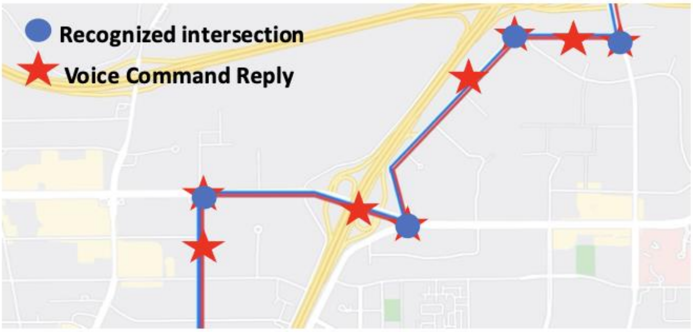
GPS trace recovery algorithm:

average/max deviation is 133 m/420 m









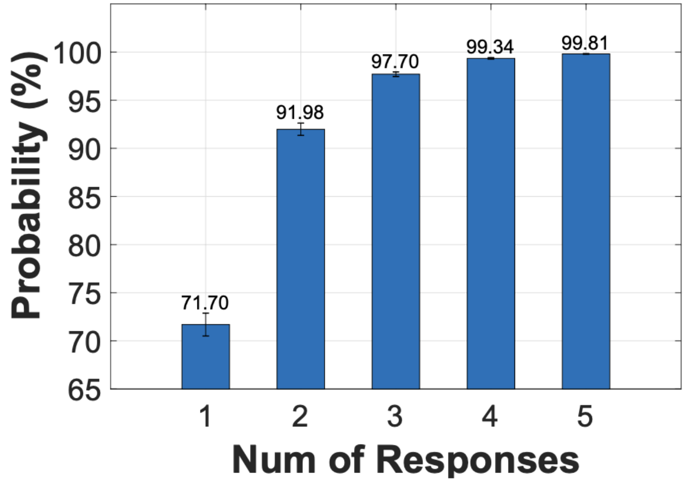
Long-Term Monitoring

Repeat the same type of voice commands in a few days, like check weather, air quality, reminder, and navigate home, etc.

Assumption: each VUI response is a single individual event.

Extract the city name from daily weather VUI response.

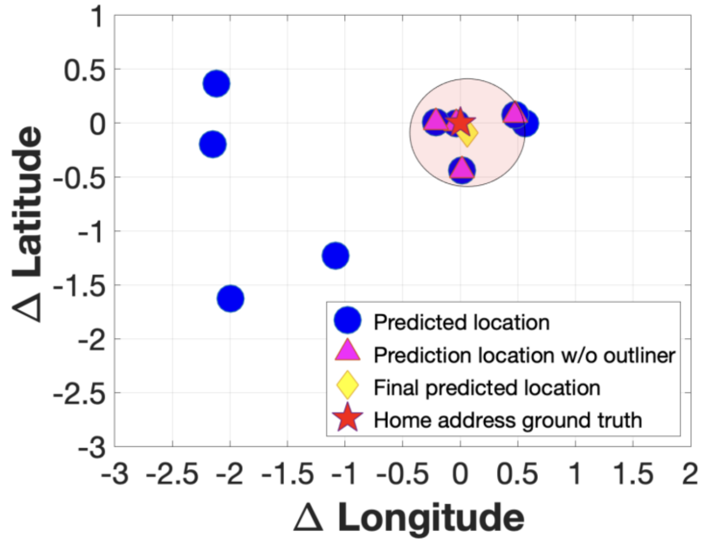
Identify the city name increases from 70% (1 inquiry) to above 98% (3+ inquiries).



Long-term Monitoring

Extract the home address from daily navigation back to home.

Achieves 11 meters home address estimation error by combining 10 attempts.



Conclusion

Uses zero-permission motion sensors to extract permission-protected private information from VUI responses.

- **Affordable** attack vector.
- Formulate it to **SLU** problem, use **cross-modal knowledge distillation strategy** to extract the private entities.
- **Short-term and long-term attack** to steal user calendar, search history, GPS trace, home address, etc.
- **Speech pre-distortion** defense mechanism.

StealthyIMU will be general to use other side channels, like RF, light, etc, to steal the private information from other VUI devices.

