

**Chandigarh Police Hackathon 2019**

**Team name – 0x90**

**Project name – Zephyr**

**Category – Malware Analysis**

**Project description- A framework to analyse web server access logs using Machine learning and to Scan executables using YARA rules.**

**Team Members-**

**Arun Chaudhary**

**Pikesh Patel**

**Ankit Bhardwaj**

**Introduction-**

The cost of security breaches is skyrocketing, bug bounty programs are getting expensive, cybersecurity professionals are difficult to find, and developers don’t get enough time to take security training. In order to prevent security breach and information theft.

We decided to develop a tool that can detect malicious request made by the attacker to the victim’s web server, this tool is we have also used machine learning to increase the efficiency to detect malicious request, with very low rate of false positives. As each sector of the society is moving towards the digital age, it increases the threat and possibilities of malware attack which can harm the organisation economically and result in privacy invasion.

We have also made use of YARA which is a rule-based approach towards malware analysis.

**Installation and Usage-**

Project was built using python 3.8 and it makes use of modules like flask, scikit learn, yara rules.

Following are the steps to set up the environment and deploy the project-

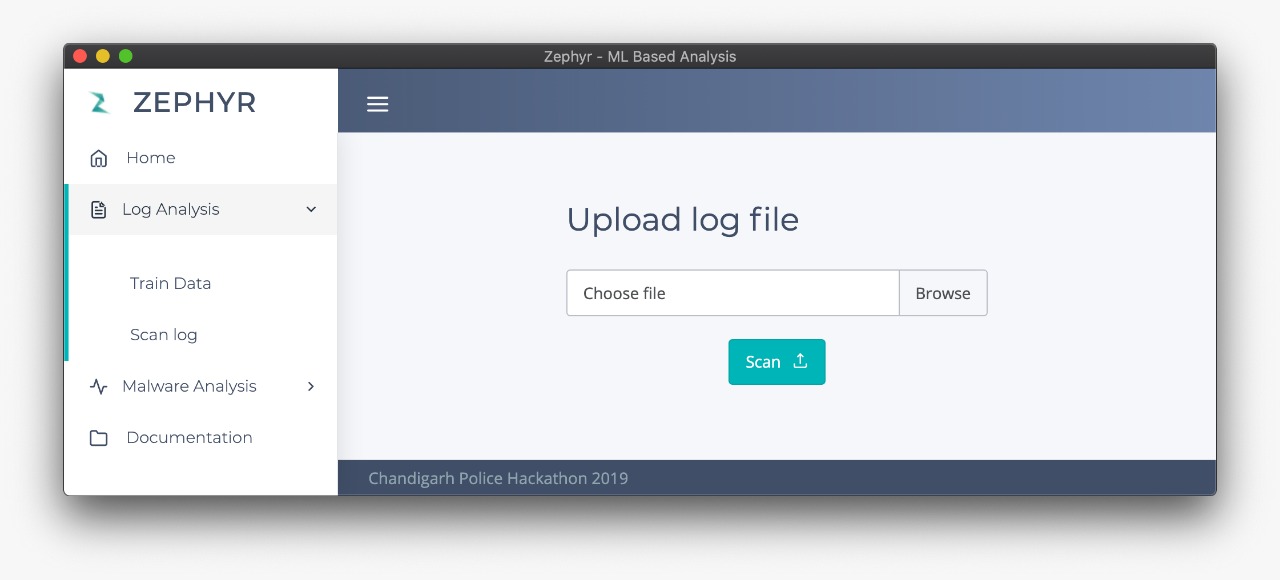
Step1: Browse to the project folder

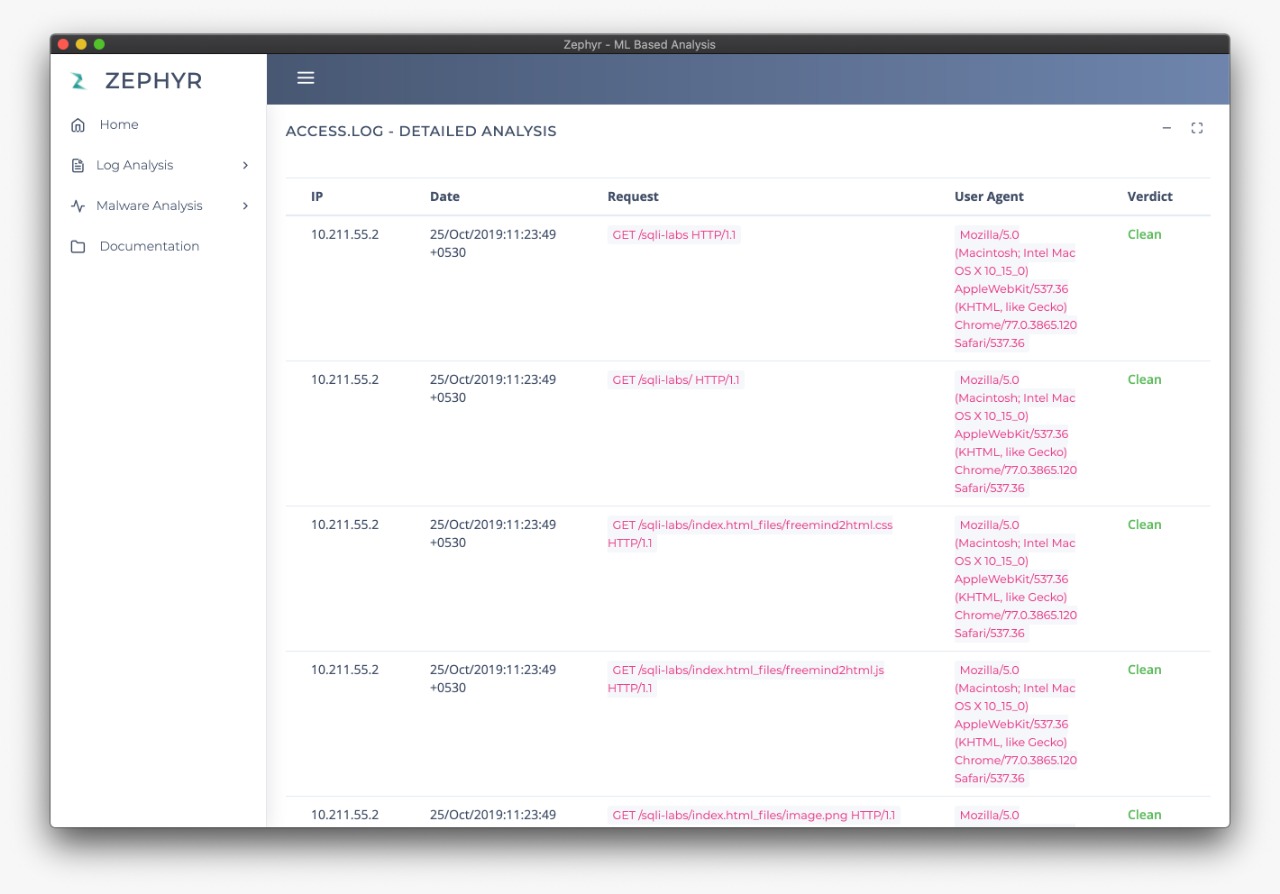
Step2: install the requirements using pip. pip install -r requirements.txt

