Privacy Preserving Adversarial Perturbations for Neural Decoders

Research Area
Machine Learning for Communication Systems

Keywords
Adversarial examples, deep learning, privacy

Description
Replacing conventional decoders with neural networks has become a popular research topic in recent years, due to the potentially low complexity of neural networks when compared with classical iterative decoders. Nevertheless, it has been shown that neural networks are particularly unstable when maliciously designed noise is added to their inputs. Such noise is known as adversarial noise or adversarial perturbation.

In this thesis we will take advantage of this phenomena to enhance the security of a communication system. Our system model consists on three agents: Bob, Alice and Eve. Bob wants to send a message to Alice without letting Eve decode it. To do so, Bob adds adversarial noise on his transmitted signal that prevents Eve's decoder from decoding the message, while at the same time allowing Alice to decode it without problems. A summary of this system model is shown in Figure 1.

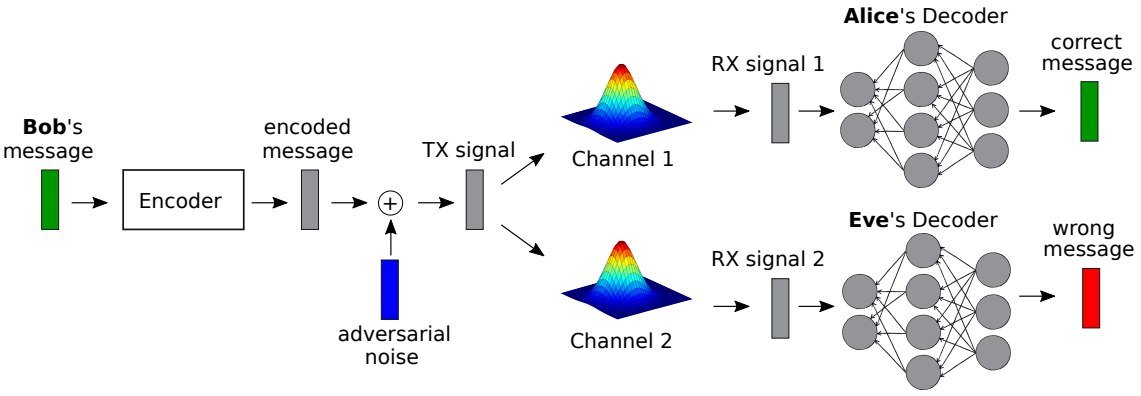

Figure 1: System Model.

The objective of this thesis is to design adversarial noise that is able to provide secrecy for Bob and Alice in different scenarios. This is carried out by modifying well known methods (for generating adversarial perturbations) for this purpose.

Requirements
- Solid experience in Python programming.
- Good understanding of basic linear algebra concepts.
- Introductory knowledge about tensorflow.

Contact
- Emilio Balda ✉ balda@ti.rwth-aachen.de ☎ +49 241 80 27704