Fighting Malware with Deep Learning

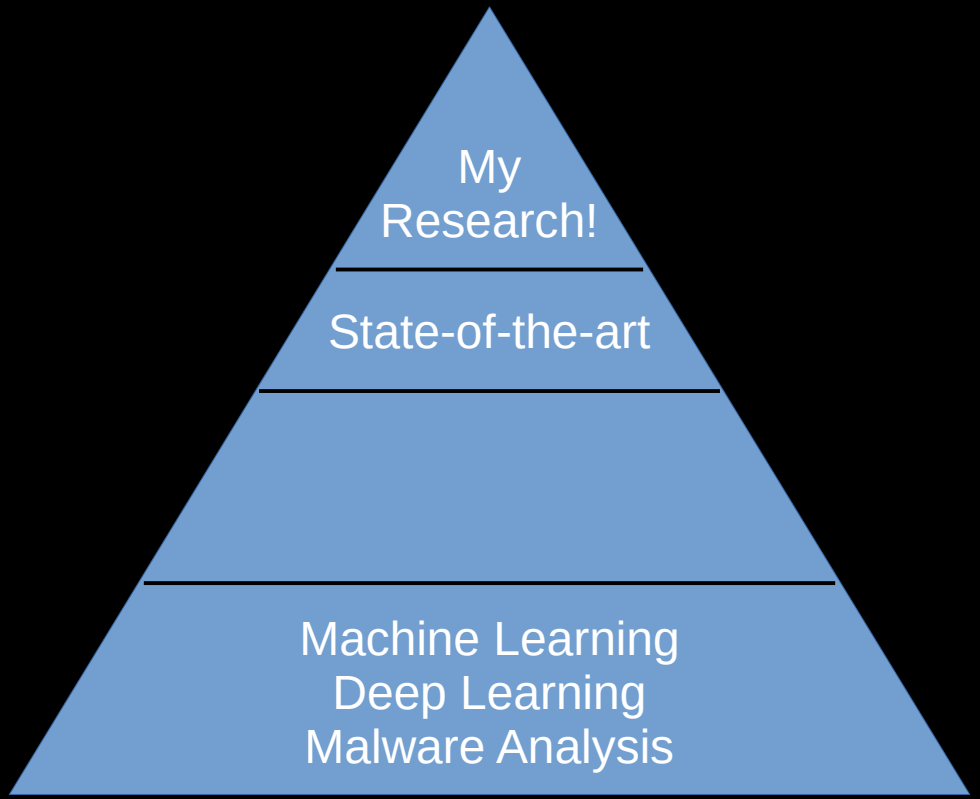
Angelo Oliveira | @ang3loliveira | www.angeloliveira.net | alpha@angeloliveira.net

