

DeepRing: Protecting Deep Neural Network with Blockchain

Akhil Goel , Akshay Agarwal , Mayank Vatsa , Richa Singh , and Nalini Ratha[†]
 IIIT Delhi and [†]IBM Research, NY, USA

Fakhi115126, akshaya, mayank, rsingh6@iiitd.ac.in, [†]ratha@us.ibm.com

Abstract

Several computer vision applications such as object detection and face recognition have started to completely rely on deep learning based architectures. These architectures, when paired with appropriate loss functions and optimizers, produce state-of-the-art results in a myriad of problems. On the other hand, with the advent of “blockchain”, the cyber-security industry has developed a new sense of trust which was earlier missing from both the technical and commercial perspectives. Employment of cryptographic hash, as well as symmetric/asymmetric encryption and decryption algorithms, ensure security without any human intervention (i.e., centralized authority). In this research, we present the synergy between the best of both these worlds. We first propose a model which uses the learned parameters of a typical deep neural network and is secured from external adversaries by cryptography and blockchain technology. As the second contribution of the proposed research, a new parameter tampering attack is proposed to properly justify the role of blockchain in machine learning.

1. Introduction

The current era of artificial intelligence and machine learning is converting several dreams to reality. AI systems are getting implemented for making recommendations in social media and e-commerce sites to assisting medical professionals in medical diagnosis and robotic surgeries, and defending personnel with technologies such as drone surveillance. Such a wide spectrum usage of these technologies requires that the algorithms are secure.

A lot of this success can be attributed to deep learning architectures such as Convolutional Neural Networks (CNN) [13, 14]. CNNs contain blocks where each block can be referred to either as a convolutional layer or a combination of the convolutional, pooling, and non-linearity layers. The first layer which is an input layer passes the input samples to the first block of CNN, and this way information passes through the network to the last layer which makes the decision. For secure and correct use of these AI systems, fault-

Figure 1. Vulnerabilities of artificial intelligence network and incorporation of blockchain for security.

less authentication of each block is a necessity. In other words, the accountability of each block which is missing in the original CNN models might be provided with the combination of the blockchain. Blockchain with its feature of data privacy, transparency, security, and authentication can help in the secure deployment of AI systems in the public domain. The data privacy in an AI system can be referred to as some information which is hidden from the general public which can be decrypted only using the private key of the authenticated owner of the system. On the other hand, the security aspect can be thought of as a guard who checks whether there has been any manipulation in the network architecture or not. The authentication feature can be referred to the property that the decision made by a particular block of the AI model would require the validation of other blocks connected with the block in concern.

Figure 1 shows the vulnerabilities of a typical artificial intelligence system. The attack on an AI system can be performed at an input level, architecture level, and decision level [17]. With the correct deployment of blockchain technology, attacks at the architecture and the decision levels can be avoided. For example, a recent algorithm,