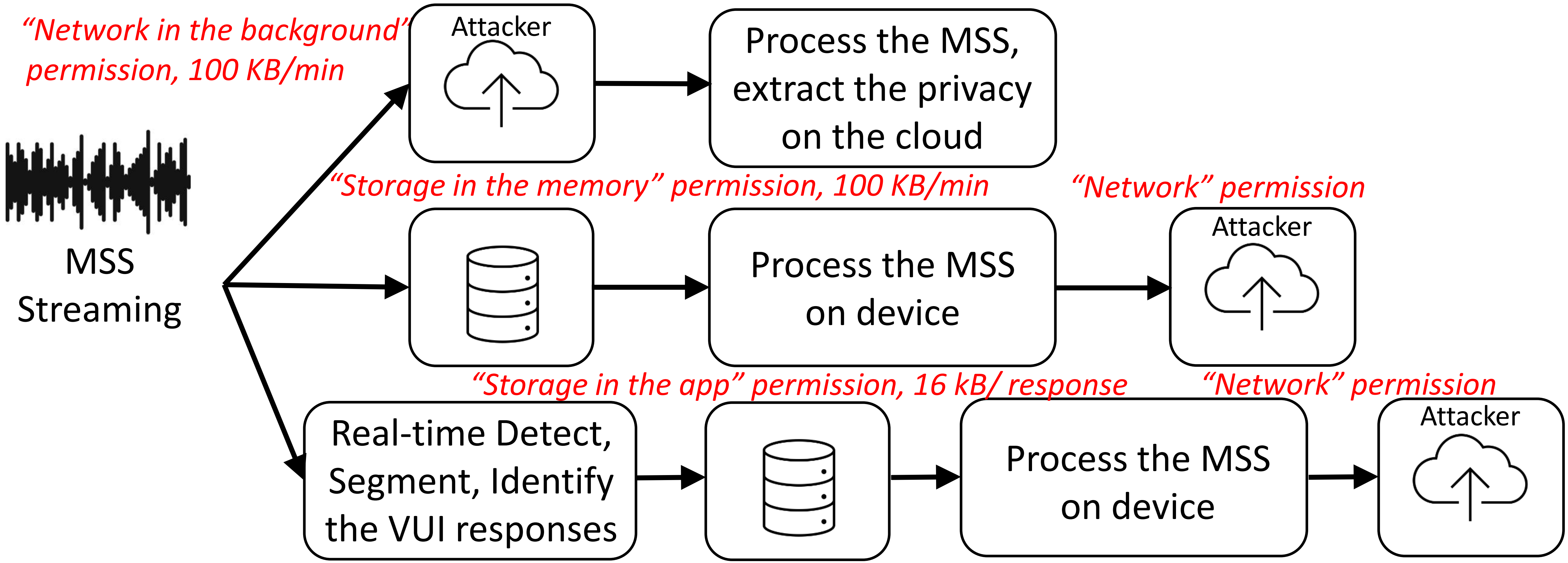
UC San Diego

Thank you

Q&A

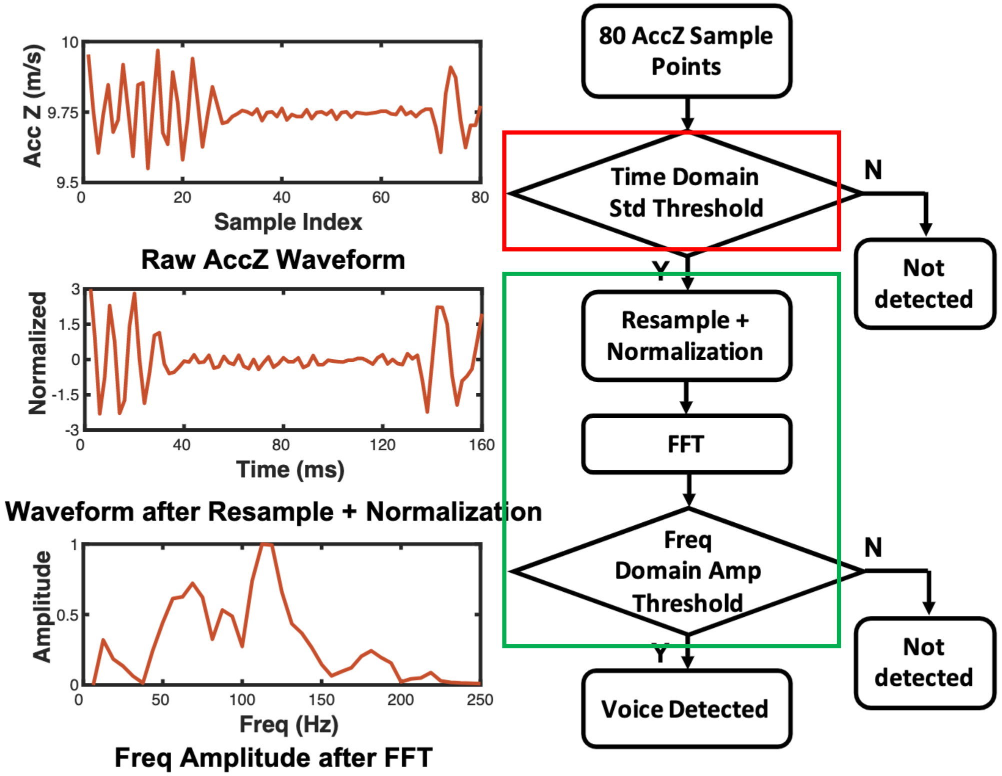
https://github.com/Samsonsjarkal/StealthyIMU