Step 3: Serve application using python app.py

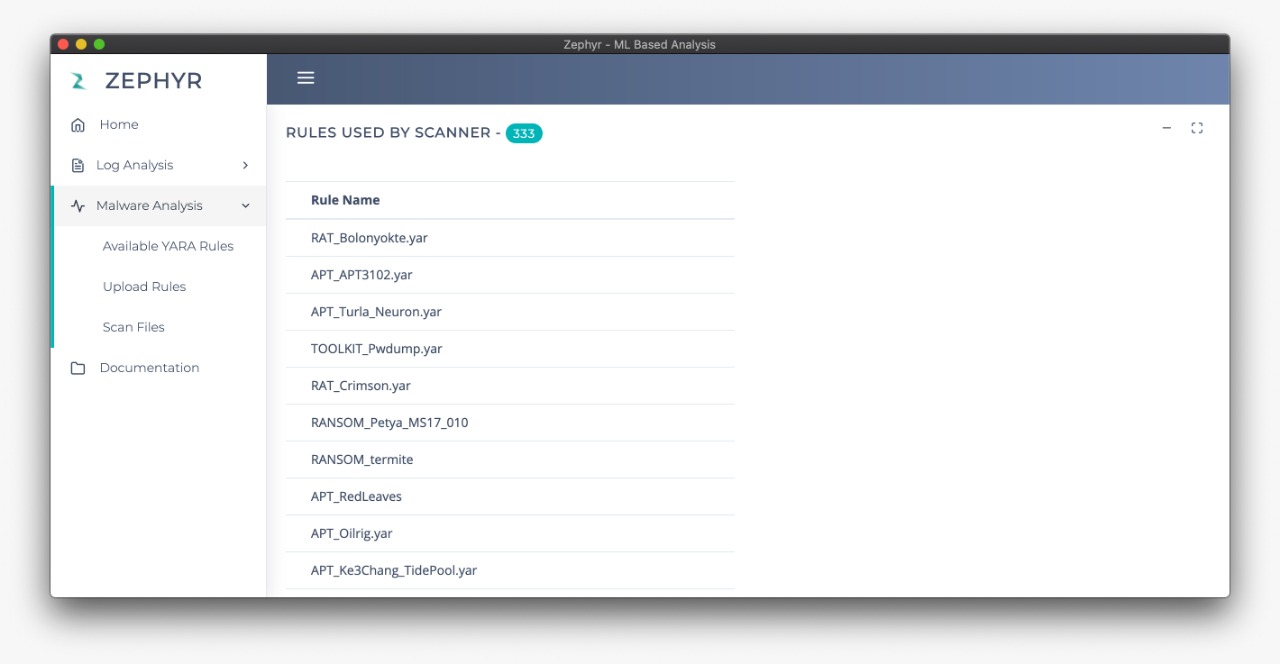
Now the tool can be used from web browser at <http://127.0.0.1:5000>

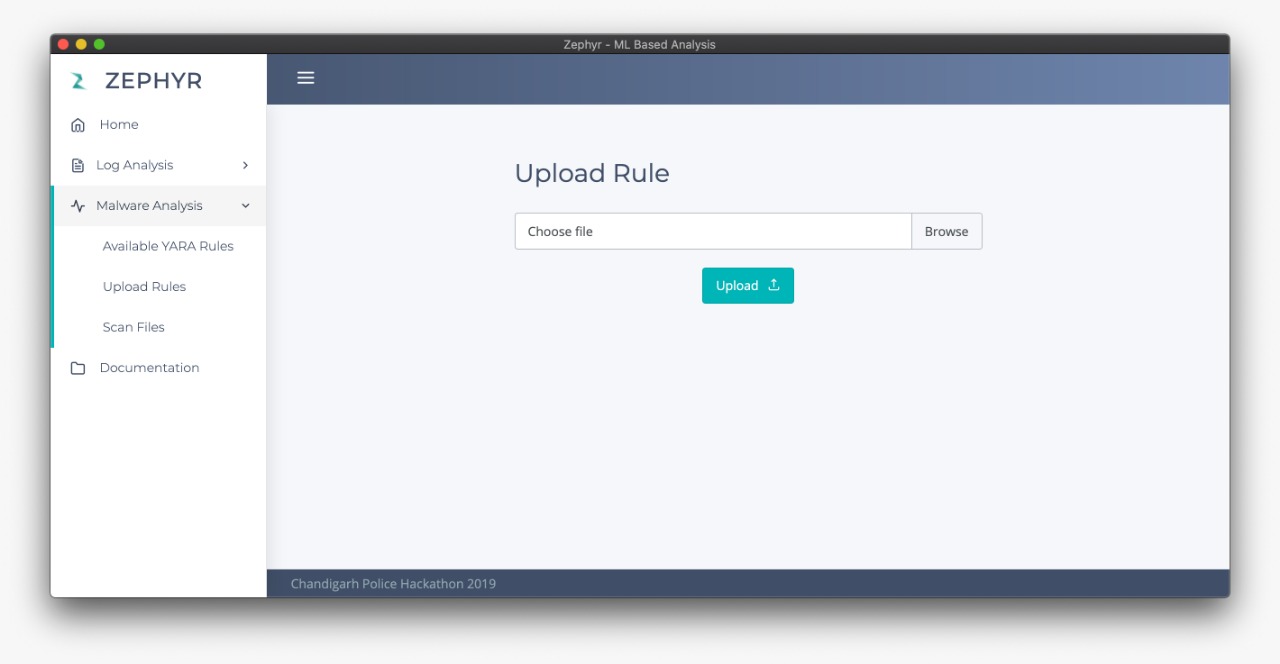
First item in navigation bar is Log Analysis. For log analysis first we have to train the model (should be done every time dataset is updated), once the model is trained click on the scan log from the navigation bar and then you can select the access log and hit scan button the uploaded file will be analysed with trained model.



