Adversarial Neural Cryptography

Michael Guarino
Marist College Department of Computer Science
MSCS 630: Security Algorithms and Protocols

April 11, 2018

Abstract

Neural networks have been proven to be universal function approximators and therefore can learn to represent a great variety of functions given appropriate parameters. This property makes neural networks excellent candidates as encryption algorithms. Due to recent advancements in quantum computing technology, research in the area of non-conventional cryptographic methods is of vital importance as current cryptographic algorithms will likely be completely obsolete. This work looks to build on the Google Brain paper ”Learning to Protect Communications with Adversarial Neural Cryptography” through exploring the novel use of neural networks in the area of symmetric encryption systems and to improve the original adversarial architecture proposed in Adbadi, et al. [1].

1 Introduction

Cryptography is concerned with algorithms and protocols that protect the secrecy of source information as it is transferred from it’s point of origin to it’s intended destination. Encryption describes the method used to convert source data or plaintext from an intelligible form to an encoded form by which can only be made intelligible through decoding the encoded form with access to a decryption key. An encryption algorithm is said to be secure if information about the plaintext cannot be extracted from cipher texts.

Neural networks have been proven to be universal function approximators and therefore can learn to represent a great variety of functions given appropriate parameters thus neural networks are excellent candidates as encryption algorithms. There has been some work in the area of using deep learning architectures to learn to encrypt source data; however, applications of deep learning in the area of cryptography have been largely unexplored.

The Google Brain paper ”Learning to Protect Communications with Adversarial Neural Cryptography” introduces an symmetric encryption system that consists of neural networks named Alice and Bob whose goal is to communicate while limiting what a third neural network, Eve, can learn from eavesdropping on their communication. Alice and Bob are trained adversarily to defeat Eve without a pre-specificed notion of what cryptosystem they may discover to accomplish this goal [1].

Generative adversarial networks (GAN) are a generative model devised by Goodfellow et al. in 2014 [2]. The GAN architecture is characterized by two differentiable functions, neural networks, that play different roles in refining the system. One differentiable function is known as the generator and the other is the discriminator. The generator learns to produce data from a learned probability distribution. The discriminator determines if the data that was produced by the generator is valid by determining if the input comes from the generator or from the actual data set. In this work we will validate that a generative adversarial network can be used to learn symmetric encryption specifically shared-key encryption.

The GAN architecture is known to be difficult to train being characterized as largely unstable for many reasons [4,5,6] the observation being that gradient descent used to update the parameters of the generator and discriminator are inappropriate when the solution to the optimization problem posed by GAN training actually constitutes a saddle point [3]. Recently there has been great improvements in theoretical explanation of GANs, improved GAN architectures, and training tricks to ensure that GANs converge.

This work aims to explore neural networks use in symmetric encryption systems and applying work done in the advancement of GANs to reduce computational requirements and or improve reliability of convergence in the architecture proposed by Adbadi, et al. [1].

2 Related Work

There has been some work in the area of using deep learning architectures to learn to encrypt source data; however, applications of deep learning in the area of cryptography have been largely unexplored. There has been more work in the area of using nerual networks to compromise encryption systems and to exploit vulnerable machine learning classifiers [7,8]. The Google Brain team paper "Learning to Protect Communications with Adversarial Neural Cryptography" is the only related work in the area of symmetric encryption system using neural networks to encrypt communications.

3 Methodology

Coming soon...

4 Experiments

Coming soon...

5 Discussion and Analysis

Coming soon...

6 Conclusion

Coming soon...

7 Bibliography

[1] Abadi, Martin, et al., "Learning to Protect Communications with Adversarial Neural Cryptography". arxiv:1610.06918, October 2016.

[2] Goodfellow, Ian, et al., "Generative Adversarial Networks". arxiv:1406.2661, June 2014.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, in Advances in Neural Information Processing Systems, 2014, pp. 26722680

[4] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, Improved techniques for training gans, in Advances in Neural Information Processing Systems, 2016, pp. 22262234

[5] M. Arjovsky and L. Bottou, Towards principled methods for training generative adversarial networks, NIPS 2016 Workshop on Adversarial Training, 2016.

[6] A. Radford, L. Metz, and S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, in Proceedings of the 5th International Conference on Learning Representations (ICLR) - workshop track, 2016

[7] Google Brain, NIPS 2017: Non-targeted Adversarial Attack. https://www.kaggle.com/c/nips
2017-non-targeted-adversarial-attack.

[8] Greydanus, Sam, "Learning the Enigma with Recurrent Neural Networks", arxiv:1708.07576v2, September 2017.