Ledger Innovation Lab
03 rue Gretry, Paris
01/03/2022

Scope Statement:
Decentralized Test of Liveness

Calic Petar, Master ANDROIDE, Sorbonne Université
Supervised by: Nicolas Bacca, Co-Founder, VP Innovation Lab

SUMMARY

1/ Introduction

This project is to be realized by Petar Calic (me) under the supervision of Nicolas Bacca, VP of Innovation lab at Ledger, in the context of an internship of 6 months. I will be working mainly alone on the project with assistance and guidance from Mr Bacca, the innovation team, the Data team, and many UPMC researchers that I will visit for consultations.
Know Your Customer (KYC) procedures are a critical function to assess customer risk and a legal requirement to comply with Anti-Money Laundering (AML) laws. Effective KYC involves knowing a customer's identity, their financial activities and the risk they pose.
Tests of Humanity are a subgroup of tests designated to distinguish real and fake persons (bots). It consists of a set of technical features to counter biometric spoofing attacks where a bot imitates a person’s biometrics or characteristics in order to deceive or bypass the identification and authentication steps given by the system. With the emergence of the internet, and with it, the daily increase of our everyday life dependence on its functionalities, increases as well the targets and vulnerabilities of third party malicious intentions. Here is one personal example: I made some digital art images that I wanted to give away freely to people that attended an online event. Two seconds after I shared the link, a bot took all 100 of the tokens. A simple humanity check could prevent that. With AI capabilities to imitate humans rising every day, the task to discover bots that don't wish to be discovered, becomes more and more challenging.
However, there are still very successful tests like Google's CAPTCHA which relies on secret images recognition, mouse movements and a top secret server classifier. Our goal is to make an Test of liveness that is : Open source (No security by obscurity), Decentralized (The test won’t be running on a local server but on an autonomous smart contract deployed on STARKNET, that would be accessible to anyone, and more user friendly than the existing tests.
There is also a more profound and general mission behind the project which is to make

advances in the field of decentralized technologies and use its potential. We should also start relying more on technologies that are verifiable rather than trustable.
An important note is that a test of liveness based on machine learning will never be 100% bot proof as the constant race of technologies between the two sides never stops (Not to mention the existence of human farms for solving CAPTCHAs). In order to be completely sure that we are dealing with a real person the only way is to request a document verification for really critical systems. That means that the development of a 100% bot detection rate algorithm is not the goal of this internship rather than the exploration of new technologies and exploiting their potential.

2/ Image analysis

- Experiment with different techniques of Image processing : Noising, Denoising, rotations, cropping for Adversarial Image Preprocessing
- Find a threshold where humans still easily recognise the image, and ML algorithms hesitate.
- Minimum 1000 images of different classes.
- Minimum 9 classes
- The nature of the test should play with the Human vision characteristics.

Test difficulty for Human and Machine (For human test done on Thomas and me, ML performance results of experiments of Article by Nazare (2017)):

- MNIST with lot of noise (sigma 0.5): Human → Easy / Machine → Easy
- MNIST with lot of noise (sigma 0.6): Human → Medium/ Machine → Easy
- MNIST: Human → Easy / Machine → Easy
- CIFAR10 with noise (0.3): Human → Difficult / Machine → Difficult
- CIFAR10: Human → Medium / Machine → Medium
- SVHN with noise (0.1): Human → Medium / Machine → Easy
- SVHN with noise (0.2): Human → Difficult / Machine → Easy
- SVHN: Human → Medium / Machine → Easy

3/ App development

The V1 will be a frontend Interface that will be optimized to work well on all platforms. It will request from the backend a 9 pictures (3x3) and a question and display it.
The pictures will be on a DB made by the Infrastructure team.
The test must be operational to begin V2
The V2 will have the DB of pictures stored on IPFS, and the back will operate on starkent via a SmartContract that will interact with the Front end interface. Chainlink for the random questions or a server(John says it's easier to begin with and cheaper).
The V3 will integrate the ML decider.

4/ Data Collection

Users that interact with the system on the V1 app will be used for building the dataset that will be used for classification training. The features need to be wisely chosen as we aren't going to be able to collect a lot of data and the information available is very restricted. We have to keep in mind that features that aren't impacting the classification are prone to make the learning process overfitting. The data will be collected approximately with this pace: 10 tests for 20 users daily → 1000 data vectors in 5 days. With 500 we can begin the cross validation tests.

For bot data we should synthesize data. We will need slightly more data because the variations of behavior are greater than for a human. Web tech study or GAN.

5/ Classifier training

Choosing a simple but effective classifier. Make synthetic responses using a GAN. More cases need to be trained: an attacking ML that sees the answers, that partially sees, that is completely blind. Or maybe choosing a classifier that would make it as hard as possible exploiting the fact that the trained classifier is open source. For example using LSTM or Cascades or Haar.

8/ Cryptographic protection of labels and classifier

The system has a vulnerability which can be exploited with medium effort if the attacker has knowledge in Machine Learning. Otherwise the test of liveness itself is very robust and we are confident that it is very hard to breach the test using individually made ML bots that recognize Objects.
The attacker needs two elements in order to breach the system: Image labels (answers to the test), and the ML classifier. If retrieved the user can train it to find a set of features which satisfies the classifier and together with visible image labels the attacker can easily “act” as a human. If we find a way to protect one or those two elements the test will be almost 100% bot proof.

As we are focusing in this project on the decentralization object, no compromise will be made and centralize some operations to make the system more secure. The V3 will be opretational and robust against front attacks. The ML Open source vulnerability will exist but a maximum will be done to make the ML inference as complicated as possible

9/ Global architecture
Global architecture

10/ Final Tests and Updates
Tests After each new version and after each merge.

11/Economic Study
Compare Cost of breach, cost of using human farms, and price of the protected asset.

12/ Bibliographie
[1]