Tracking Mobile Web Users Through Motion Sensors: Attacks and Defenses

Anupam Das (UIUC),
Nikita Borisov (UIUC),
Matthew Caesar (UIUC)

NDSS '16
SYMPOSIUM

Real World Digital Stalking

MOTHERBOARD Watch Machines Discoveries Space Futures Gaming Earth

Written by

How are they tracking devices?

- Device Fingerprint ~ *Set* (unique device properties)

IEEE

RISK ASSESSMENT / SECURITY & HACKTIVISM

Why fingerprint devices?

- Targeted Advertisement (tracking usage pattern)

Top Websites Secretly Track Your Device Fingerprint

By Jeremy Hsu
Posted 11 Oct 2013 | 20:46 GMT

Mobile Ad Expenditure

Mobile Internet Ad Spending Worldwide, 2013-2019

	2013	2014	2015	2016	2017	2018	2019
Mobile internet ad spending	$19.20	$42.63	$68.69	$101.37	$133.74	$166.63	$195.55

There are multiple companies such as TapAd and AdTruth that utilize device fingerprinting to build cross-device user profile.

Note: includes display (banners, video and rich media) and search; excludes SMS, MMS and P2P messaging-based advertising; ad spending on tablets is included
Source: eMarketer, March 2015

186887 www.eMarketer.com

Targeted ad can help increase the Return On Ad Spend.

Device Fingerprinting Techniques

How are device fingerprints generated?

Device Fingerprint

Software Variations

- Difference in Protocol Stack/Network Stack
- Difference in Firmware and Device Driver
- Difference in installed Software
- MAC Headers

Hardware idiosyncrasies

- Difference in spectral property of Radio Signal Transmitters
- Difference in emitted radio frequency of NIC
- Unique and constant clock skews in network devices

Exploit small deviations in either the software or hardware characteristics of the device.

Example: Browser Fingerprinting

https://amiunique.org

Are you unique?

Yes! (You can be tracked!)

34.62 % of observed browsers are **Chrome**, as yours.

0.25 % of observed browsers are **Chrome 48.0**, as yours.

16.53 % of observed browsers run **Linux**, as yours.

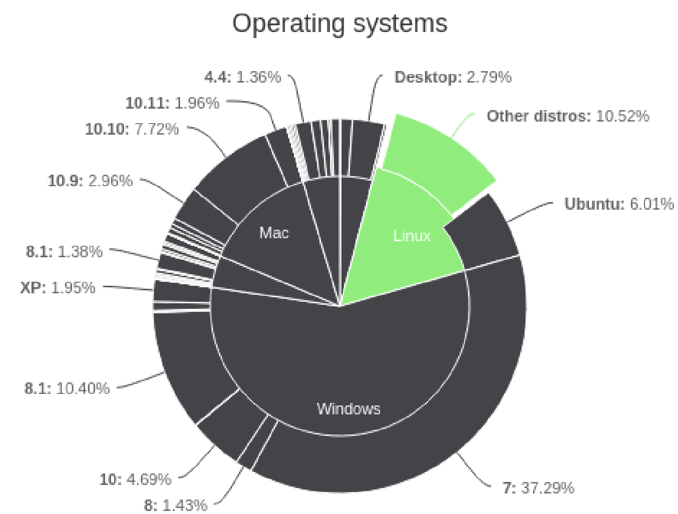
63.26 % of observed browsers have set **"en"**as their primary language, as yours.

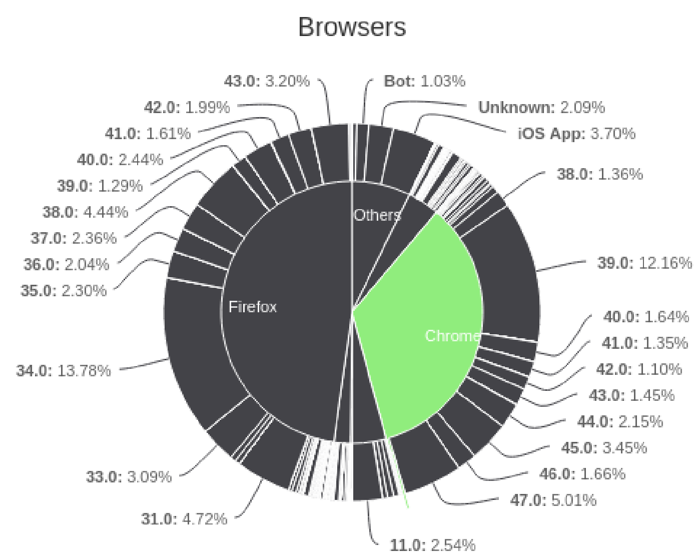
3.83 % of observed browsers have **UTC-6** as their timezone, as yours.

