Bachelor thesis proposal

Julian van der Horst

October 31, 2022

My thesis topic will be about Backdoor attacks on Deep neural networks. My focus will be on a backdoor on speech recognition using a tone that humans can't hear. I will combine the work of [1] and [2] to create a backdoor attack that is only audible to a microphone and will be in the range of human hearing. This means that even when applying filters, to filter out the ultrasonic frequencies, the trigger will still be recorded.

Backdoor attacks on deep neural networks are ways to make the deep learning model behave normally when normal data is provided, but it will behave differently when special poisoned data is provided. The poisoned data in my case will be the tone which be the result of two ultrasonic tones.

This backdoor can be practically undetectable and therefore has a lot of impact.

References

[1] Stefanos Koffas et al. *Can You Hear It? Backdoor Attacks via Ultrasonic Triggers*. 2021. DOI: 10.48550/ARXIV.2107.14569. URL: https://arxiv.org/abs/2107.14569.

[2] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. "BackDoor: Making Microphones Hear Inaudible Sounds". In: *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. MobiSys '17. Niagara Falls, New York, USA: Association for Computing Machinery, 2017, pp. 2–14. ISBN: 9781450349284. DOI: 10.1145/3081333.3081366. URL: https://doi.org/10.1145/3081333.3081366.