Hearing your touch: an acoustic side-channel on smartphones

Ilia Shumailov, Laurent Simon, Jeff Yan, Ross Anderson

Department of Computer Science and Technology, University of Cambridge

Threat model

- Attacker has an application running on target phone
- Targets phone has access to microphone(s)
- Attacker knows the model of the phone used
- Attacker wants to steal PIN-codes and text entered on the phone in another application

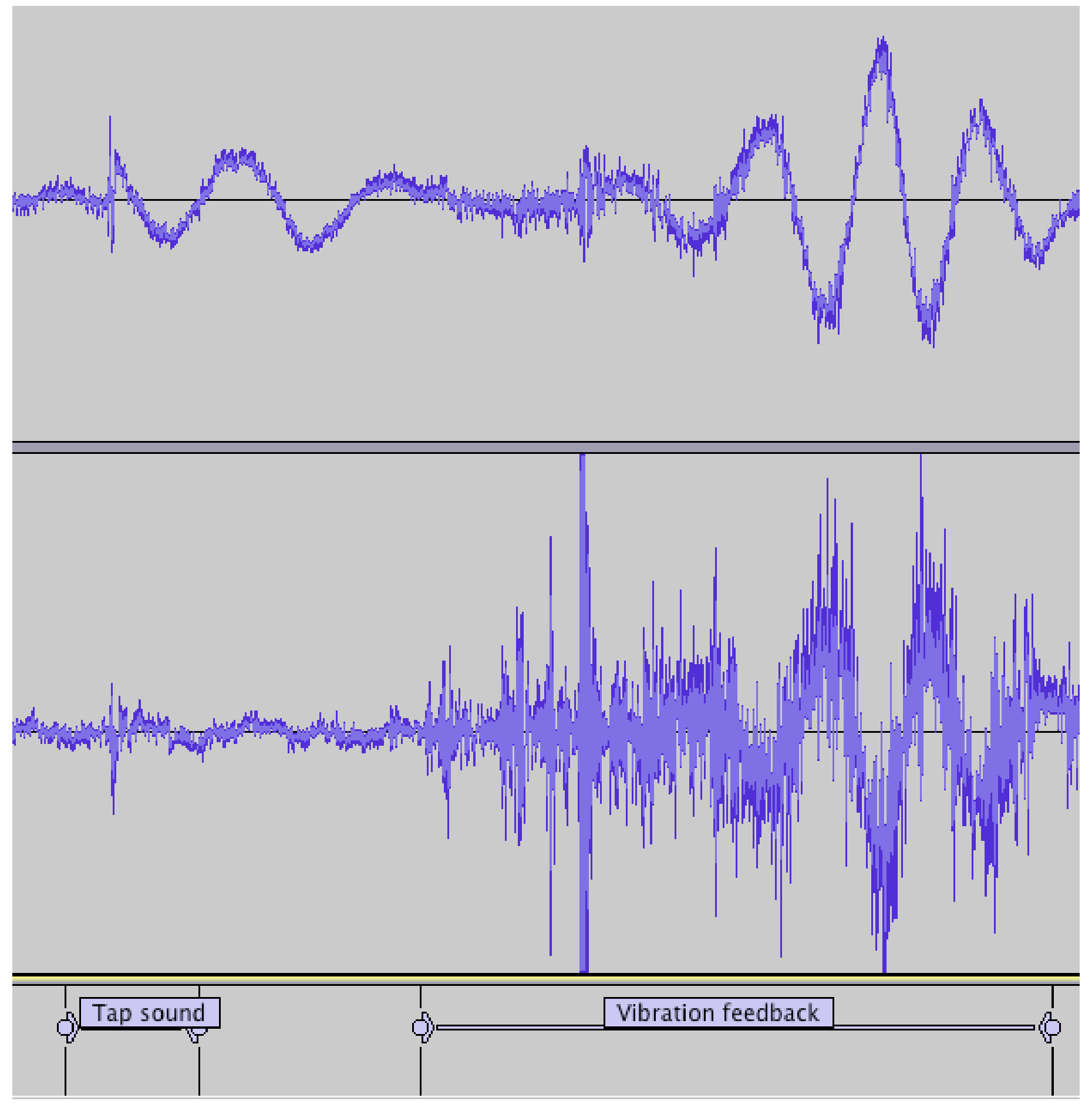

Figure: Vibration and Sound feedback comes long after the tap

Why does it work?