However, your full fingerprint is unique among the 134529 collected so far. Want to know why?

Click here

View more details | View graphs | Force fingerprinting





Fingerprinting Smartphones

Can traditional approaches be applied to fingerprint smartphones?

Smartphones are somewhat less susceptible to software-based fingerprinting approaches due to a stable software base.

https://amiunique.org

Browser Characteristic	% of fingerprints sharing same value	
	Laptop (ThinkPad L540)	Smartphone (iPhone 5)
User agent	<0.1%	<0.1%
List of plugins	0.28%	17.05%
List of fonts	<0.1%	23.72%
Screen resolution	9.83%	0.95%
Canvas	0.34%	0.11%

How are Smartphones Different?

Smartphones are equipped with a wide range of sensors.

- accelerometer
- gyroscope
- magnetometer
- Bluetooth
- GPS
- WiFi + cellular
- humidity
- temperature

We focus on exploiting onboard sensors to generate unique fingerprints.

Applications:

- Motion detection

ction

n

earby devices

- Navigation
- etc.

Our Contribution

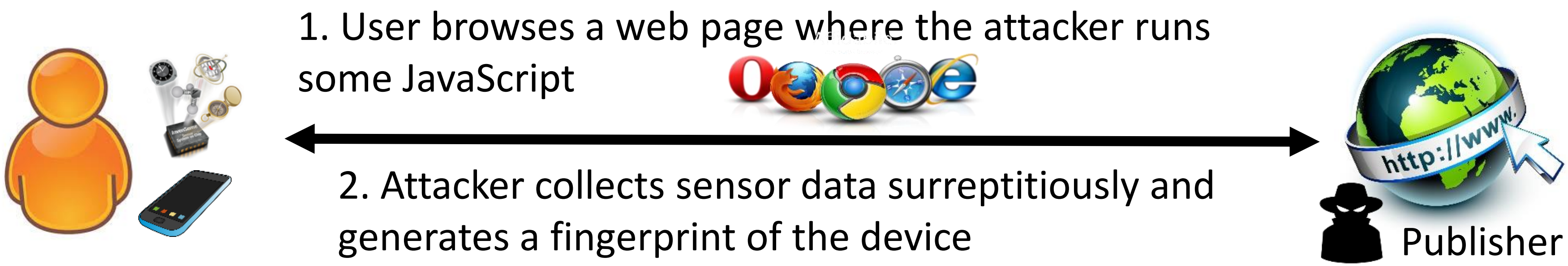
We'll look at addressing the following questions:

- ➢ Can smartphones be fingerprinted using motion sensors?
- ➢ Are there ways to mitigate such fingerprinting techniques?
- ➢ Are there any implications of such mitigation techniques?

Fingerprint Motion Sensors

Fingerprint smartphone using accelerometer and gyroscope.

Attack Scenario



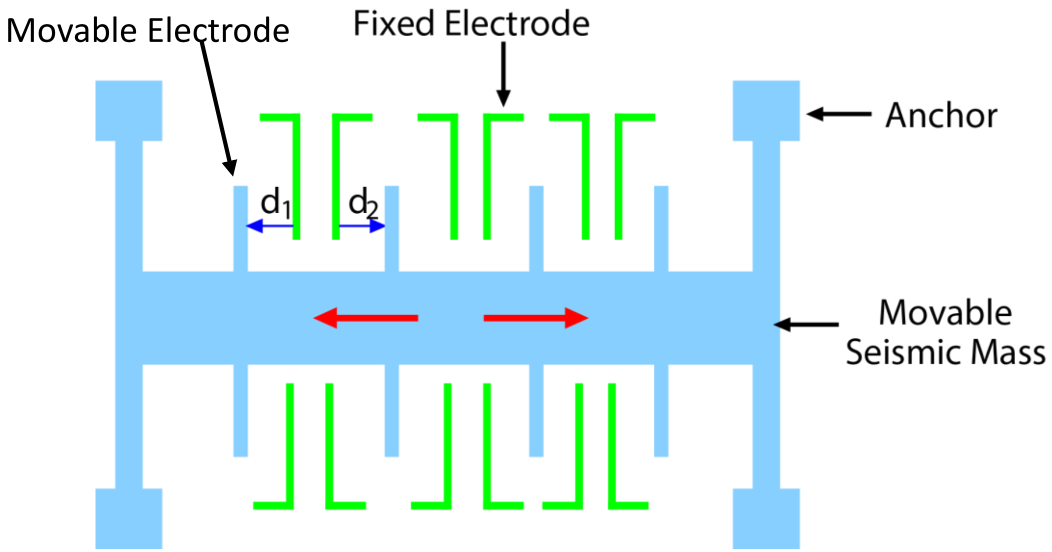
<u>Device Position</u>:
On Desk: Devices kept on top of a desk
In Hand: Devices kept in the hand of the user while user is sitting in a chair

Requires No Explicit Permissions!!!

