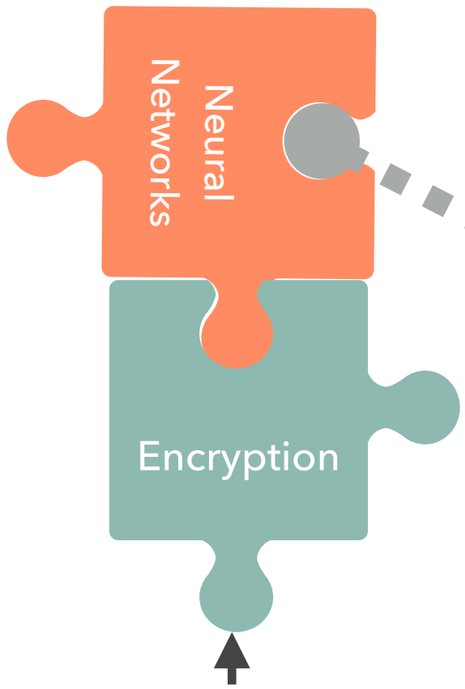
Problem Statement:

We here discuss whether we can employ neural nets to secure our communication channels. We take classic Alice, Bob and Eve example where Alice is trying to send a message to Bob and Eve is eavesdropping. Instead of using rigid symmetric encryption algorithms like AES, DES, we will create an adversarial neural network where Alice will be trained to encrypt using a shared key, Bob will be trained to decrypt using a shared key and Eve will be trained to reconstruct the message without a shared key.

Indian Institute of Information Technology, Nagpur



Can Neural Networks learn to use secret keys to protect information?

1 How Does this fit in our System?

The Architecture of Alice, Bob, and Eve

Alice, Bob, and Eve are all neural networks, message P is an input to Alice. Alice processes this input, to produce an output C. Both Bob and Eve receive C and attempt to recover P. We represent what they compute by P-Bob and P-Eve, respectively. Alice and Bob have an advantage over Eve: they share secret key K.

2 Related Past research

Literature Survey

Paper Learning to protect communications with adversarial neural cryptography by Martin Abadi and David G. Andersen

3 Implemented Model

The Architecture of Model

Each network has a Fully Connected layer of size 2n (n = number of message/key bits) which is then followed by six 1-dimensional convolutional layers

Activation functions:
`leaky_relu, sigmoid, relu, tanh`

4 Analysis

Reconstruction errors start high. After a period of time, Alice and Bob start to communicate quite effectively, but in away that allows Eve to improve its understanding as well. Then, around step 2,000 Alice and Bob counter Eve's progress. By about step 2,500 the training goals are effectively achieved

Conclusion

neural networks can learn how to perform forms of encryption and decryption, and also how to apply these operations selectively in order to meet confidentiality goals

Final Year Project
Mentor: Dr. Nidhi Lal

LEARNING TO PROTECT COMMUNICATIONS WITH ADVERSARIAL NEURAL CRYPTOGRAPHY

Shreyash H. Turkar
BT17CSE026