Do You Hear What I Hear? Fingerprinting Smart Devices Through Embedded Acoustic Components

A.Das, N.Borisov, M.Caesar
CCS 2014

Presented by
Siddharth Murali

Fingerprinting smartphones

- Being able to uniquely identify a smartphone
- Why is this important?
  - Tracking mobile phones
  - User based advertising

Fingerprinting smartphones

› Being able to uniquely identify a smartphone

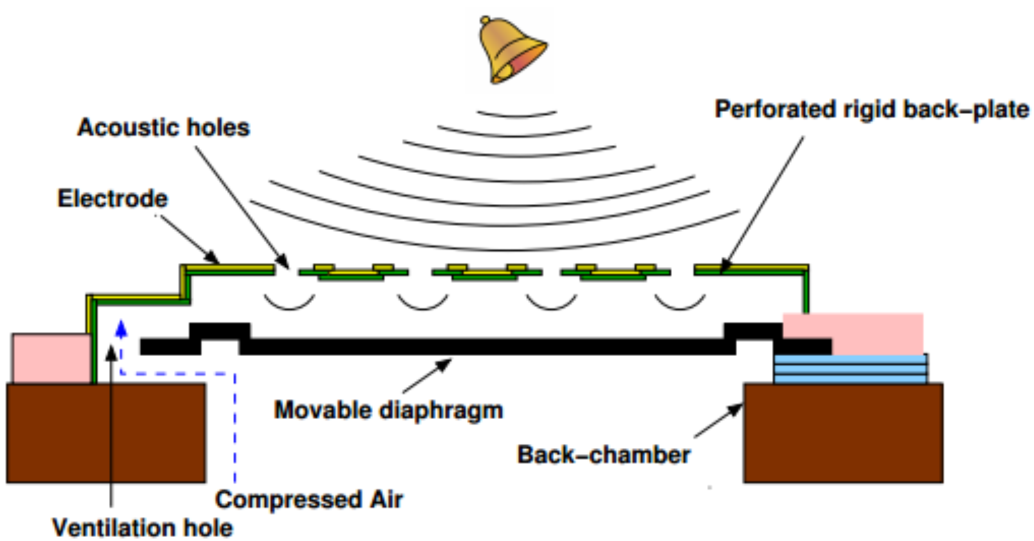
› Software methods
  – Timing analysis of network packets
  – Fonts installed in browsers
  – Browsing history
  – Nmap, Xprobe, able to identify unique responses from the networking stack

Fingerprinting smartphones

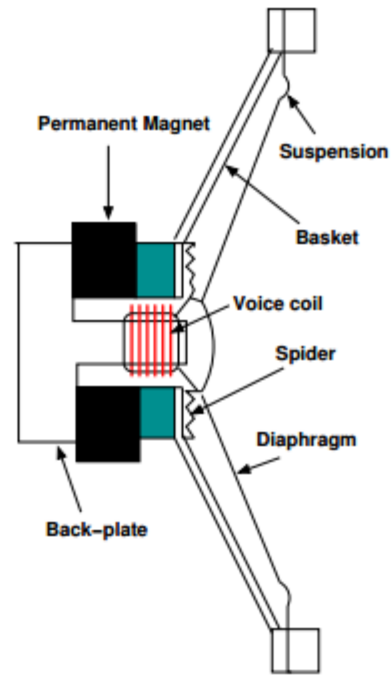
› Hardware methods
  – Using clock skews of network devices
  – Radio transmitters
  – Network interface cards
  – Smartphone accelerometers
  – Now, acoustic components like speakers, microphones

Microphones and Microspeakers

› Based on MEMS technology



Microphone



Microspeaker

Classification Algorithms

› k-Nearest Neighbors
  – Computes distance to learned data points, and classifies our data point based on nearest k data points.

