Radboud University Nijmegen

Faculty of Science

Backdoor attack on deep neural networks using inaudible triggers

Dolphin attack trigger

Thesis BSc Computing Science

Author:
Julian van der Horst

Supervisor:
Stjepan Picek
Stefanos Koffas

December 2022

Contents

1 Introduction

2 Background

2.1 Automatic Speech Recognition (ASR)

Introduce speech to text, today the main way to do this translation is by using deep neural networks. Introduce mfcc's and the reasoning behind them.

2.2 Backdoor attacks

Introduce the idea of a backdoor attack and especially with audio neural networks

2.3 Microphone

Explain shortly how modern microphoens work and why a MEMS michrophone is special

2.4 BackDoor

[1]

Explain the idea of the BackDoor paper and how we will create the trigger

2.5 Threat model

Explain the transmitter, reciever and gray box data poisening. Also add

3 Method

References

[1] 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, pages 2–14, New York, NY, USA, June 2017. Association for Computing Machinery.