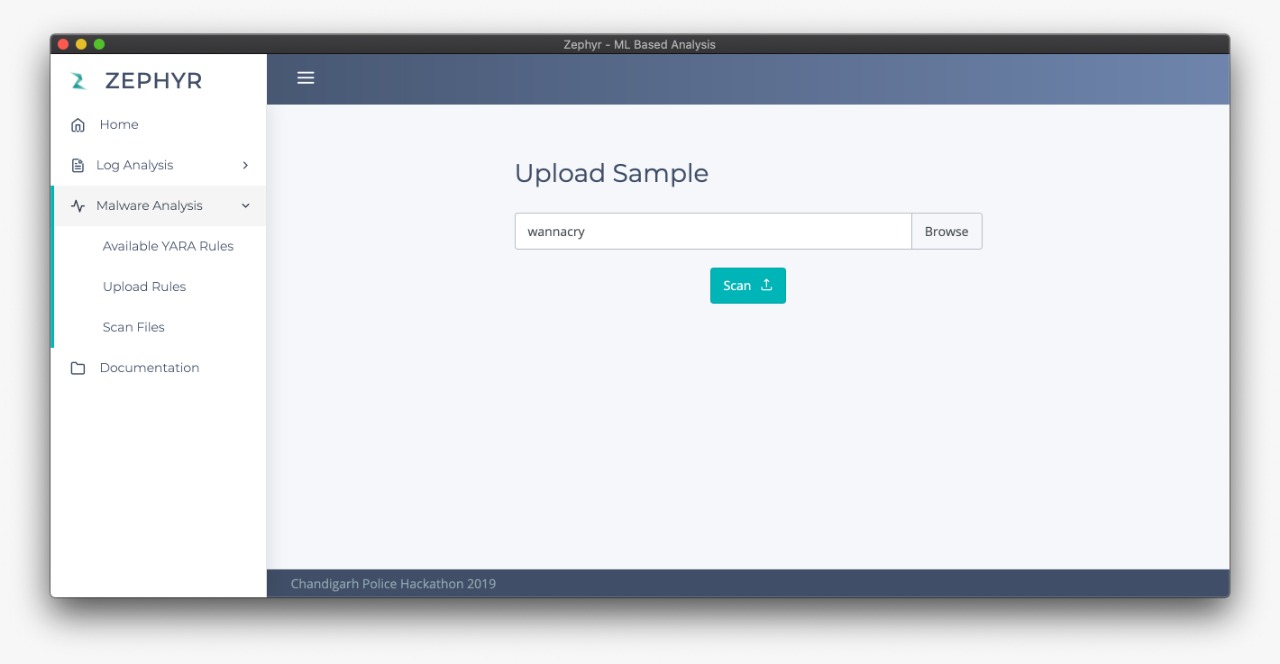
The output generated contains the details in tabular form which contains the IP, Date and Time, Requested URL, User Agent and verdict telling the user whether it is Clean or Malicious. Below is the snippet of sample output. 

Next, we have Malware analysis related functionality, first we have list of available YARA rules which are used to scan the uploaded sample.