- Fixed plate vibrates upon pressure
- Speed in Gorilla Glass 3 is about 4154.44$\frac{m}{s}$
- Modern microphones support sampling rates up to 44.1 kHz
- There are multiple microphones to perform noise cancellation

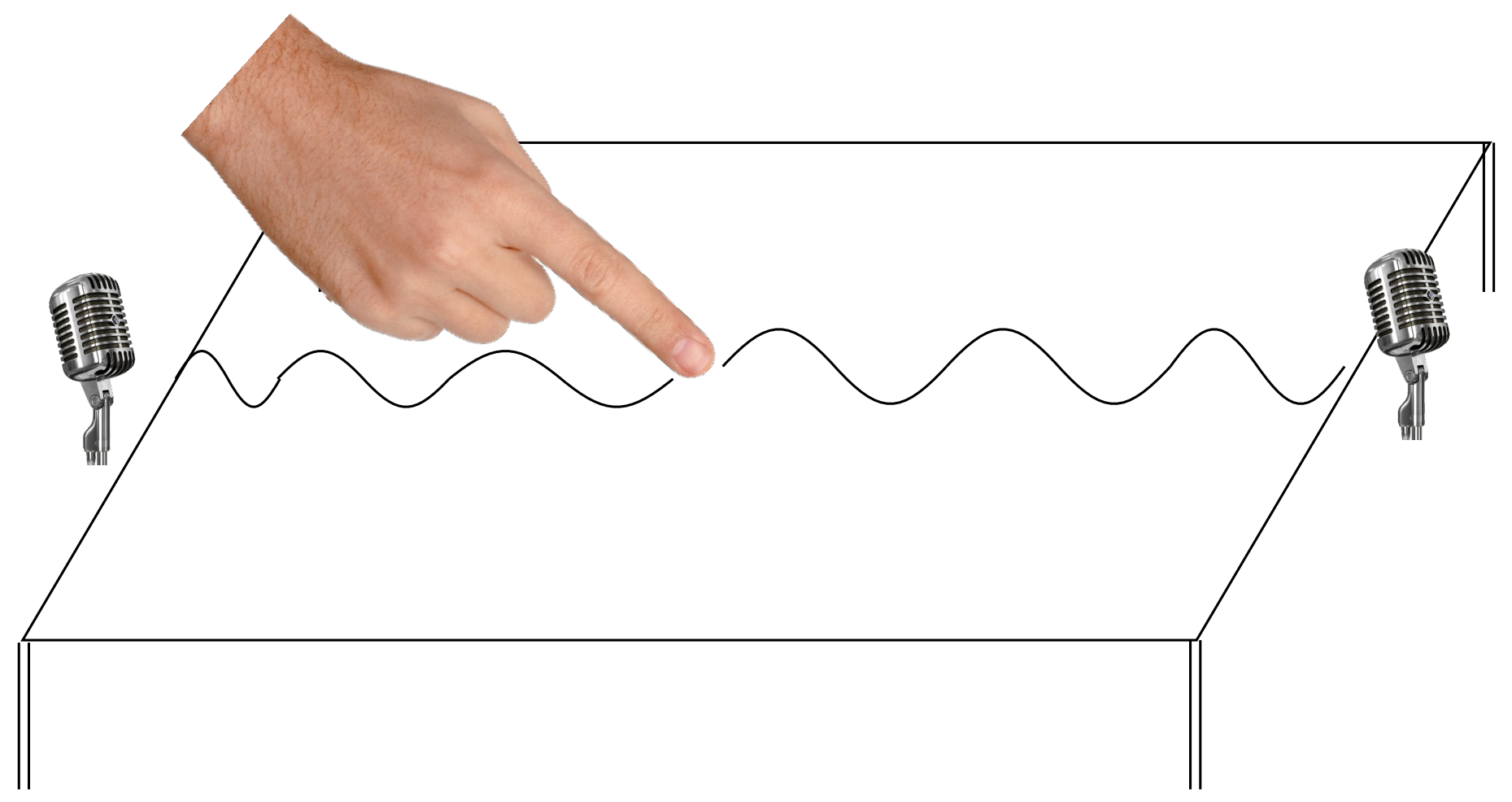

Figure: Screen is a fixed plate that vibrates upon pressure

Time Difference of Arrival(TDOA)

Smartphones provide access to high resolution synchronised data. Common TDOA estimation techniques work!

