


Ideation Phase

Brainstorm & Idea Prioritization Template

Date	26 th October 2023
Team ID	Team-593456
Project Name	Project - Adversarial Attacks and Defenses
Maximum Marks	4 Marks

Brainstorming Protocol for Adversarial Attacks and Defenses



Brainstorm & Idea prioritization

Problem Statement
Discussing the pros, cons and various perspectives of Adversarial Attacks and Defenses.

PROBLEM: ADVERSARIAL ATTACKS AND ADVERSARIAL DEFENSES

Rules we're following while brainstorming
To run an smooth and productive session

- Stay on topic.
- Encourage wild ideas.
- Defer judgment.
- Listen to others.
- Go for volume.

Brainstorm
Noting ideas provided by the entire team.

Person 1

- Security Information and Event Management (SIEM):**
Use SIEM tools to collect, analyze, and correlate log data from various sources.
- Denial-of-Service (DoS) Protection:**
Implement DoS protection mechanisms to prevent or mitigate the impact of DoS attacks.
- Network Segmentation:**
Divide the network into segments to limit the spread of a potential attack, in case a network segment is compromised.

Person 2

- Threat modeling**
– Formalize the attacker's goals and capabilities to the target system.
- Attack simulation** – Formalize the optimization problem the attacker tries to solve according to possible attack strategies.
- Information laundering**
– Alter the information received by adversaries (for model stealing attacks)

Person 3

- Feature Squeezing:**
Reducing precision of input, making it more challenging for attackers to exploit small differences.
- Training a model with the correct information and possible incorrect information in prior to increase model immunity.**
- Regular Model Updating:** training on new data and defense mechanisms consistently to adapt to new attack techniques and evolving threats.

Person 4

- Adversarial training**
: Training a model to clean and dirty data to make model more tough and immune to similar data.
- Keeping the machine learning models and their architecture private / hidden from public access**
UNLESS the project is open source.
- Multi-instance training**
: Training multiple models and combining their predictions for better resilience of a model.

Group Ideas
Sorting all determined group ideas under a common sector and segment.

Enhancing Defensive Measures

- Security Information and Event Management (SIEM):**
Use SIEM tools to collect, analyze, and correlate log data from various sources.
- Regular Model Updating:** training on new data and defense mechanisms consistently to adapt to new attack techniques and evolving threats.
- Adversarial training**
: Training a model to clean and dirty data to make model more tough and immune to similar data.

Analysing Attackers and Attack Vectors

- Threat modeling**
– Formalize the attacker's goals and capabilities to the target system.
- Training a model with the correct information and possible incorrect information in prior to increase model immunity.**
- Attack simulation** – Formalize the optimization problem the attacker tries to solve according to possible attack strategies.

Overall Security Suggestions

- Multi-instance training**
: Training multiple models and combining their predictions for better resilience of a model.
- Denial-of-Service (DoS) Protection:**
Implement DoS protection mechanisms to prevent or mitigate the impact of DoS attacks.
- Keeping the machine learning models and their architecture private / hidden from public access**
UNLESS the project is open source.

4

Prioritize

Prioritising which factors stand important and which factors lack significance.