Source of Uniqueness

MEMS Accelerometer:

Mechanical Energy ➡ Capacitive Change ➡ Voltage Change



Gap ~ 1.3µm

Sensitivity ~ 20pm

Possible source of idiosyncrasies:

- Slightest gap difference between the structural electrodes
- Flexibility of the seismic mass

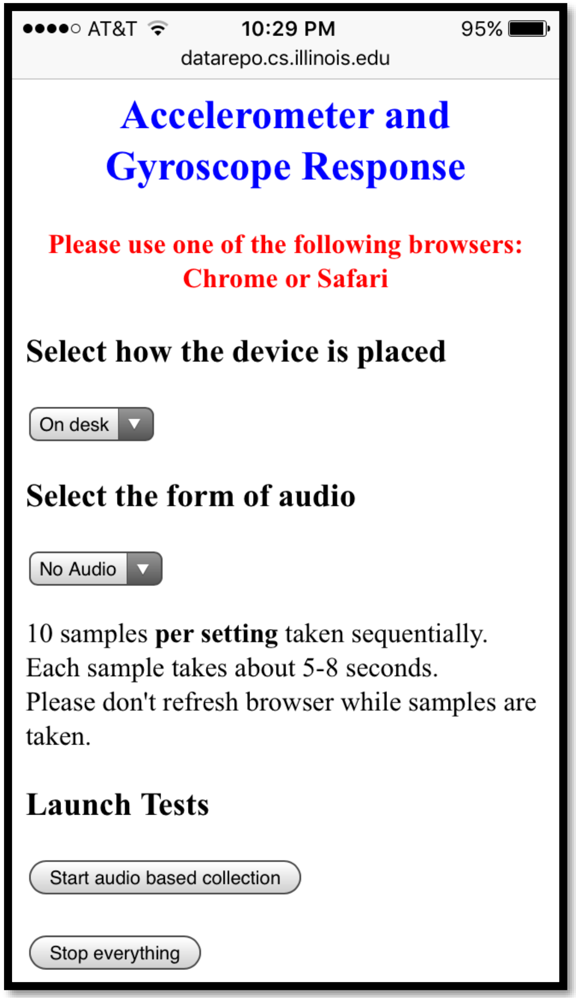
Data Collection Setup

Using JavaScript we collected sensor data through the web browser.

OS	Browser	Sampling Freq. (Hz)	Sensors Accessible*
Android 4.4	Chrome	100	A,G
	Android	20	A
	Opera	40	A,G
	UC Browser	20	A,G
	Standalone App	200	A,G
iOS 8.1.3	Safari	100	A,G
	Chrome	100	A,G
	Standalone App	100	A,G

*A=Accelerometer, G=Gyroscope

Chrome being the most popular mobile browser, we collect lab-data using the Chrome browser.



Experimental Setup

Devices:

Maker	Model	#
Apple	iPhone 5	4
	iPhone 5s	3
Samsung	Nexus S	14
	Galaxy S3	4
	Galaxy S4	5
Total		30

Data Streams:

Four data streams are considered:

1. Accelerometer Magnitude
2. Gyroscope X-axis
3. Gyroscope Y-axis
4. Gyroscope Z-axis

Samples:

- 10 samples per device per setting
- Each sample is around 5-8 second

Settings:

Stimulation Type	Description
No Audio	No audio is being played through the speaker
Inaudible Audio	20kHz Sine wave is being played through the speaker
Popular Song	A popular song is being played through the speaker

Features

25 features were explored.

#	Temporal Feature
1	Mean
2	Standard Deviation
3	Average Deviation
4	Skewness
9	Zero Crossing Rate
10	Non-Negative Count

#	Spectral Feature
1	Spectral Root Mean Square
2	Spectral Spread
3	Spectral Low-Energy-Rate
4	Spectral Centroid
5	Spectral Entropy
10	Spectral Rolloff
11	Spectral Brightness
12	Spectral Flatness
13	Spectral Flux
14	Spectral Attack Slope
15	Spectral Attack Time

Joint-Mutual-Information (JMI) is used for feature exploration to determine the best subset of features for classification.

For Spectral Features, cubic-spline interpolation is used to obtain a sampling rate of 8kHz.

Evaluation Algorithms & Metrics

Tested several supervised classifiers:

- SVM,
- Naive-Bayes classifier,
- Multiclass Decision Tree,
- k-NN,
- **Bagged Decision Trees**.