› Gaussian Mixture Model
  – Computes probability distribution for each class, and determines maximal likely association

Testing and results

› For analysis of the audio, they used MIRToolbox, Netlab, Audacity, Hertz

› Each sample audio was recorded 10 times, 50% for training and 50% for testing

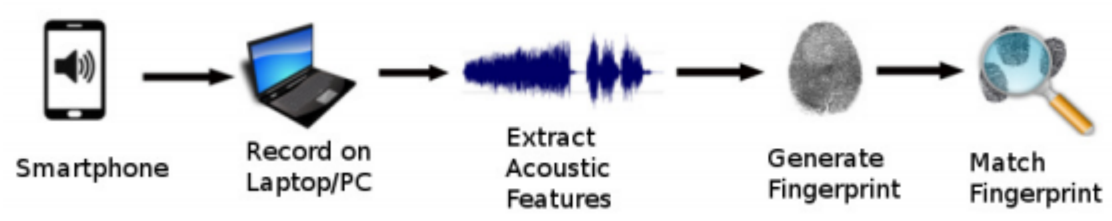
Table 3: Types of audio excerpts

Type	Description	Variations
Instrumental	Musical instruments playing together, e.g., ringtone	4
Human speech	Small segments of human speech	4
Song	Combination of human voice & instrumental sound	3

Maker	Model	Quantity
Apple	iPhone 5	1
Google	Nexus One	14
	Nexus S	8
Samsung	Galaxy S3	3
	Galaxy S4	10
Motorola	Droid A855	15
Sony Ericsson	W518	1
Total		52

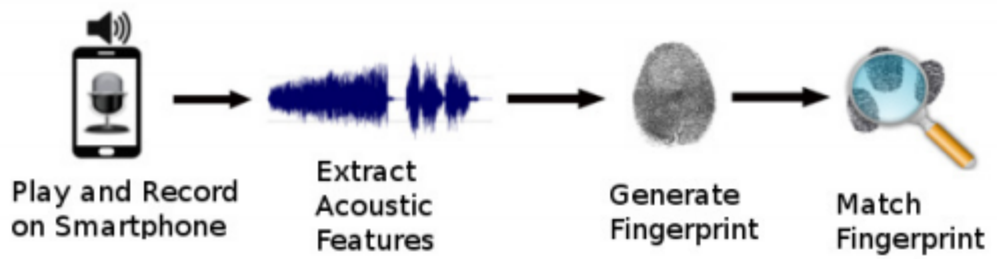
Testing and results

› Fingerprinting the speaker



› Fingerprinting the microphone

› Fingerprinting both speaker and microphone



Testing and results – Different model and make

Table 5: Fingerprinting different smartphones using speaker output

Audio Type	k-NN				GMM			
	Features*	$AvgPr$	$AvgRe$	$AvgF1$	Features*	$AvgPr$	$AvgRe$	$AvgF1$
Instrumental	[1,7]	97.6	97.1	97.4	[13]	100	100	100
Human speech	[13]	95.2	94.3	94.8	[13]	100	100	100
Song	[15]	97.6	97.1	97.4	[13]	100	100	100

Table 6: Fingerprinting different smartphones using mic

Audio Type	k-NN				GMM			
	Features*	$AvgPr$	$AvgRe$	$AvgF1$	Features*	$AvgPr$	$AvgRe$	$AvgF1$
Instrumental	[13,1]	95.2	94.3	94.8	[13,1,7]	100	100	100
Human speech	[15,9,1]	95.2	94.3	94.8	[13,15,11]	97.6	97.1	97.4
Song	[13,1,12]	97.6	97.1	97.4	[13,1,9]	100	100	100

Table 7: Fingerprinting different smartphones using mic & speak

Audio Type	k-NN				GMM			
	Features*	$AvgPr$	$AvgRe$	$AvgF1$	Features*	$AvgPr$	$AvgRe$	$AvgF$
Instrumental	[10]	96.7	96	96.3	[13]	100	100	100
Human speech	[12]	96.7	96	96.3	[13]	100	100	100
Song	[10]	96.7	96	96.3	[13]	100	100	100

Testing and results – Same model and make

Table 9: Fingerprinting similar smartphones using speaker output

Audio Type	k-NN				GMM			
	Features*	$AvgPr$	$AvgRe$	$AvgF1$	Features*	$AvgPr$	$AvgRe$	$AvgF1$
Instrumental	[13,14]	96.7	96	96.3	[13,14]	98.4	98.1	98.3
Human speech	[13]	98.9	98.7	98.8	[13,14]	98.9	98.7	98.8
Song	[13,7]	93.2	92	92.6	[13,14]	95.6	93.3	94.5

Table 10: Fingerprinting similar smartphones using microphone

Audio Type	k-NN				GMM			
	Features*	$AvgPr$	$AvgRe$	$AvgF1$	Features*	$AvgPr$	$AvgRe$	$AvgF1$
Instrumental	[13,8,12]	95.9	94.7	95.3	[13,8,12]	96	94.7	95.3
Human speech	[13]	98.9	98.7	98.8	[13,14]	100	100	100
Song	[13,14,10]	96.4	96	96.2	[13,14]	96.5	95.7	96.1