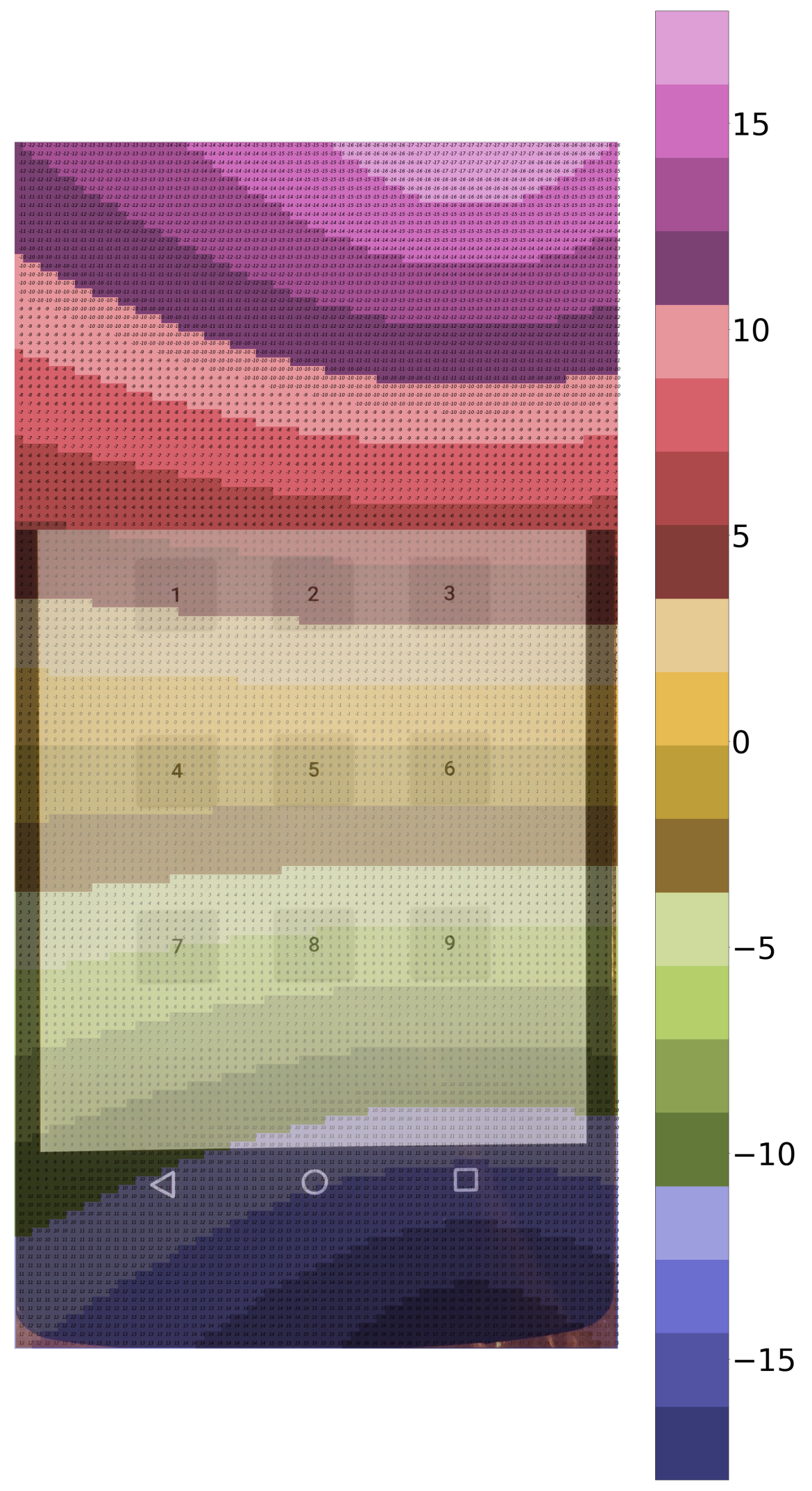

Figure: Theoretical recognisability for Nexus 5 phone. From Microphone 1 to Microphone 2 the difference is about 32 samples.

