PRINCIPLES OF INFORMATION SECURITY

TEAM NAME : INFOSEC

Table of Contents

1

LITERATURE REVIEW SUMMARY

ZKCNN: ZERO KNOWLEDGE PROOFS FOR CONVOLUTIONAL NEURAL NETWORK PREDICTIONS AND ACCURACY

This paper focuses on applying ZKPs to the field of machine learning. It explores how to create ZKPs that prove the correctness of predictions made by a convolutional neural network (CNN) without revealing the network's internal details or the training data used.

zkCNN, which stands for zero-knowledge proofs for convolutional neural network predictions and accuracy, is a relatively new technique in the field of privacy-preserving machine learning.

Here are some key benefits of zkCNN:

- **Proving Prediction Origin**: Given a data point and a corresponding prediction made by the CNN model, zkCNN allows generating a proof that convinces the verifier that the prediction indeed originated from the specific CNN model in question. This is achieved without revealing any details about the model's architecture, weights, or training data.
- **Public Dataset Accuracy Proofs**: zkCNN can also be used to demonstrate the accuracy of a CNN model on a publicly available dataset. This is useful when you want to convince someone about the model's performance on a specific dataset without compromising the confidentiality of the model itself. The verifier can gain confidence in the model's generalizability without learning anything about the inner workings of the model.

The system achieves its efficiency through two key techniques:

- **New Protocol for Mathematical Proofs:** zkCNN introduces a new protocol for proving mathematical operations like Fast Fourier Transforms (FFTs) and convolutions. These operations are essential computations within CNNs. The new protocol allows for efficient proofs that convince the verifier of the correctness of these computations without revealing the underlying data.
- **zk-SNARKs for Succinct Proofs**: zkCNN leverages zero-knowledge Succinct Non-interactive ARgument of Knowledge (zk-SNARKs) schemes. These are cryptographic tools that enable generation of very concise proofs and fast verification times. In zkCNN, zk-SNARKs are used to create proofs regarding the model's computations on the input data point, leading to the predicted output.

2

ACADEMIC TRENDS

Based on your project statement, your topic aligns with research on applying zero-knowledge proofs (ZKPs) to deep learning models, specifically focusing on Denoising Autoencoders (DAEs). Here's a breakdown of relevant areas in current academic work:

- **Privacy-Preserving Denoising Autoencoders:** Similar to your project's goal, research explores using ZKPs to prove the functionality of DAEs without revealing the internal details of the model or the training data. This is crucial when dealing with sensitive data, as traditional DAEs expose information during the training process.
- **Verifiable Denoising Performance:** Current research investigates ZKPs to demonstrate the denoising efficiency of a DAE model. This allows verification of the model's ability to remove noise from images without revealing the model itself or the original noise injection process.

- **ZKPs for Secure DAE Training:** Academia is exploring ZKPs to enable secure collaborative training of DAEs. This allows multiple parties to contribute data to the training process while keeping their individual data private. The ZKPs ensure each party contributes valid data without revealing its content.

Additional Considerations:

- Current research focuses on applying ZKPs to different deep learning architectures beyond DAEs. You might consider exploring the applicability of existing ZKP techniques designed for convolutional neural networks (CNNs) to DAEs.
- Efficiency is a challenge in ZKP implementations for deep learning models. Investigate recent advancements in zk-SNARKs (zero-knowledge succinct non-interactive argument of knowledge) to optimize proof generation and verification times for your DAE model.

RESEARCH INTEREST

Recently, a paper was published verifying Zero Knowledge Proofs for Convolutional Neural Networks. It was one of the papers that focused on applying Zero-Knowledge Proofs for Machine Learning and Deep Learning. The main theme of the paper is about using zero-knowledge proofs (ZKPs) to ensure privacy and verifiability in the context of Convolutional Neural Networks (CNNs).

1. zk-SNARKs for Autoencoder Training and Inference:

This approach focuses on using zk-SNARKs (Zero-Knowledge Succinct Non-Interactive Arguments of Knowledge) to prove that the autoencoder training and inference processes were conducted correctly, without revealing the model parameters or the training data. This ensures the privacy of both the model and the data while maintaining verifiable results.

2. Lightweight Zero-Knowledge Proofs for Efficient Verification:

Research is ongoing to develop lightweight and efficient zero-knowledge proof systems specifically tailored for autoencoder architectures. This aims to reduce the computational overhead associated with generating and verifying proofs, making zero-knowledge autoencoders more practical for real-world applications.

