Information Security

Spring 2022

Project Proposal

Topic: Backdoor Detection

Group Members:

Joshua Naeem 22-10367

Eraj Khurshid 22-10457

Project Description:

We are working on the backdoor detection. Deep neural networks (DNNs) are proved to be vulnerable against backdoor attacks. A backdoor is often embedded in the target DNNs through injecting a backdoor trigger into training examples, which can cause the target DNNs misclassify an input attached with the backdoor trigger. A backdoor is a mechanism surreptitiously introduced into a computer system to facilitate unauthorized access to the system. While backdoors can be installed for accessing a variety of services, of particular interest for network security are ones that provide interactive access. These are often installed by attackers who have compromised a system to ease their subsequent return to the system.

We will be using CIFAR models and as well poisoned datasets to identify the backdoor if it is created. Our program will tell us the detected backdoor and will display some plots to help us visualize the poisoned dataset and as well clean dataset.

Existing backdoor detection methods often require the access to the original poisoned training data, the parameters of the target DNNs, or the predictive confidence for each given input, which are impractical in many real-world applications, e.g., on-device deployed DNNs. We address the black-box hard-label backdoor detection problem where the DNN is fully black-box and only its final output label is accessible.

Steps:

To implement our project, we need to first train the dataset on any notebooks available on the internet like Jupyter, Google Colab etc. After that we will implement those trained dataset into our code and then it will help us identify the backdoor if there is any. Based on this observation, we will then try to improve the accuracy of the trained dataset and we can then try to find any backdoors hidden in the networks as well. Tools that we will use will be Vs Code for IDE, Jupyter Notebook for training dataset, Kali Linux and Routersploit if we needed to check vulnerabilities and backdoors in a network.

Outcome:

This project will help us to secure more of our devices from trojan horses like backdoor, we can detect backdoors and we will be able to improve on the code and try to bring fresh perspective on why backdoors are important to identify as we get connected to internet many of our devices if they have backdoor, they can be accessed by hackers online. To make this project stand out importance of why creating a backdoor is unethical and it can be dangerous as if you get detected using a backdoor there are many consequences.