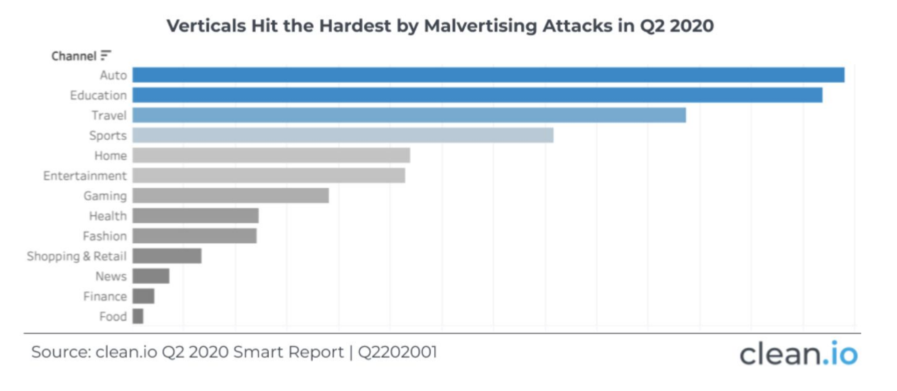
Machine Learning for Malvertising Detection

Aashish Ananthanarayan, Marc Riccione

What is Malvertising?

- Uses online advertising to spread malware
- Easy way to inject malware into a computer by attracting users to advertise a product
- Fairly new concept and hard to combat, the first ever malvertising attack took place around 2007

Statistics

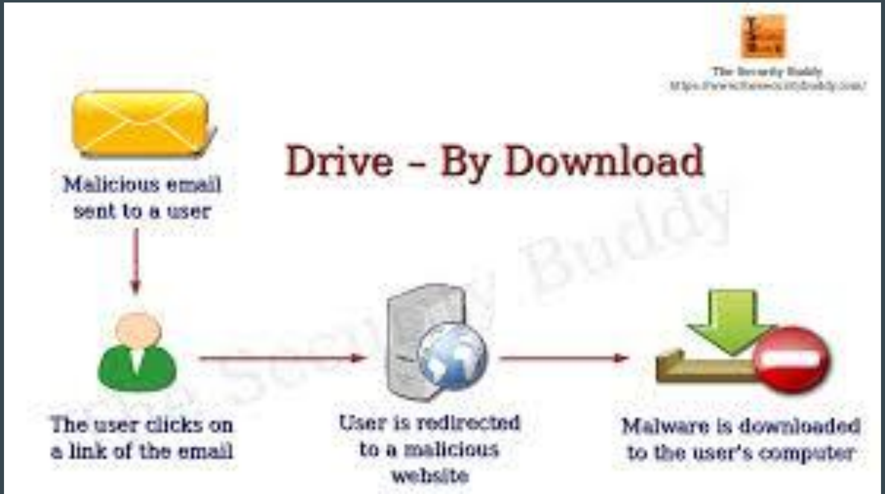


Verticals Hit the Hardest by Malvertising Attacks in Q2 2020

Source: clean.io Q2 2020 Smart Report | Q2202001

- Malvertising is found in 1 in every 100 Ad Impressions
- Roughly 61% of malicious ads target Windows users
- Attacks shift due to demands
- During COVID-19, increased threats were caused due to an increase in work-from-home lifestyles

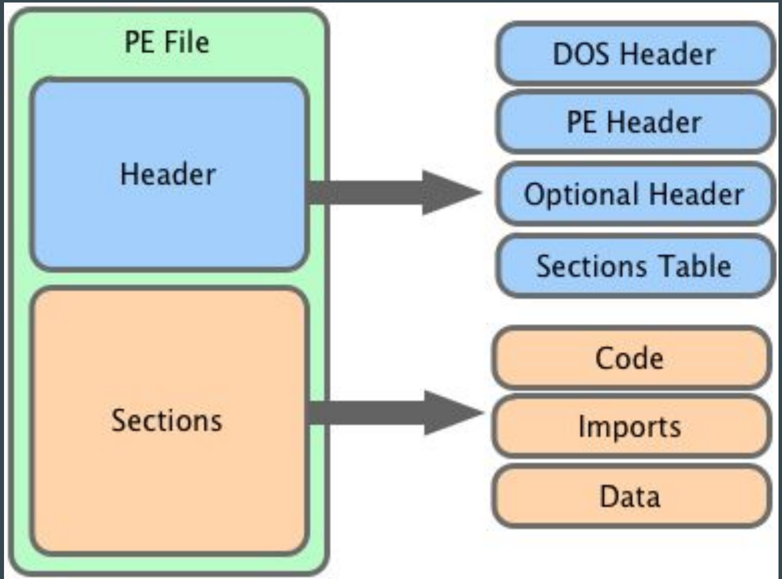
Our Proposed Idea


The Security Buddy
https://www.thesecuritybuddy.com/
Drive - By Download
Malicious email sent to a user
The user clicks on a link of the email
User is redirected to a malicious website
Malware is downloaded to the user's computer
Security Buddy

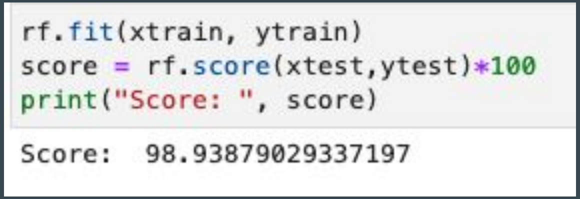
Data

Data:

- 137597 samples of PE header files
- 41323 of “legit” file data;
- 96274 samples of malware infected file data
- Original number of features: 56
- Used 8 most important features



Model



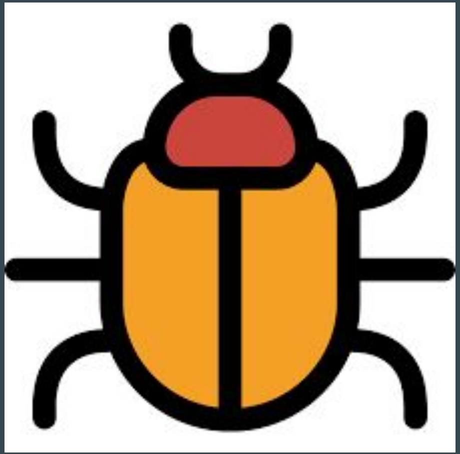
```
rf.fit(xtrain, ytrain)
score = rf.score(xtest,ytest)*100
print("Score: ", score)
```

```
Score:  98.93879029337197
```

- Train size: 80%; Test size: 20%
- Model utilized: Random Forest
- Number of estimators: 50
- Model final score: 98.9%

Chrome Extension

- Simple Chrome extension that parses through the URL and returns malicious files
- These files are then checked through the model
- The output is then returned as to whether the file is safe or not



Prototype Demo

Additional Work - Final Package

- Integrate the model and the extension to automate scanning of files
- Work on improving file interception
- Add a response for the classification in the browser

Q&A