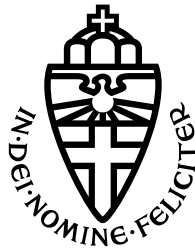


RADBOD UNIVERSITY NIJMEGEN



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
2	Background	2
2.1	Automatic Speech Recognition (ASR)	2
2.2	Backdoor attacks	2
2.3	Microphone	2
2.4	BackDoor	2
2.5	Threat model	2
3	Method	2

1 Introduction

2 Background

2.1 Automatic Speech Recognition (ASR)

2.2 Backdoor attacks

2.3 Microphone

2.4 BackDoor

[1]

2.5 Threat model

3 Method

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.

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

Explain shortly how modern microphones work and why a MEMS microphone is special

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

Explain the transmitter, receiver and gray box data poisoning. Also add

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.