3. Applications in Privacy-Preserving Machine Learning:

Researchers are actively exploring applications of zero-knowledge autoencoders in various domains, including medical imaging, financial data analysis, and genomic research. These applications leverage the privacy-preserving capabilities of zero-knowledge proofs to enable secure and trustworthy machine learning on sensitive data. These key ideas represent the current frontiers in zero-knowledge autoencoders. As research progresses, we can expect to see further advancements in efficient proof systems, secure computation techniques, and novel applications of this technology in various privacy-sensitive domains.

PROPOSED PROBLEM STATEMENT

We are planning to do our Project on Zero-Knowledge proofs for Denoising Auto Encoders. Denoising Auto Encoders are a subclass of the Encoder-Decoder Model in which we are giving Noisy Images as input and “denoising” the images, i.e. removing the noise injected in the images and then evaluating the accuracy of the classification tasks. It involves an encoder model that first projects the Overall information contained in the image to a smaller dimension (also referred to as Bottleneck Layer) which allows the model to learn the noise injected in the image and then projecting it back to the same dimension using opposite operations or the De-Convolution Operators. After this, the Model Output is tested against a standard Loss Function between Original Image and Reconstructed Image whose optimization leads to the model denoising the noise of the images. In this project, we are basically evaluating whether the output that is provided by the Auto-Encoder is indeed evaluated by it only.

Threat Model

The definition of security that we aim to achieve for the proposed solution can be modeled as an Indistinguishability Game. In this game, the Verifier will randomly choose an image, which may or may not be generated by the Auto Encoder. The Distinguisher's task is to determine whether the output is a randomly generated image or an image generated by the Auto Encoder.

Network Assumptions

For simplicity, we will assume a single prover-verifier communication model. The prover owns the DAE model and generates ZKPs for the verifier. This model can be extended to a multi-party setting in future work.

Here are some of the relevant cryptographic assumptions:

- Hardness of Discrete Logarithm Problem: zkCNN relies on the difficulty of computing the order of random elements within a specific mathematical group. This assumption is considered secure if solving the Discrete Logarithm Problem (DLP) is intractable. The RSA assumption hinges on this very hardness [1].
- Resistance to Hash Collisions: The scheme also assumes that finding collisions in a cryptographic hash function is infeasible. In other words, it should be extremely difficult to find two different inputs that result in the same hash output.

These cryptographic assumptions are standard in building secure zero-knowledge proofs and are not specific to any network.

5

ESSENTIAL READINGS

ZERO KNOWLEDGE IN CNN

zkCNN lets someone prove a convolutional neural network's (CNN) prediction is correct without revealing the model itself. It uses zero-knowledge proofs, where a "prover" convinces a "verifier" of a computation result (prediction) without leaking the secret input (model). zkCNN even allows proving a hidden CNN's accuracy on public data! This relies on a special zk-SNARK proof system for fast computations within CNNs.
https://eprint.iacr.org/2021/673.pdf

AUTO ENCODERS

An autoencoder is a neural network that learns to compress data into a smaller representation and then recreate it. Imagine it like a sketch artist - it captures the essence of the data (like a face) in a smaller form (the sketch) and uses that to rebuild (draw) a good approximation of the original. Variants exist! Some force the compressed data to be sparse (few active neurons) for better feature extraction, while others might focus on specific data aspects like color or edges.
https://ieeexplore.ieee.org/document/8616075

CNN

A Convolutional Neural Network (CNN) is an image recognition pro inspired by animal vision. It works like a layered filter system. Small filters scan the image, finding edges and shapes. These are stacked to build more complex features. Later layers combine these features to identify objects, even in different locations. This layered approach lets CNNs "learn" to see!
https://ieeexplore.ieee.org/document/8308186

PROJECT PLAN

STEP-1:

KNOWLEDGE UNDERSTANDING

- Firstly understand the zkCNN Zero Knowledge Proofs and its functionalities.
- Then, we will research about Autoencoders, a type of ML model, examining existing research on integrating zero-knowledge proofs with them to uncover established methods and potential challenges.

STEP-2:

DEVELOPING ZKP FRAMEWORK

- Design a ZKP scheme that caters to zk-Autoencoders, focusing on proving specific properties without revealing the model itself.
- Formalize the ZKP scheme by translating our Zero Knowledge Proofs system into precise mathematical structure to ensure both security and efficiency of predictions and accuracy.

THIS PROJECT PLAN HELPS US IN DEVELOPING A ZKP SCHEME THAT PRESERVES THE PRIVACY AND EFFICIENCY OF AUTO ENCODERS

THANK YOU

AKHIL GUPTA
2021101012

HARSH BANSAL
2021101027

MANUJ GARG
2021101047

PRANAV GUPTA
2021101095

YASH SHIVHARE
2021101105