Evaluation metrics:

$$Precision = \frac{TP}{TP + FP}$$

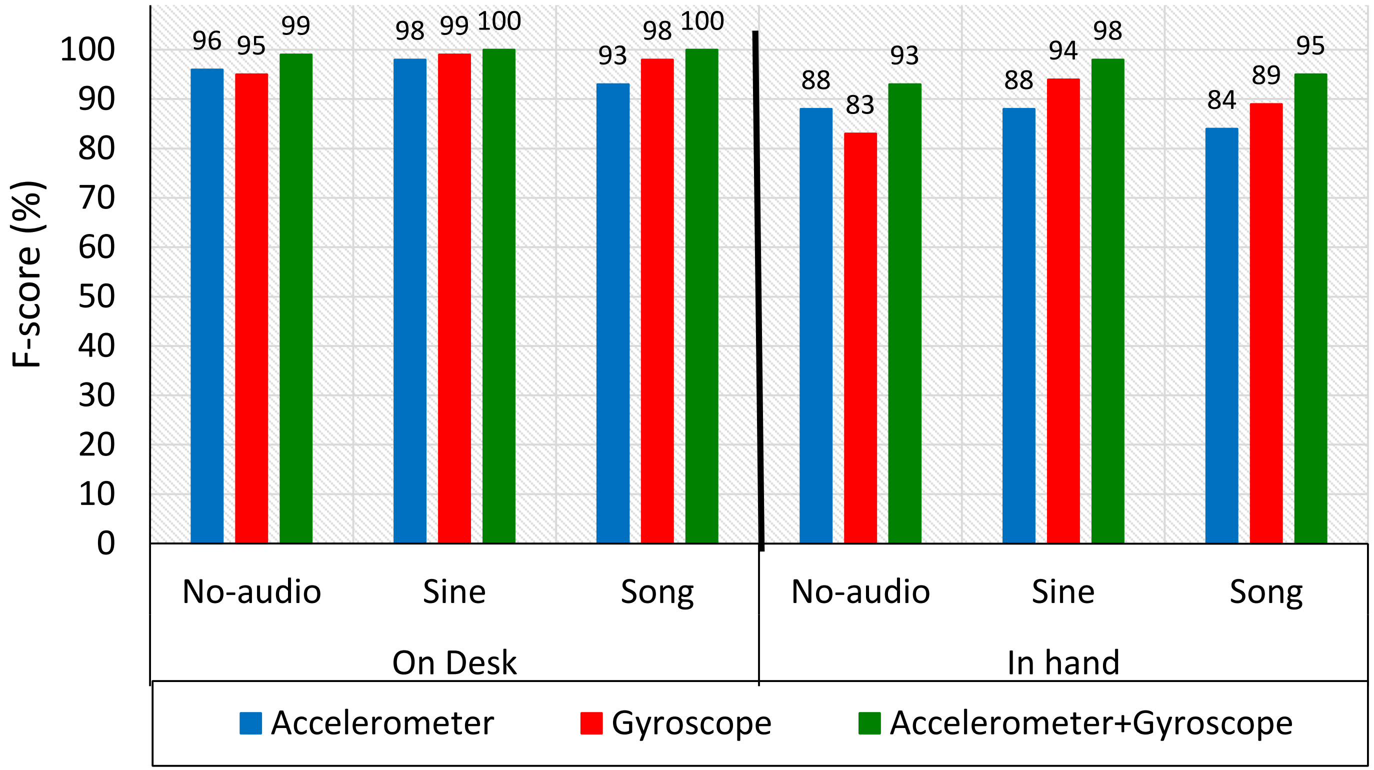
$$Recall = \frac{TP}{TP + FN}$$

$$F\_Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

TP: True Positive
FP: False Positive
FN: False Negative

Randomly portioned 50% of the data for training and testing.
Reported the average of 10 iterations.