StealthyIMU Attacking Pipeline



Real-time Detect, Segment VUI Responses

- Two-stage detection
  - Stage 1: Lightweight detection
    - 15 mW, 2% CPU
  - Stage 2: Resample + FFT + detection
    - 35 mW, 5% CPU
- Segmentation
  - 2~8 s duration for each VUI response
  - 16 KB per VUI response

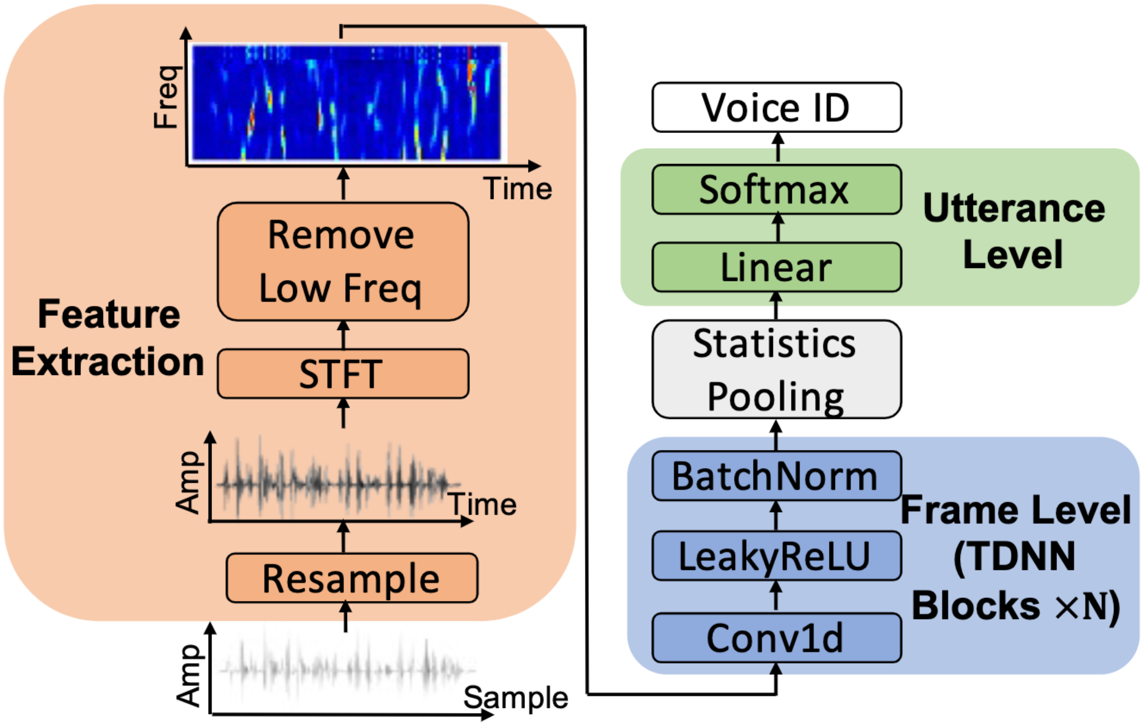


Offline Identify VUI Responses

Motivation: Identify the VUI response-induced MSS.