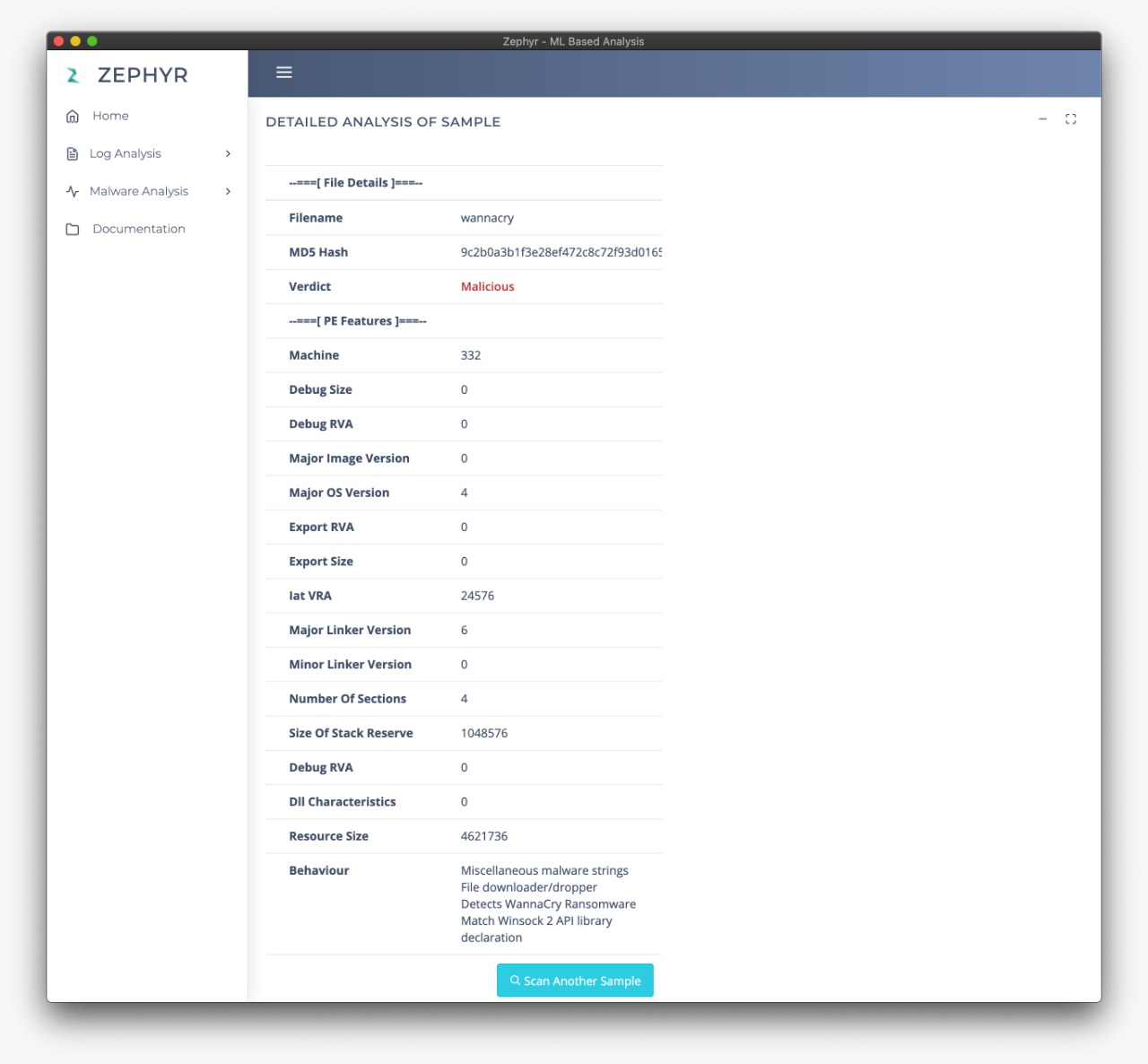


Second sub menu is Upload New YARA rules to the existing library of rules.

And the last option is Scan files, which is used to analyse the sample that was uploaded with the YARA rules available in our library. Below is the snippet to upload sample.