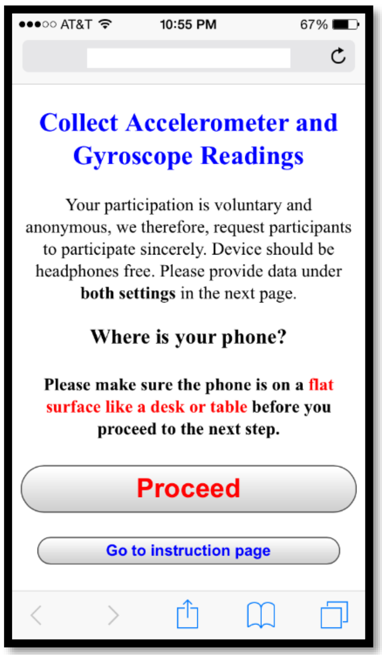
Results: Lab Setting

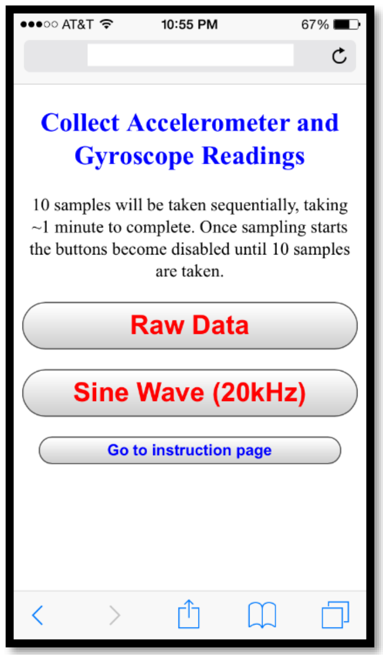


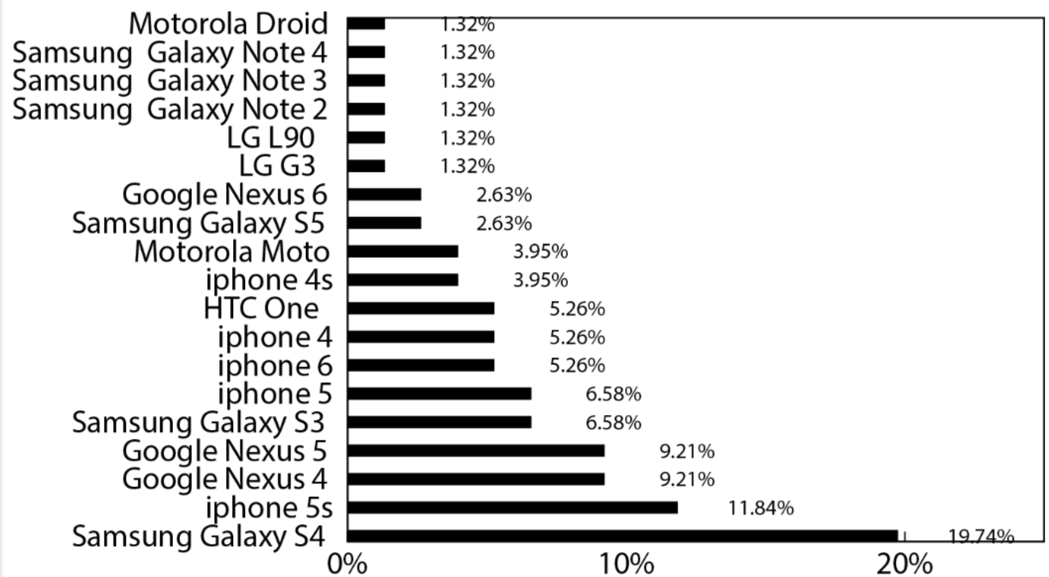
Combining features from both accelerometer and gyroscope yielded the best results.

Real-World Data

Invited people to voluntarily participate in our study.