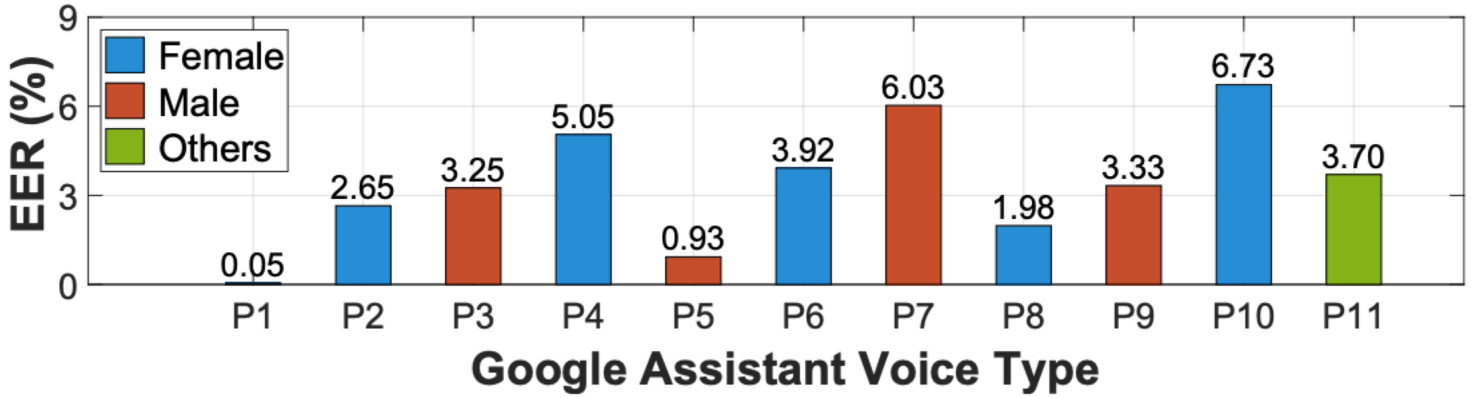
Assumption: VUIs only has a limited set of voice profiles.

Solution: ***Lightweight voice identification DNN models***.



Evaluation: Detect, Segment, and Identify VUI Responses

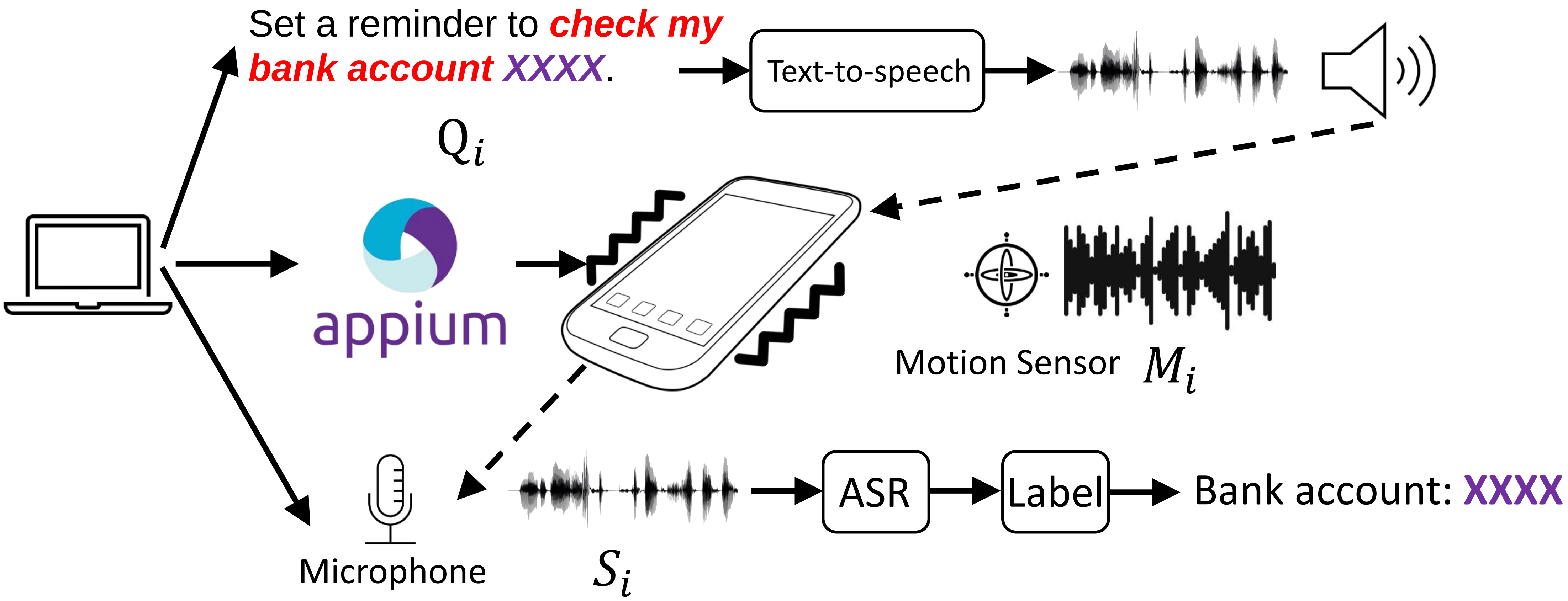
- Equal Error Rate (EER) → both acceptance and rejection errors are equal.



