# Tutorial for regenerate the model and other.

By following this tutorial, you would get your new labeled data, features (csv and pickle), and the new model

# Part Zero: Use the existing tool: main.py and predict\_for\_FirstWatch

HOW to use the existing TOOL:

Main.py

python main,py –h

python main.py -p ~/labdata/rigData/423925dce27f6acdfc0582f2484eb44d

ls

cat console.txt

cat current\_prediction.csv

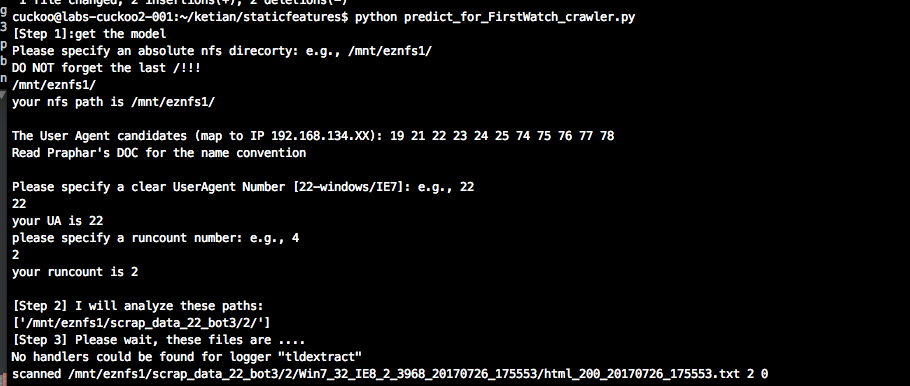
rm console.txt

rm current\_prediction.csv

python main.py –f –p ~/labdata/rigData/423925dce27f6acdfc0582f2484eb44d

python main.py –u

Python predict\_for\_FirstWatch\_crawler.py

cuckoo@labs-cuckoo2-001:~ /ketian/staticfeatures$ Python predict\_for\_FirstWatch\_crawler.py

# Part one: label file and retrain the model

Out testing benign dataset is Alexa10K, the malicious data set is htmlvt\_alpha

--

1 ssh to cuckoo

ssh [cuckoo@192.168.134.120](mailto:cuckoo@192.168.134.120)

your pwd

Data labeling

screen

cd /home/cuckoo/kt\_labtools/dataLabeling

python check\_iframes.py -t 1 -r ~/labtools/htmlvt\_alpha/

[choose n when you see the console, use control+a to run on background]

mv final-labled-data.txt /home/cuckoo/labdata/lableled\_data/tutorial/m-htmlvt-alpha.txt

(benign-alexa100k-labeled-data.txt is alreay there in tutorial folder)

Extract Features

cd ~/ketian/staticfeatures

python analysis.py ~/labdata/lableled\_data/toturial/benign-alexa100k-labeled-data.txt

python analysis.py ~/labdata/lableled\_data/toturial/m-htmlvt-alpha.txt.txt

(get csv and pickle files)

mv these pickles into train\_data and rename them as m-tutorial-htmlvt-alpha.pickle

and b-tutorial-alexa100k.pickle

Train the data

Comment all the existing code in read\_alaxa\_malicious():

And hard code these:

*def* read\_alaxa\_malicious():  
 *import* random  
 random.seed(RANDOM\_SEED)  
 mal = transfor\_feature\_dictionaries\_into\_trainable\_arrays("train\_data/m-tutorial-htmlvt-alpha.pickle")  
 beg = transfor\_feature\_dictionaries\_into\_trainable\_arrays("train\_data/b-tutorial-alexa100k.pickle")

The model file path is alreay hardcoded as

save\_the\_Model\_to\_model\_folder(rforest, x, y,name="model/tutorial\_randomforest.model")

call:

python train.py

you get the new model.