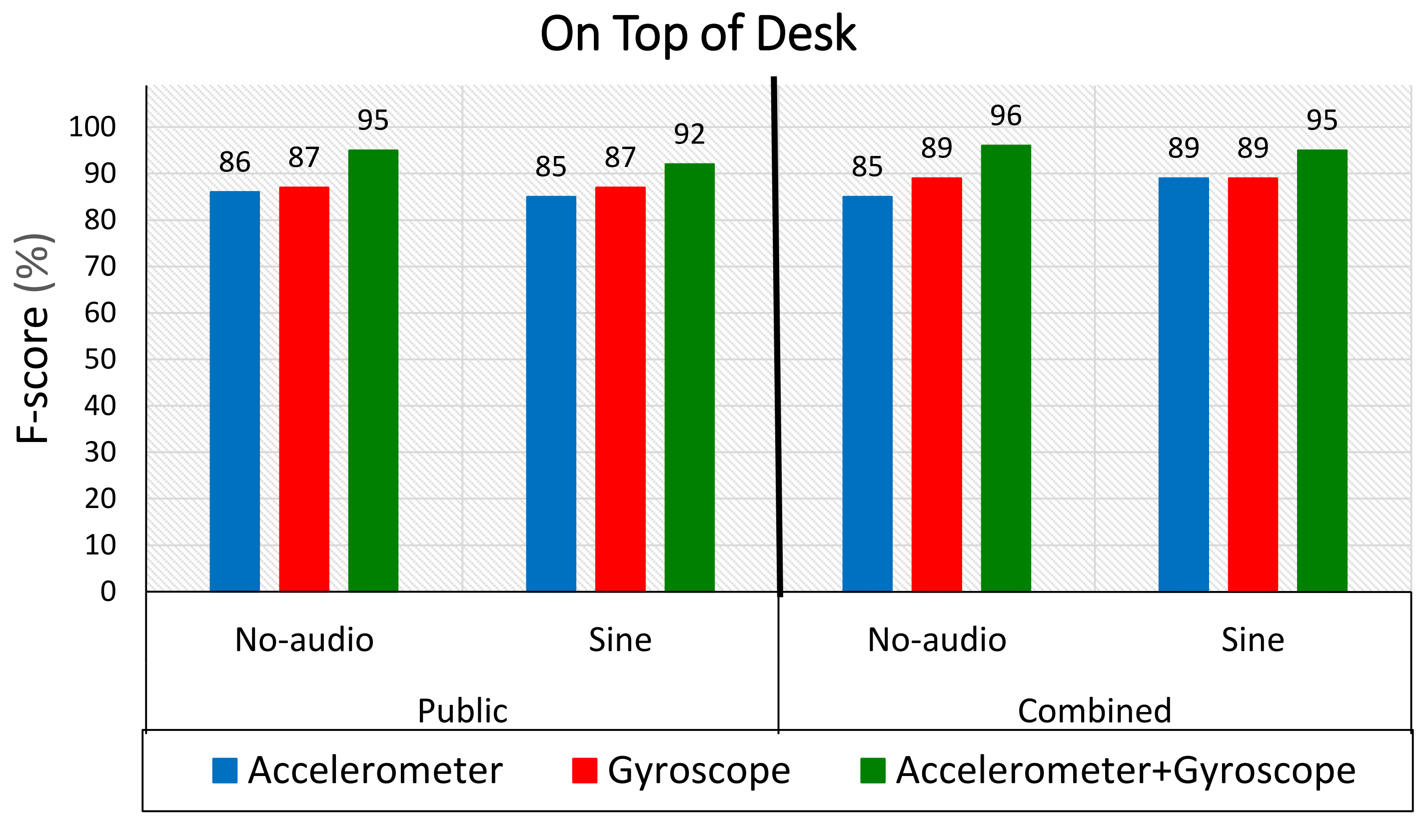






76 participants visited our web page in two weeks but only 63 of the devices actually provided any form of data.

Public and Combined Setting



Public setting : F_score of 95%
Combined setting: F_score of 96%

Mitigation Techniques

We explore two types of countermeasure techniques:

- Sensor Calibration
  -- Computing offset and gain error of sensors.
- Data Obfuscation
  -- Adding noise to data to obfuscate data source.

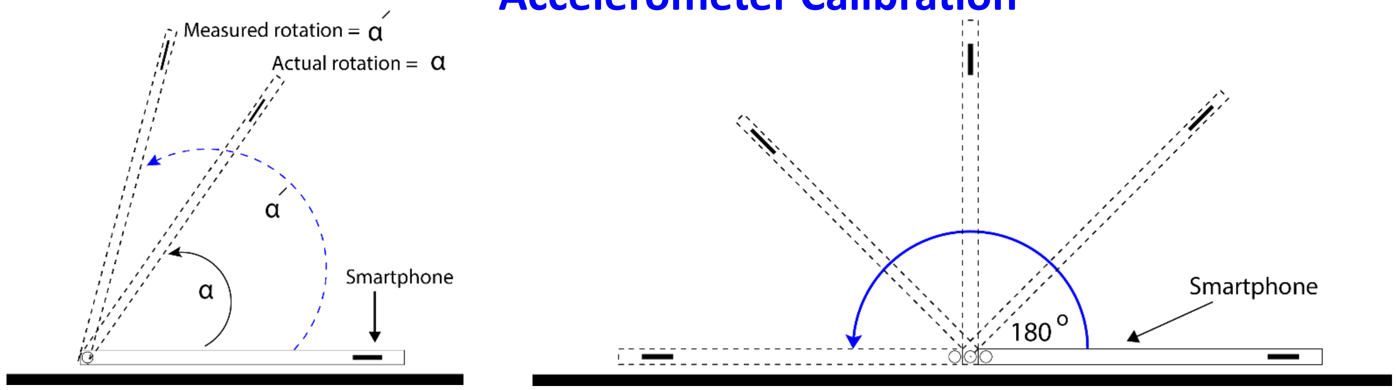
Two extreme approaches:

Sensor Calibration: Map every device to the same point.
Data Obfuscation: Scatter the same device to different points.

Sensor Calibration

Measured sensor value $a^M = O + S.a,$ where $\boldsymbol{O}$ and $\boldsymbol{S}$ represent the offset and gain error along an axis respectively.