An average 3.5% EER for different voices

	Mobile Data	Segment Memory	Peak CPU	Power
Cloud-based attack	100 KB/min	/	/	/
StealthyIMU	/	16 KB/Seg	5%	35mW

MSS and Corresponding Speech Collection



Extract the Ground Truth Private Entities

Intents:

- Turn left/ right
- Take the next left/ right
- Slight left/ right
- Keep left/ right
- Continue
- Stay
- Take (highway and exit)
- Towards
- Make a U-turn
- Merge
- Follow sign
- Arrive at destination

Name:

- Road
- Highway
- Exit name

Use the right lane to **take** **exit 1b** **toward** **B drive** then **turn left** onto **C street**

Type: navigation

Entities: **type: intent | filler: take**, **type: name | filler: exit 1b**;

type: intent | filler: towards, **type: road | filler: B drive**;

type: intent | filler: turn left, **type: name | filler: C street**;

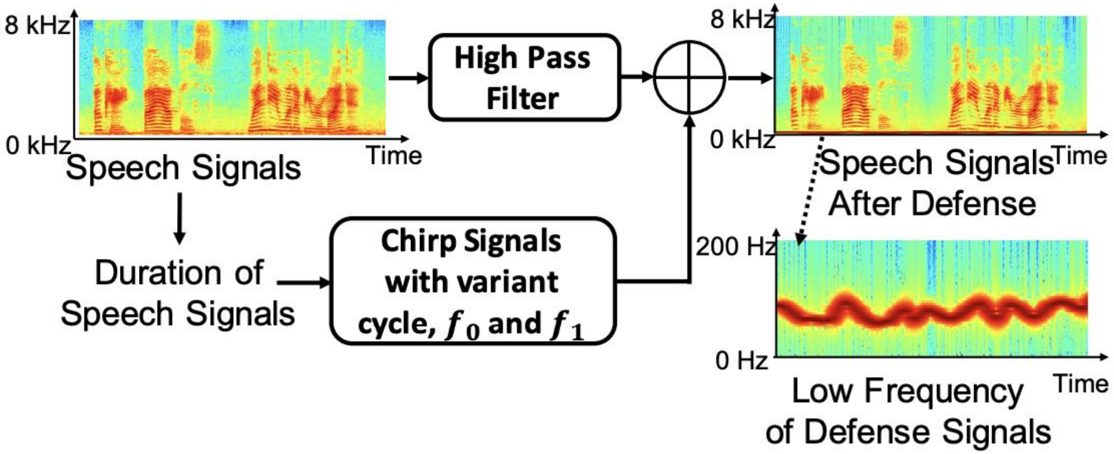
Evaluation: VUI Response Private Entity Recognition

	Model Size	Peak CPU	APP Memory	Time (s/Seg)	Energy (mAh/Seg)
SLU	9.1 MB	13%	54.5 MB	4.60	0.65
SLU	3.9 MB	12%	34.5 MB	1.19	0.17

Our SLU model can recognize 176 voice-associated segments with less than 1% battery consumption.

Defense

- Predistortion of Speech Signals
  - Assumption: the highest MSS sampling rate < 500 Hz
  - Insight: modifying the low modifying the low-frequency components of speech signals
    - Will significantly impact the MSS.
    - Will not affect the human perceivable speech quality.