TOTVS

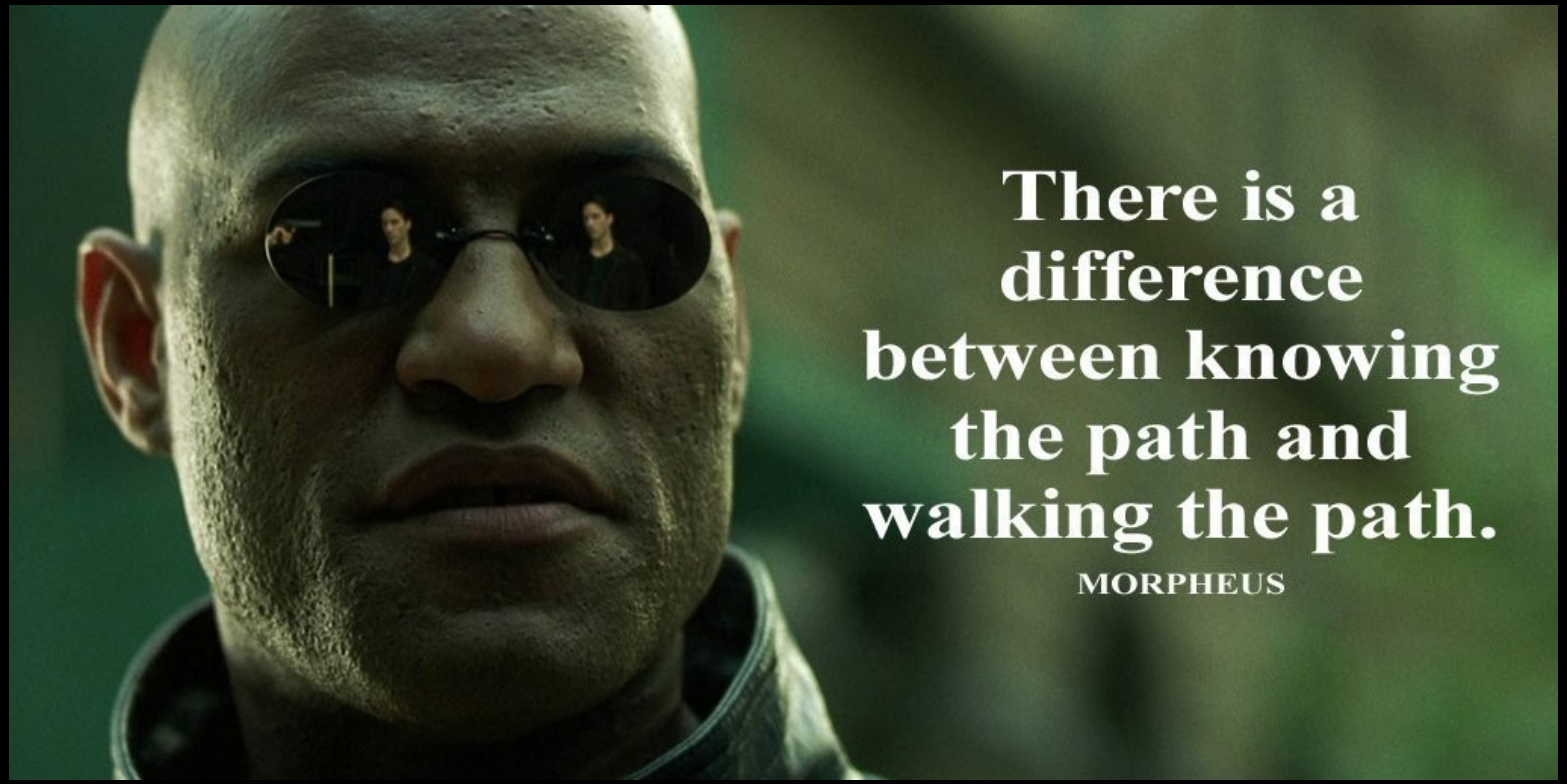
whoami

- Ethical Hacker @ TOTVS
- Part-time Data Scientist
- Part-Part-time PhD student
- Interested in Deep Learning and Data Science applied to malware detection and classification

The Gap


My
Research!
State-of-the-art
Machine Learning
Deep Learning
Malware Analysis

Filling the Gap



Filling the Gap

- Build your own baseline models “from the scratch” as a starting point for doing research
- Build simple, working models using the techniques you want to master and research
- You need to not contribute first to be able to contribute later: Always be humble

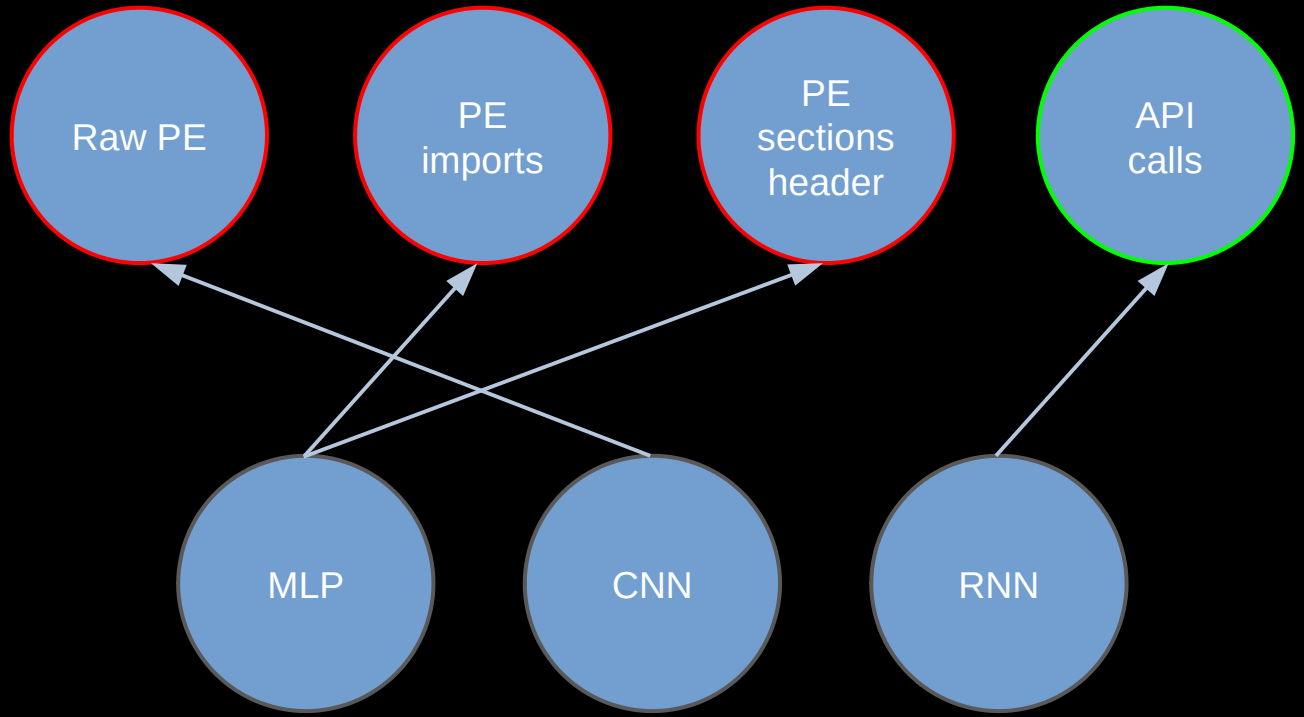
We need malware, lots of malware

- Malware
  - VirusShare.com: ~34M malware examples labelled with VirusTotal!
- Goodware
  - Executable files in Windows directories
  - portableapps.com: ~2K goodware (freeware) examples
    - Easy to run in a sandbox
  - Local Agent