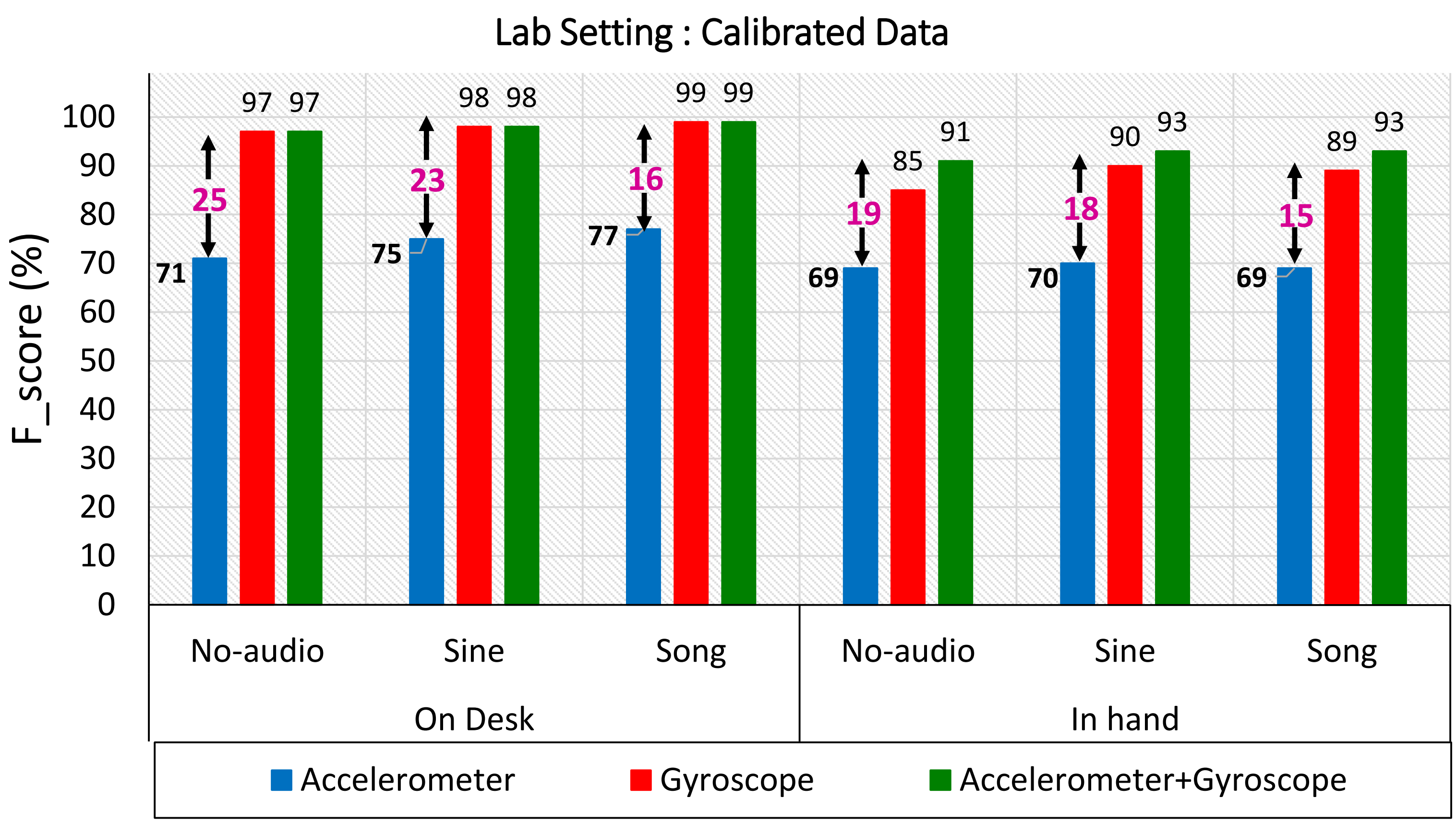
Accelerometer Calibration



Gyroscope Calibration

Measurements along all six directions $(\pm x, \pm y, \pm z)$ are taken.

Results: Calibrated Data



F_score reduces by approximately 15–25% for accelerometer data but not much for the gyroscope data.

Data Obfuscation

Instead of removing the calibration errors, we can add extra noise to hide the miscalibration.

We explore the following 3 techniques:

- **Uniform noise**: highest entropy while having a bound.
- **Laplace noise**: highest entropy which is inspired by Differential Privacy.
- **White noise**: affecting all aspects of a signal.

Uniform Noise

To add obfuscation noise, we compute $a^o = O^o + S^o a^M$
Here, S^o and O^o are the obfuscated gain and offset error.

