Project Design Phase-II
Data Flow Diagram & User Stories

Date	03 October 2022
Team ID	PNT2022TMIDxxxxxx
Project Name	Malware detection and Classification
Maximum Marks	4 Marks

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

In our proposed workflow, the user initiates the process by converting executables from a specified directory into images using our provided exe-to-img script. If needed, the segregation of these executables into training, validation, and testing datasets can be simplified using the auxiliary extract-exe script. This script allows the user to copy executable files from a user-defined directory to a target directory.

It's important to note that if data segregation using the extract-exe script is required, it should be done before the conversion to images. Once the dataset is successfully segregated, the cnn-model script comes into play. This script is responsible for training the CNN Model on the prepared training and validation datasets. To validate the model's performance, the user can utilize the "Make Prediction" option within the cnn-model script.

Here's a step-by-step breakdown of the process:

1. Execute the extract-exe script to extract executable files:

   Command: `python3 extract-exe.py`

2. Utilize the exe-to-img script to convert executable files to images:

   Command: `python3 exe-to-img --folder [Source Directory] --target [Target Directory] [--benign | --malware]`
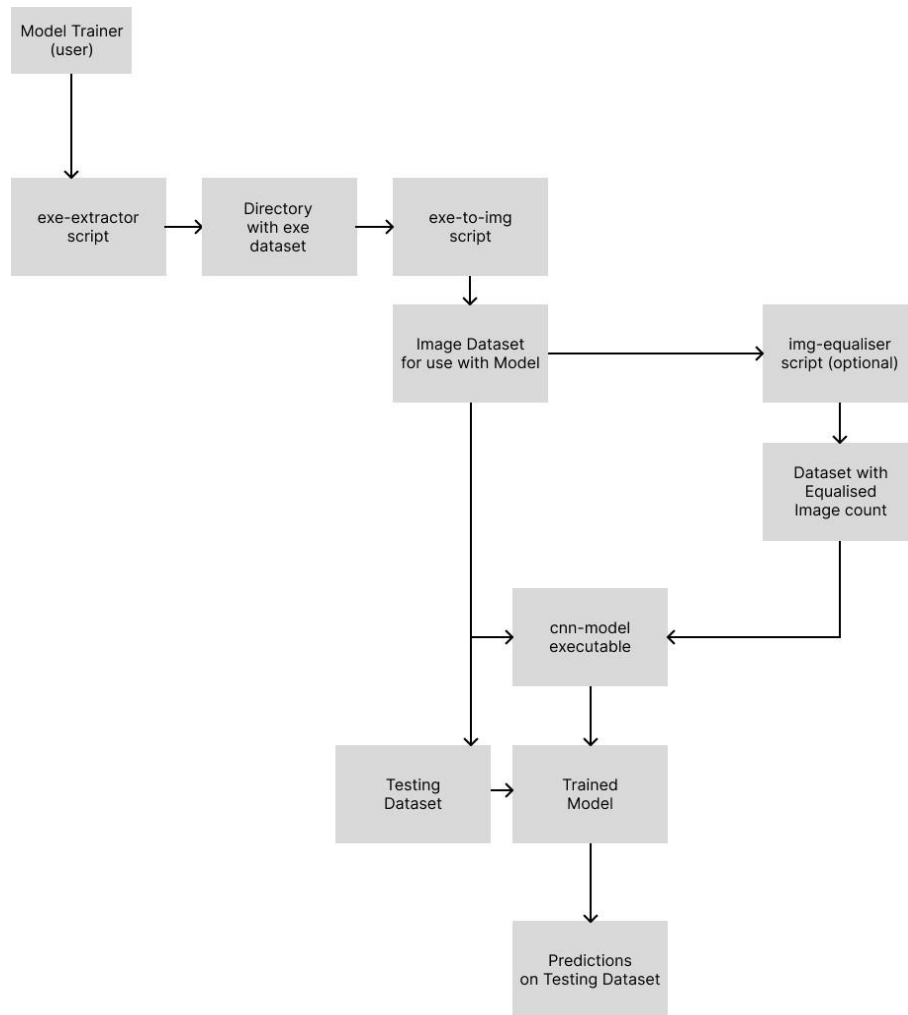
3. Run the cnn-model script for model training and validation. Use the following options in the cnn-model main menu (order is not critical):

   - Train the model using the "Train Model" option (execute it first when in refresh mode).

   - Test the model's performance with the "Make Prediction" option.

   - Modify environment variables if image storage locations differ from the defaults:

     i. Model location: `./.model/`

     ii. Training and validation dataset location: `../Generated Images/`

     iii. Testing dataset location: `../Equalised Images/`

Model Trainer
(user)

exe-extractor
script

Directory
with exe
dataset

exe-to-img
script

Image Dataset
for use with Model

img-equaliser
script (optional)

Dataset with
Equalised
Image count

cnn-model
executable

Testing
Dataset

Trained
Model

Predictions
on Testing Dataset

User Stories(Projected)

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance Criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can create an account by providing my email, setting a password, and confirming it.	After registration, I can access my dashboard.	High	Sprint-1
Customer (Mobile user)	Registration	USN-2	As a user, I should receive a confirmation email upon successful registration.	I can receive the email and click the confirmation link to activate my account.	High	Sprint-1
Customer (Mobile user)	Login	USN-3	As a user, I can log into my account by entering my email and password.	I can access my dashboard after a successful login.	High	Sprint-1
Customer (Mobile user)	Dashboard	USN-4	As a user, I can view real-time scan progress and receive notifications at the end of the scan.	I receive notifications and can monitor the scan's progress in real time.	Medium	Sprint-2
Customer (Web user)	Registration	USN-5	As a web user, I can register for the application through the web interface.	I can create an account by providing my email, setting a password, and confirming it.	High	Sprint-1
Customer (Web user)	Login	USN-6	As a web user, I can log into the application using the web interface.	I can access my dashboard after a successful login.	High	Sprint-1
Customer (Web user)	Dashboard	USN-7	As a web user, I can access and view detailed logs and real-time AI communication to monitor malware detection progress.	I can access and view the logs, and chat in real-time with an AI for explanations.	Medium	Sprint-2
Customer Care Executive	Dashboard	USN-8	As a customer care executive, I can access the dashboard to monitor the status of malware detection for users.	I can view the real-time progress of malware detection for all users.	High	Sprint-3
Customer Care Executive	Incident Response	USN-9	As a customer care executive, I can access an incident response integration feature for crisis situations.	I can access incident response features and receive real-time notifications during crises.	Medium	Sprint-4
Administrator	User Management	USN-10	As an administrator, I can manage user accounts, including creation, modification, and deactivation.	I can create, update, and deactivate user accounts.	High	Sprint-5
Administrator	Data Usage Limits	USN-11	As an administrator, I can set data usage limits and analysis options for the system.	I can configure data usage limits and analysis options.	Medium	Sprint-6