Defense

- Redesigning the Permissions
  - If vendors unleash the sampling rate limitation to 4000 Hz

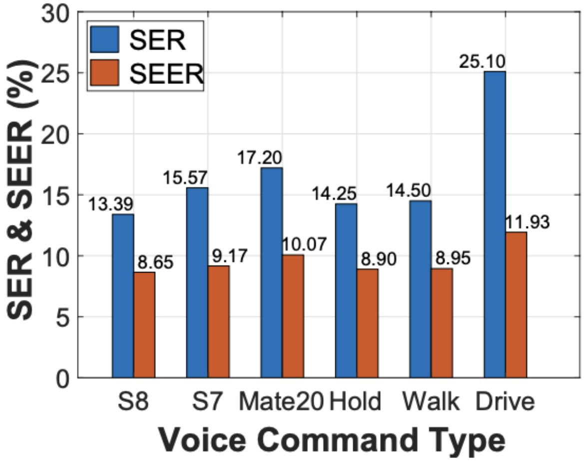
Even if future smartphone OS restricts the motion sensor permission, the StealthyIMU attack can still work—it can pretend to be an innocuous app that needs the motion sensor permission alone.

Evaluation: Generalization

- Different Sampling Rate, Smartphone models

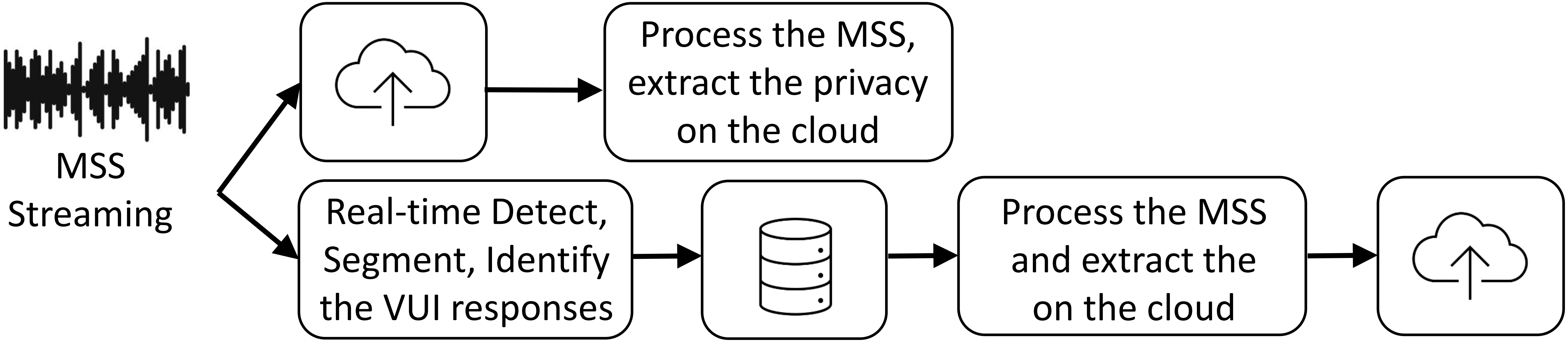
Phone	OS Version	Sampling Rate	SER	SEER
OnePlus	Android 11	440 Hz	13.85%	8.99%
Samsung S8	Android 9	400 Hz	13.39%	8.65%
Samsung S8	Android 9	200 Hz	62.69%	38.30%
Samsung S8	Android 9	100 Hz	84.01%	52.50%
Huawei Mate 20	Android 9	500 Hz	17.20%	10.07%
Samsung S7	Android 8	420 Hz	15.57%	9.17%

- Different Motion Artifacts Interference



Evaluation: Implementation & Overhead

Attack Implementation:



- Voice detection and segmentation overhead (running in the background)

	Mobile Data	Segment Memory	Peak CPU	Power
1	100 KB/min	/	/	/
2	/	16 KB/Seg	5%	35mW

Evaluation: System Overhead

- On-device DNN overhead (running when the app is active)

	Model Size	Peak CPU	APP Memory	Time (s/Seg)	Energy (mAh/Seg)
ID	79.6 KB	5%	4.1 MB	$9.8e-3$	$6.5e-3$
ID	1.7 MB	5%	7.4 MB	$53.3e-3$	$1.1e-3$
SLU	9.1 MB	13%	54.5 MB	4.60	0.65
SLU	3.9 MB	12%	34.5 MB	1.19	0.17

Overall, the on-device implementation of StealthyIMU is unlikely to be distinguishable from an innocuous app.