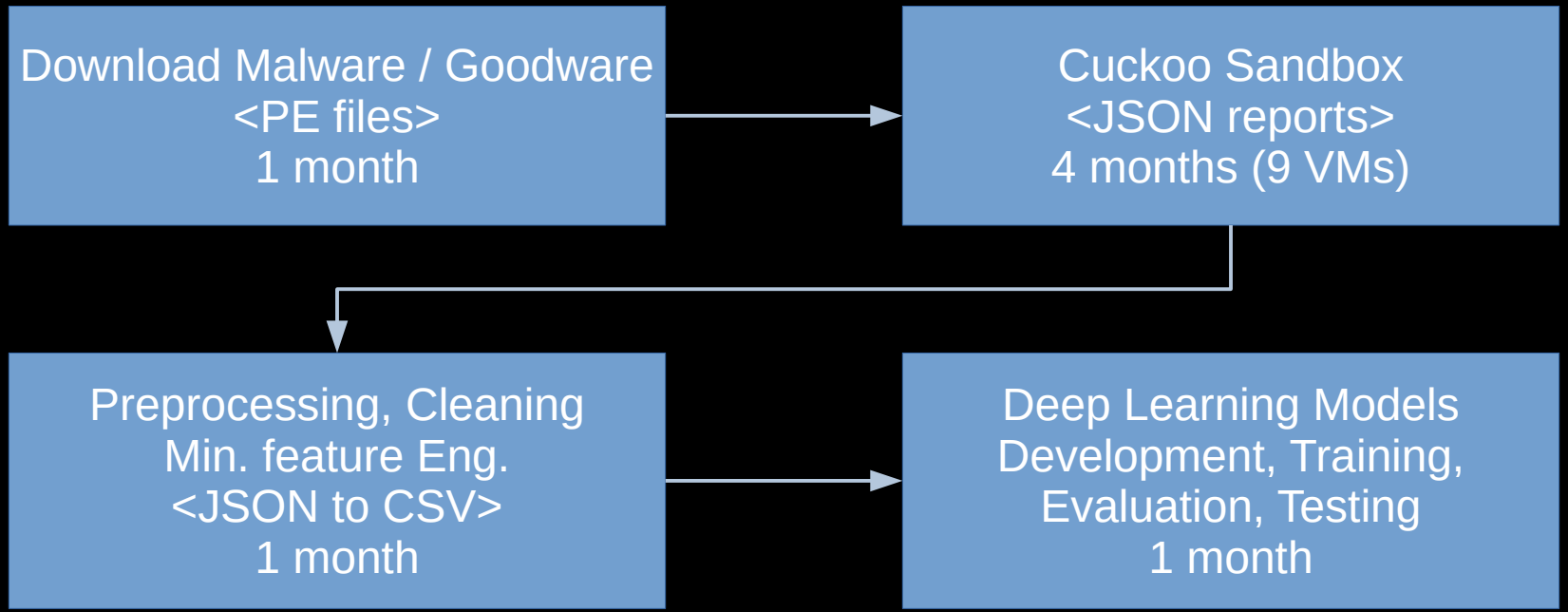
Static, Dynamic or Both?

- Static Analysis: What can we learn from a malware instance without running it?
  - Signatures (hashes, parts of the (malicious) code, etc)
  - File structure and content: PE headers
    - pypi.org/project/pefile/
    - Cuckoo Sandbox (cuckoosandbox.org)
- Dynamic Analysis: What can we learn from a malware instance by running it?
  - Behaviors (API (System) Calls, I/O, network traffic, etc)
  - Cuckoo Sandbox (cuckoosandbox.org)

“Natural” Use Cases



ML / DL / DS Pipeline



Talk is cheap. Show me the code.

(Linus Torvalds)

izquotes.com

My Research

- Malware behavior (Dynamic Analysis data)
  - Specialized data augmentation methods
  - Specialized representation learning techniques
- Improvements in detection and classification of zero-days, polymorphic and metamorphic malware

THANK YOU! :)

Slides and Jupyter Notebooks will be available at github
Datasets will be available at kaggle
Soon!

@ang3loliveira

Angelo Oliveira | @ang3loliveira | www.angeloliveira.net | alpha@angeloliveira.net

TOTVS