Practical TDOA

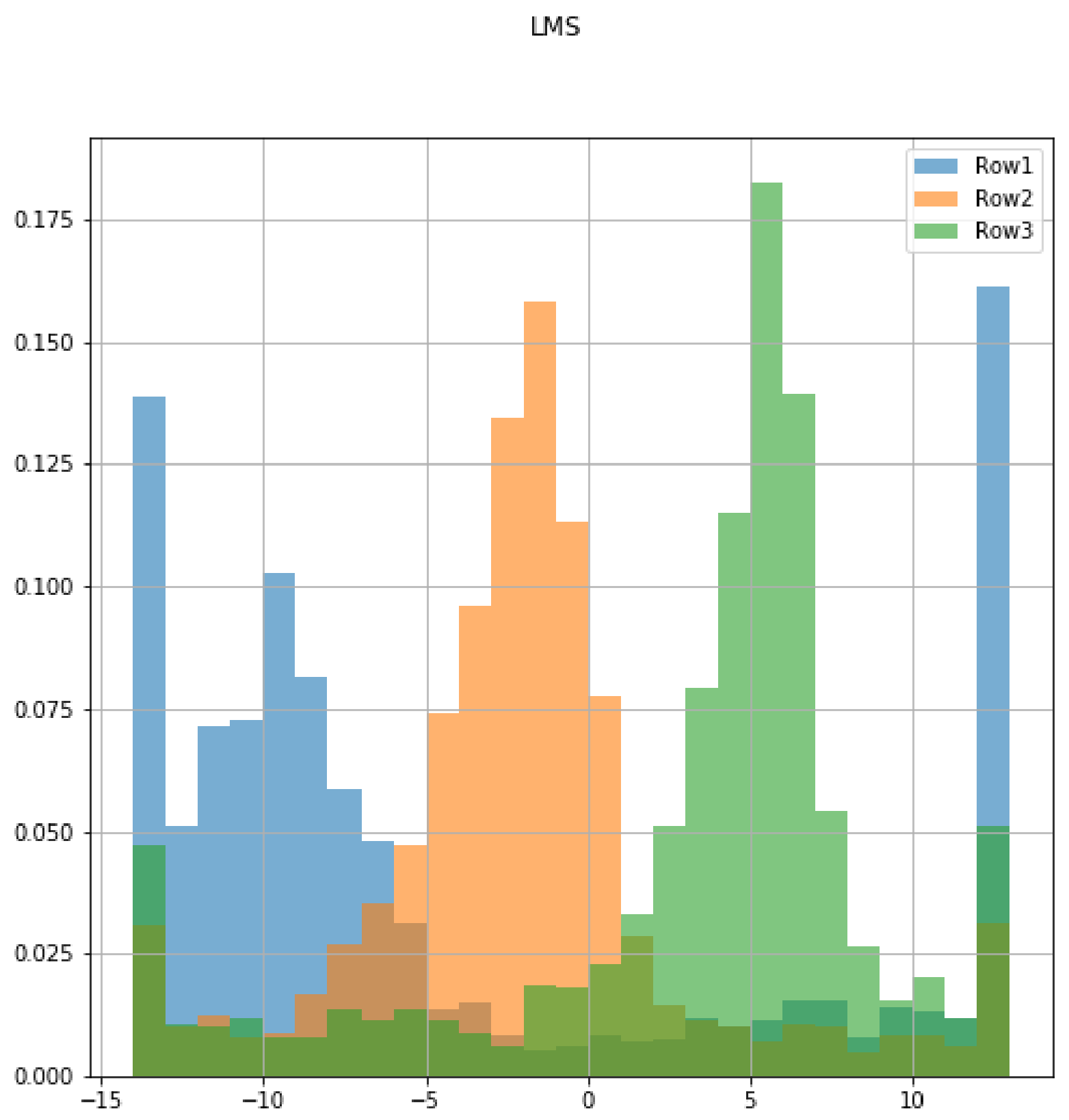

Figure: In practice the best we can do is recognise taps on different pin rows.

PIN entry acoustic attack

Table: PIN Attack performance comparison. We report the best performing classifiers in single and double configurations.

Attack by	set size	10^{th} try	20^{th} try
Our best single	50	42%	50%
Aviv et al. [1]	50	55%	-
Our best double	50	**55**%	61%
Simon and Anderson[6]	50	61%	84%
Spreitzer [7]	50	79%	-
Shukla [5]	50	94%	-
Our best single	100	41%	49%
Simon and Anderson[6]	100	48%	58%
Our best double	100	**51**%	59%
Our best single	150	40%	48%
Simon and Anderson[6]	150	44%	53%
Our best double	150	**52**%	61%
Simon and Anderson[6]	200	40%	53%
Our best single	200	43%	48%
Our best double	200	**53**%	61%

Soft-keyboard acoustic attack

Table: 27 corn-cob words of size 7-13 benchmark. We report the best performing classifiers in single and double configurations.

Attack by	10-attempts	50-attempts
Phone/best single	21%	30%
Phone/best double	25%	34%
Marquardt et al.[4]	43%	56%
Berger et al. [2]	43%	73%
Tablet/best single	43%	55%
Liu et al.[3]	63%	82%
Sun et al.[8]	63%	93%
Tablet/best double	70%	80%

What does that mean?

- Microphones provide comparable accuracy to existent side channel attacks, despite being purely acoustics based.

Can we make the attack better?

Language models can aid the performance of text prediction!

