CAAD CTF 2018 Rules

June 21, 2018

Version 1.1

The organizer will invite 5 teams to participate CAAD CTF 2018. We will have it in Las Vegas on Aug. 10th, 2018. The rules details are below:

1. Each team will submit their defense classifier to organizer. Organizer will test the classifier with a secret image dataset, verify the accuracy, to ensure the accuracy rate above 75%. The deadline of change is Aug. 5th. Before the deadline, the defense classifier is allowed to change. After that, it cannot be modified before/in the competition. In the competition, original team logo images and some other images will randomly be sent by the organizer to the defense classifiers, to make sure the defenses are working in good conditions. If the behaviors are abnormal for testing images, judges will decide if there are cheating behaviors. If yes, the team may lose chances to win.

2. Organizer will prepare 5 different images which can be classified correctly by all the participants submit defense classifiers. The images will be kept secret before the competition. The whole competition will be divided into 5 rounds. In each round, each team will get a different image. For example, team 1 gets an image of cat. Team 2 gets an

image of dog. Then, each team will use their own image as source image to create adversarial examples. For example, team 1 will modify the cat image, send the modified image to team 2 with the purpose that team 2 will classify the image as a dog. If indeed the defense classifier of team 2 recognizes the image as a dog, the attack is successful. Team 1 gets score and team 2 loses score. If it is not recognized as a dog, the attack fails. No team gets or loses score.

3. When participant creates adversarial examples, different perturbation value can be used. They are 32, 16, 8, 4, 2 and 1. Different value means different score. The smaller the perturbation value is, the bigger the score. If team 1 attacks team 2 successfully with perturbation value 32 first, after that, team 1 tries perturbation 8 and succeed again, team 1 will get score for perturbation 8. The score for perturbation 32 will be deducted. Each time an attacker team gets score, same score will be deducted from the defender team.

Perturbation Value	32	16	8	4	2	1
Score	10	15	25	40	60	100

4. Each round lasts for 30 minutes. The total score of 5 rounds will determine the final score. Teams will have prizes based on the score.

5. In the competition, organizer will be assigned a virtual machine with 24GB RAM and a single NVIDIA Tesla P100 board in cloud. Participants can use the VM to perform attack. It is not allowed to use other compute resources during the competition.

Detailed requirements of defenses

1. The defense classifiers will run on machines with 24GB RAM and a single NVIDIA Tesla P100 board in cloud. It is required the defense can classify one image in 3 seconds.

2. The defense classifiers are required to have accuracy rate above 75% against a secret test dataset.

3. The defense classifiers can run in a Docker container. The Docker image can be any which can be downloaded from public Internet. The defense classifier cannot use network.

4. The defense can monitor an input folder, when there is a new image exists, it will classify the image and put result in the output folder. The input and output folders are mapped folders on host. The exact folder paths will be passed to the defenses by parameters.

Detailed requirements of attack

1. In the competition, participant will send attack packet to a controller machine. Each attack contains information: target team, adversarial example image, perturbation value. Controller will forward the attack to corresponding defense, then respond attacker with information: attack result (success or fail), class label the defense classified.

2. Each team can perform attack on other teams at the same time. Attacking its own defense is not allowed. For a same target team, it is not allowed to attack too frequently. Controller will control the frequency. 6 seconds are required interval time between two attacks to same target.

Prize

First Prize, 1 team, USD$ 5000

Second Prize, 1 team, USD$ 3000

Third Prize, 3 teams, USD$ 1000

Team requirement

Each team can have up to 5 members, but only 2 members can participate the competition at Las Vegas.

Organizer will cover traffic and hotel fees for 2 team members.

Registration

We are now inviting teams to participate this CAAD CTF. If you are interested in it, please send email to us at caad@geekpwn.org before June 30th, please also introduce what you did in the area of Adversarial Examples in the email. Judges will decide the invitation list.

To know more about GeekPwn, please visit http://geekpwn.org