BTP REPORT

Name: Kandi Vishnu Vardhan Reddy

Roll No: 18CS30022

Privacy Breach in Deep Learning through Computer Architecture

Motivation:

The rapidly emerging Deep Learning technology has recently triggered a substantial amount of interests in the computer security community. Since the Success of such applications depends on the amount of Data used to efficiently DL models, Data breaches are one of the top cybersecurity problems affecting the digital economy.

Project Definition:

There are lots of possible threats to deep learning for intentional or unintentional exposure of sensitive information. This information can be the training data, inference queries or model parameters or hyperparameters.

As opposed to attacks that exploit vulnerabilities in software or algorithm implementations, side channel attacks utilize information leaks from vulnerabilities in the implementation of computer systems. In this project we investigate the attacks on DL applications which use computer "architectural explorations".

Examples of Attack:

Here our adversary may not need to query the victim model (Deep Recon).

Deep Recon : An attack that reconstructs the architecture of the victim network using the internal information extracted via Flush+Reload, a cache side-channel technique. Once the attacker observes function invocations that map directly to architecture attributes of the victim network, the attacker can reconstruct the victim's entire network architecture.

Due to modern microprocessor architecture that shares the last-level cache (L3 cache) between CPU cores,these cache side-channel attacks have become more readily available to implement.

Or by using a constant number of queries.

Model Extraction via Timing Side Channels: A model extraction attack in a black box setting where he exploits timing side channels and efficiently reconstructing a substitute model architecture with functionality close to the target model.

Work Done and Further Plans:

Explored different attacks in cryptography and also the required topics in neural networks to understand and get a better view at different research papers. Went through different research papers to get to know about the project. Working on Implementation hasn't been completely done but the analysis of the documentation and prerequisites required to work on the project are completed.

Further plan for the project is to implement the attacks encountered till now and to get know more about side-channel attacks and prevent them.

Sources:

PRIVACY IN DEEP LEARNING: A SURVEY (Fateme Sadat Mireshghallah , Mohammad Kazem Taram, Praneeth Vepakomma , Abhishek Singh , Ramesh Raskar , Hadi Esmaeilzadeh : University of California San Diego, Massachusetts Institute of Technology)

SECURITY ANALYSIS OF DEEP NEURAL NETWORKS OPERATING IN THE PRESENCE OF CACHE SIDECHANNEL ATTACKS (Sanghyun Hong, Michael Davinroy, Yigitcan Kaya, Stuart Nevans Locke, Ian Rackow, Kevin Kulda, Dana Dachman-Soled, Tudor Dumitras: University of Maryland Swarthmore College Rochester Institute of Technology Blair High School Baylor University)

STEALING NEURAL NETWORKS VIA TIMING SIDE CHANNELS (Vasisht Duddu, Debasis Samanta, D. Vijay Rao, Valentina E. Balas : Indraprastha Institute of Information Technology, Delhi, India Indian Institute of Technology, Kharagpur, India Institute for Systems Studies and Analyses, Delhi, India Aurel Vlaicu University of Arad, Arad, Romania)

MASTIK : A Micro-Architectural Side-Channel Toolkit (Yuval Yarom: The University of Adelaide and Data, CSIRO Adelaide, Australia)