Adversarial Gameplay Strategies Report

Chua Sheng Xin[1]
[1]School of Information Technology, Monash University Malaysia, Kuala Lumpur, Selangor 47500 Malaysia

This report outlines the various adversarial gameplay strategies employed by the author throughout the semester. It encompasses team-based tactics, individual strategies, and powers of observation, all aimed at securing a competitive edge.

***Index Terms*—Adversarial Gameplay**

I. Introduction

THROUGHOUT the semester, my primary objective was to maximize marks and gain advantages over other students and teams. This report chronicles the varied strategies I adopted to achieve this goal.

October 21, 2023

A. Team Based Gameplay Strategies

1) Circumvention of Other Teams' Techniques

From what I have learned in the lectures and tutorials about black box attack, I attempted to develop a system that could detect vulnerabilities in other teams' defense techniques and perform attack on their models. This involves using a substitute model for DNN and then using momentum iterative method (MIM) to attack the model. However, this failed as the research paper on the black box attack did not provide any code for the model and the model that we found on another paper was not suitable for momentum iterative method. My idea if succeeded would have a decent transferability rate between models and as such would most likely succeed against other team's defense models without prior knowledge of them. After being informed that this implementation was not possible, my team decided to focus on the research paper that provided us the model and we changed to spectrum simulation attack (SSA) that could transform the input in the frequency domain. It aims to craft adversarial examples that are more transferable across various defense models, which I am sure would be great against the teams that are implementing defense models. In both versions of our team's colab (MIM and SSA), I put in effort to ensure that the code would run and the descriptions and are written by me with assistance of my helpful teammate. After learning that MIM would not work for the model in the paper regarding SSA, I changed the colab to perform SSA including: deleting the parts that import and perform MIM and utilizing the code provided in the research paper to import and perform SSA.

2) Deceptions of other teams

Unfortunately there is very little means of deception towards other teams as my team was busy trying implement the unfruitful MIM attack. But I figured it would not be very helpful to deceive other teams that are implementing defense models as a black box attack does not require information about the defense model. As such, even if the adversary changes the implementation of their model our attack should still work as improving transferability is the main focus of black box attack,

Manuscript revised October 21, 2023. Corresponding author: Chua Sheng Xin (email: [shengxin1122@gmail.com]).

B. Individual Based Gameplay Strategies

1) Blindsides

Throughout the semester I have attempted to reach out to other teams and see what adversarial gameplay strategies that I can possibly employ. However this yielded little results as our progress was slower than most of the teams because of the fact that we were simply trying something new that had little chance of success. I contacted other teams and tried to gain information on their progress and aim. Then I tried to collaborate with one of the teams which is also doing black box attack but it has not materialized due to the change from MIM to SSA. In general, it was hard for me to employ any adversarial gameplay strategies.

II. Conclusion

Adversarial gameplay requires a balance of proactive strategy and reactive adaptability. Through a combination of technical prowess, collaboration and observation, I tried to navigate the semester's challenges in order to gain maximal advantage regarding the attack and defense aspects of things. However, I think that things did not go so smooth for me as my team have met many challenges throughout the development phase of our implementationl.

Acknowledgment

I would like to acknowledge my teammate Riddhi Boodnah for her trust and collaboration. Thanks also goes to the lecturer, tutor and especially teaching assistant who provided much needed technical support when it comes to the coding aspect as I was very much unexperienced with using colab. The helps were very useful and instrumental in the development of the implementation, even though in the later days we learned that MIM was largely not possible because of compatibility.

References

[1] R. Boodnah and S. X. Chua, *Practical BlackBox Attack Against ML: Spectrum Simulation Attack*, Google Colab, [Online]. Available: https://colab.research.google.com/drive/1fWv_jGxeO7-9kFAKnSpRN-KvTep-1J4c?authuser=1#scrollTo=3ngBNydVWaj7&uniqifier=1

[2] Y. Long, *SSA: Self-Supervised Learning via Adversarial Training*, GitHub Repository, [Online]. Available: https://github.com/yuyang-long/SSA, 2023.

Chua Sheng Xin Chua Sheng Xin is an undergraduate student at Monash University Malaysia who is studying the degree for Bachelor of Advanced Computer Science. He is the author of this report submitted for the Malicious AI and Dark Side unit that delves into cybersecurity. Currently, he has been busying himself in rushing due dates and it is pretty scary to put one in a situation where one has three assignments due on the same day in the last week of the semester, life can be tough.