Table 11: Fingerprinting similar smartphones using mic & speaker

Audio Type	k-NN				GMM			
	Features*	$AvgPr$	$AvgRe$	$AvgF1$	Features*	$AvgPr$	$AvgRe$	$AvgF1$
Instrumental	[13]	100	100	100	[13]	100	100	100
Human speech	[13]	100	100	100	[13]	100	100	100
Song	[13]	100	100	100	[13]	100	100	100

Testing and results – All combinations

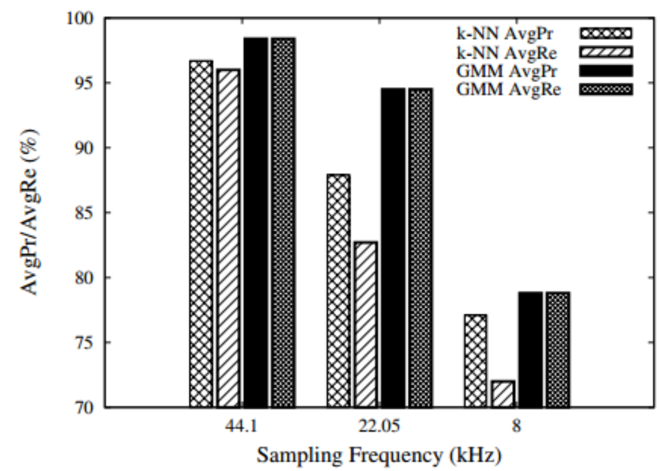
› Results show that malicious applications that have access to mic and speakers can fingerprint smartphones with an accuracy of over 98%

Table 13: Fingerprinting all smartphones using mic & speaker

Audio Type	k-NN				GMM			
	Features*	$AvgPr$	$AvgRe$	$AvgF1$	Features*	$AvgPr$	$AvgRe$	$AvgF1$
Instrumental	[13]	99.3	98.8	99	[13]	98.6	98.1	98.3
Human speech	[13]	99.7	99.6	99.6	[13]	99.4	99.2	99.3
Song	[13]	99.7	99.6	99.6	[13]	100	100	100

Sensitivity analysis

› Impact of sampling rate
  – Lower sampling rate led to reduced accuracy



› Impact of training size
  – Lower training size also led to reduced accuracy

Table 14: Impact of varying training size

Training samples per class	k-NN Features [13,14]*			GMM Features [13,14]*		
	$AvgPr$	$AvgRe$	$AvgF1$	$AvgPr$	$AvgRe$	$AvgF1$
1	42	49.3	45.3	50	53.3	51.6
2	79.2	80	79.6	80.4	80	80.2
3	91.3	89.3	90.2	91.7	89.3	90.5
4	95.3	94.7	95	95.6	94.7	95.1
5	96.7	96	96.3	98.4	98.1	98.3

Sensitivity analysis

- Varying distance between speaker and recorder

Table 15: Impact of varying distance

Distance (in meters)	k-NN Features [13,14]*			GMM Features [13,14]*		
	$AvgPr$	$AvgRe$	$AvgF1$	$AvgPr$	$AvgRe$	$AvgF1$
0.1	96.7	96	96.3	98.4	98.1	98.3
1	92.7	91.5	92	95.2	94.7	94.9
2	88.2	87.6	87.9	94.5	92	93.2
3	76.7	76	76.3	78.9	84	81.4
4	70.2	64	67	76.8	76	76.4
5	64.5	62.7	63.6	77	73.3	75.1

- Ambient background noise

Table 16: Impact of ambient background noise

Environments	SNR (dB)	k-NN Features [13,14]*			GMM Features [13,14]*		
		$AvgPr$	$AvgRe$	$AvgF1$	$AvgPr$	$AvgRe$	$AvgF1$
Shopping Mall	15.85	88.8	85.3	87	95.1	93.3	94.2
Restaurant/Cafe	17.77	90.5	89.7	90.1	92.5	90.7	91.6
City Park	15.43	91.7	90	90.8	95.2	94.1	94.6
Airport Gate	14.92	91.3	89.5	90.4	94.5	93.3	93.9

Discussion

- Key contributions of the paper?
- Limitations/criticisms of the paper?
- Accelerometer vs Acoustic for fingerprinting
- Can we use permissions to prevent this? Other methods?