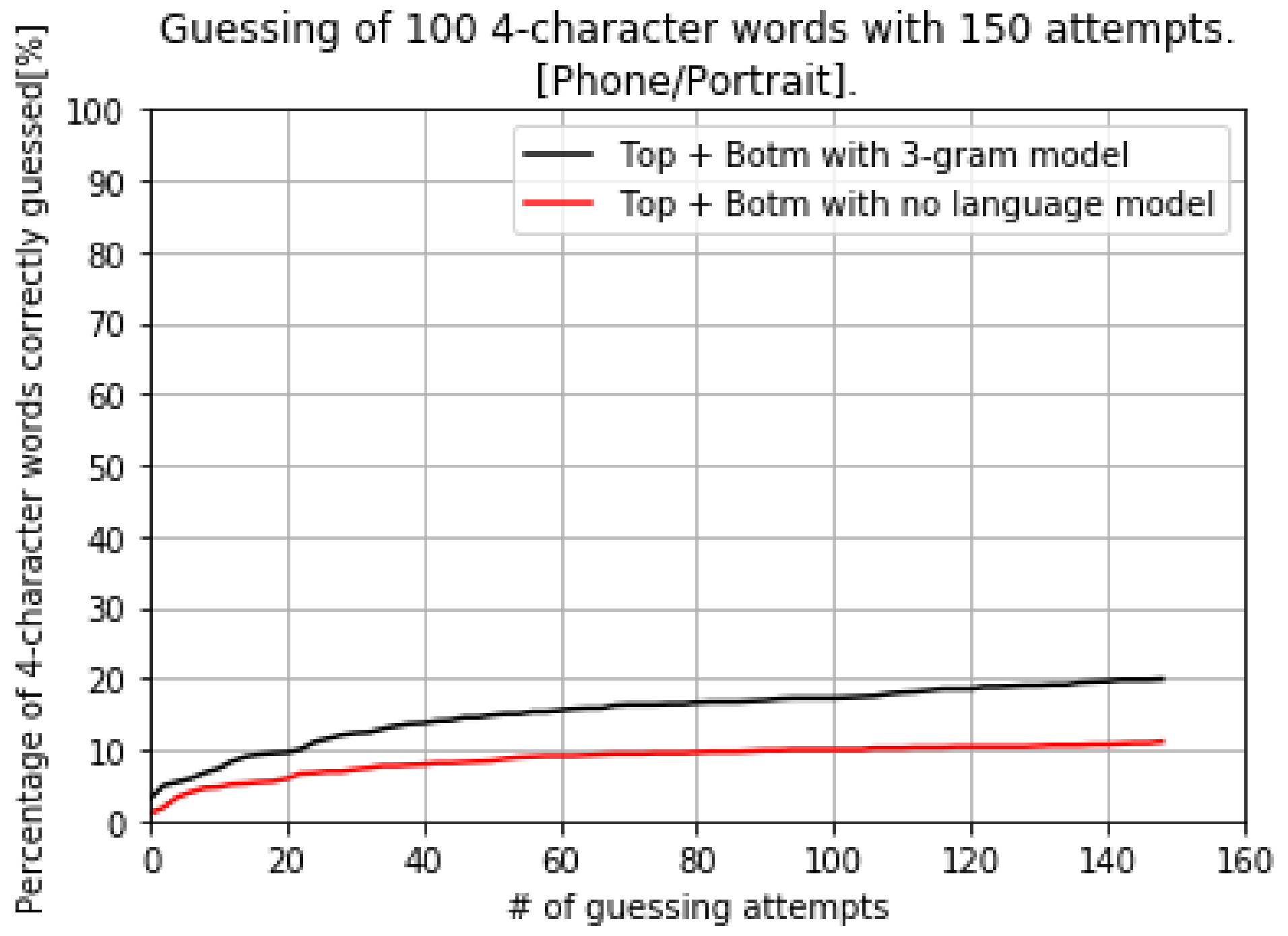

Figure: Use of language model to aid classification.

Conclusion

- Yet again the hardware configuration is underestimated
- Protection mechanisms are fairly hard to design, however, a simple capability for stereo audio access should make the attack less scary
- We believe that there is a need for *secure attention sequence* mode to be introduced to modern smartphones

References

[1] A. J. Aviv, B. Sapp, M. Blaze, and J. M. Smith. Practicality of accelerometer side channels on smartphones.

[2] Y. Berger, A. Wool, and A. Yeredor. Dictionary attacks using keyboard acoustic emanations.

[3] X. Liu, Z. Zhou, W. Diao, Z. Li, and K. Zhang. When good becomes evil: Keystroke inference with smartwatch.

[4] P. Marquardt, A. Verma, H. Carter, and P. Traynor. (sp)iphone: Decoding vibrations from nearby keyboards using mobile phone accelerometers.

[5] D. Shukla, R. Kumar, A. Serwadda, and V. V. Phoha. Beware, your hands reveal your secrets!

[6] L. Simon and R. Anderson. Pin skimmer: Inferring pins through the camera and microphone.

[7] R. Spreitzer. Pin skimming: Exploiting the ambient-light sensor in mobile devices.

[8] J. Sun, X. Jin, Y. Chen, J. Zhang, Y. Zhang, and R. Zhang. Visible: Video-assisted keystroke inference from tablet backside motion. 2016.