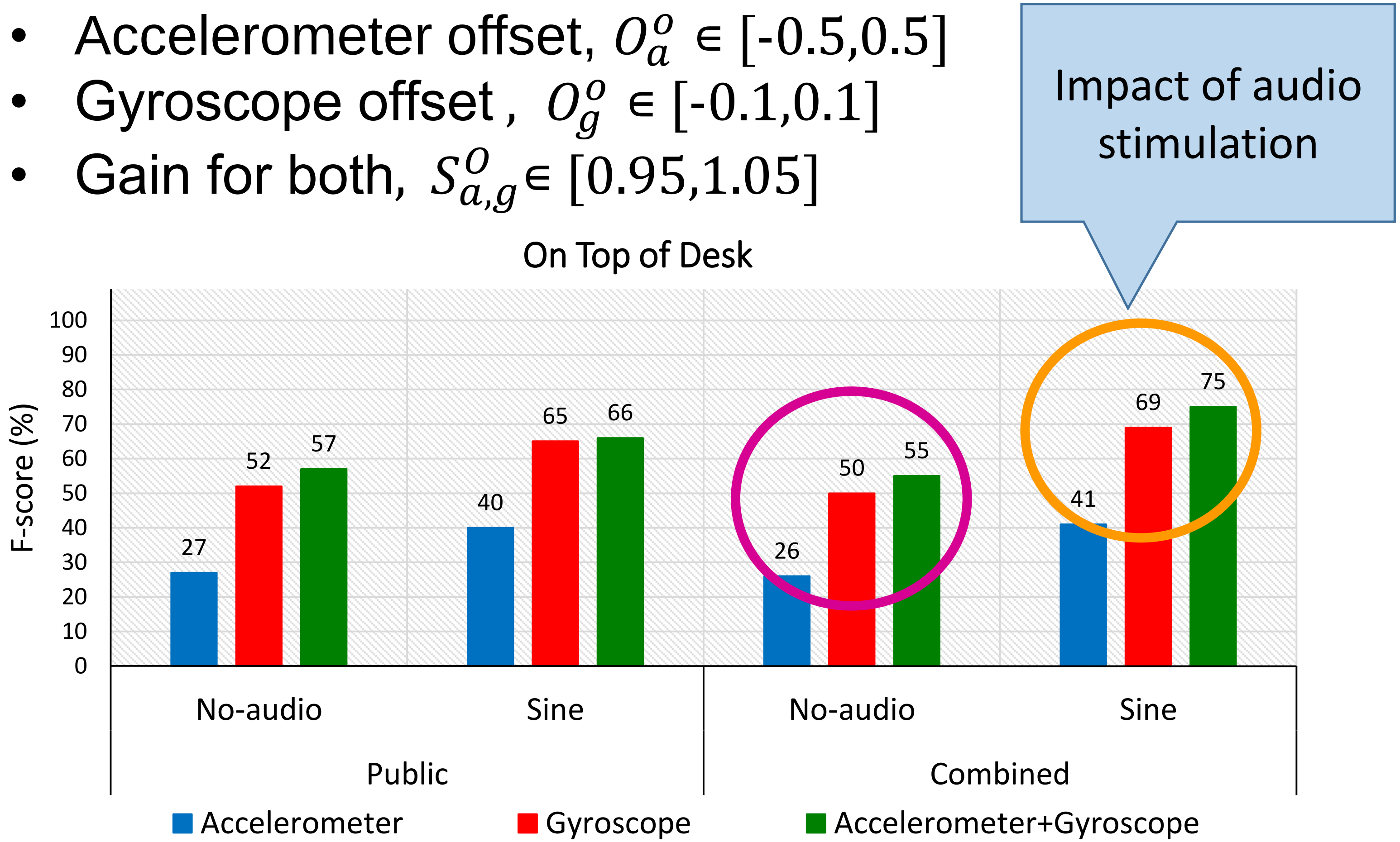
We explore three variations of adding uniform noise:

- **Basic Obfuscation**
- Increased Range Obfuscation
- Enhanced Obfuscation

Basic Obfuscation

Based on the calibration errors found from our lab phones we set the base error ranges as follows:

- Accelerometer offset, $O_a^o \in [-0.5,0.5]$
- Gyroscope offset , $O_g^o \in [-0.1,0.1]$
- Gain for both, $S_{a,g}^O \in [0.95,1.05]$



Impact of Mitigation Techniques

We prototype a simple application like step-counter.

Participant takes 20 steps and the process is repeated 10 times.

Data Stream	Step Count	
	Mean	Std Dev
Raw Stream	20	0
Calibrated	20.1	0.32
Basic Obfuscated	20.1	0.32
Increased Obfuscated Range	19.9	1.69
Enhanced Obfuscated	25.1	4.63

- Both calibration and basic obfuscation seem to be benign.
- Both increased and enhanced obfuscation scheme seem to have an adverse affect.

Recommendation

- Request explicit user permission.
- Data is always obfuscated unless the user explicitly allows an application to access unaltered sensor data. This enforces developer to request explicit permissions for legitimate usage.

Thank You

If you would like to participate in our study or learn more about our work please go to the following link

http://hatswitch.org/phonestudy

Contact Info:
das17@illinois.edu
http://web.engr.illinois.edu/~das17/