Output of scanned sample contains verdict based on detection with Yara rules as well as features related to the executable’s header.



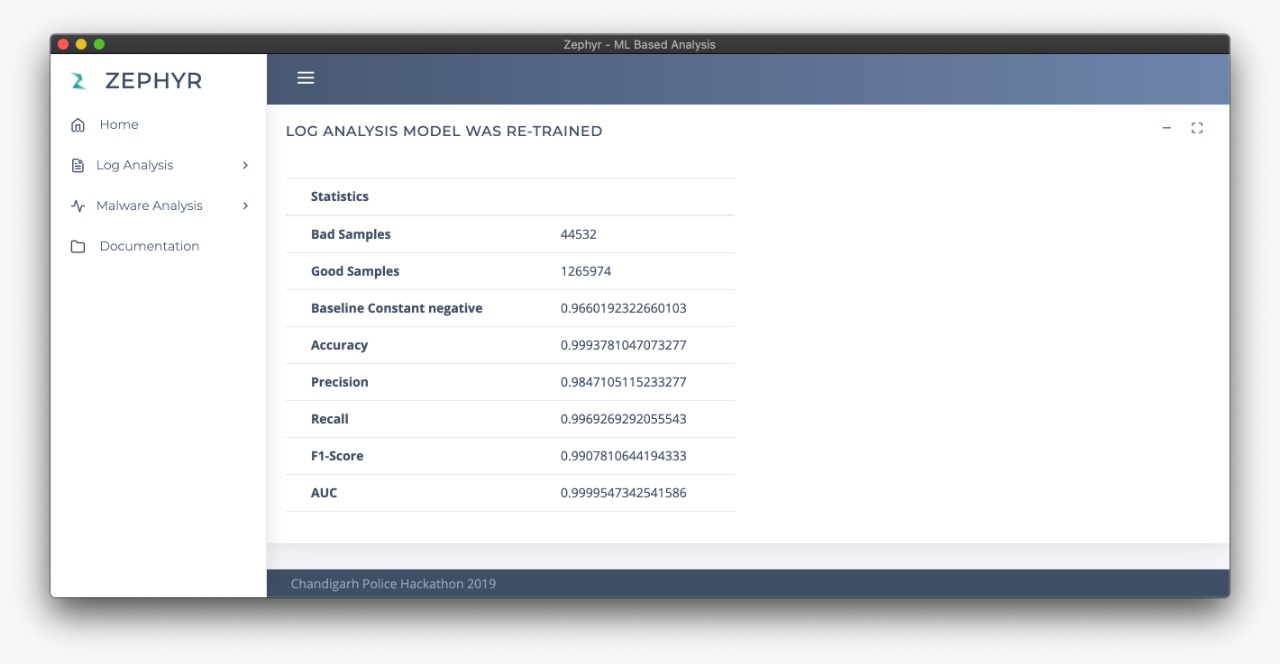
**Working-**

For log analysis we used Bad Queries containing payloads which are used to exploit different vulnerabilities and Good Queries containing clean requests. then we converted this data to vectors so that the data can be used for creating machine learning model.

In this case we have used Logistic Regression which comes under supervised machine learning. In which we have provided vectorized data for creating the model.

We evaluated the model based on baseline constant negatives, accuracy, precision, recall, f1-Score and AUC.

After training the model we pickled the trained model so that it would be easier to use the trained model for log analysis in future, instead of training the data each and every time which time is consuming.

Following is the snippet of performance of trained model using logistic regression.

Based on evaluation output we can see that the model’s accuracy is pretty high( ~99% ).

For malware analysis we scanned the uploaded sample with the YARA rules available in the library which are compiled on the runtime, based on detection with YARA rules we extract the behaviour of the sample that caused the detection.

After scanning the YARA rule we extract the features using the pefile module which is used to parse the header of executables.

**Future Scope-** Tool Zypher is presented with machine learning powered Malicious log analyser and Static malware analysis using YARA rules.

In future we intend to implement an AI powered neural network to detect windows malware based in Ergo and LIEF( library to instrument executable formats).

Ergo prefers GPU cores instead of CPU in order to significantly speed up the training process.