Unit V – ML Application to Cyber Security

Artificial Intelligence

School of Cyber Security & Digital Forensics

M. Sc. Cyber Security (Semester-I)

Malware Detection and Classification

- **Traditional AV solutions**
  - **Signature based**
  - **Heuristic based**
- **Requires Malware Analysis**
  - **Static analysis (faster but evades detection if malware obfuscated or concealed)**
  - **Dynamic analysis**
- **Need for ML approach**
  - **A large amount of data due to large attacks leads to high computation for searching and matching used in traditional AV approach**

Taxonomy of Malware

- **Adware**
- **Backdoor**
- **Bot**
- **Downloader**
- **Launcher**
- **Ransomware**
- **Rootkit**
- **Spyware**
- **Trojan**
- **Virus**
- **Worm**

Malware Analysis: Static Analysis

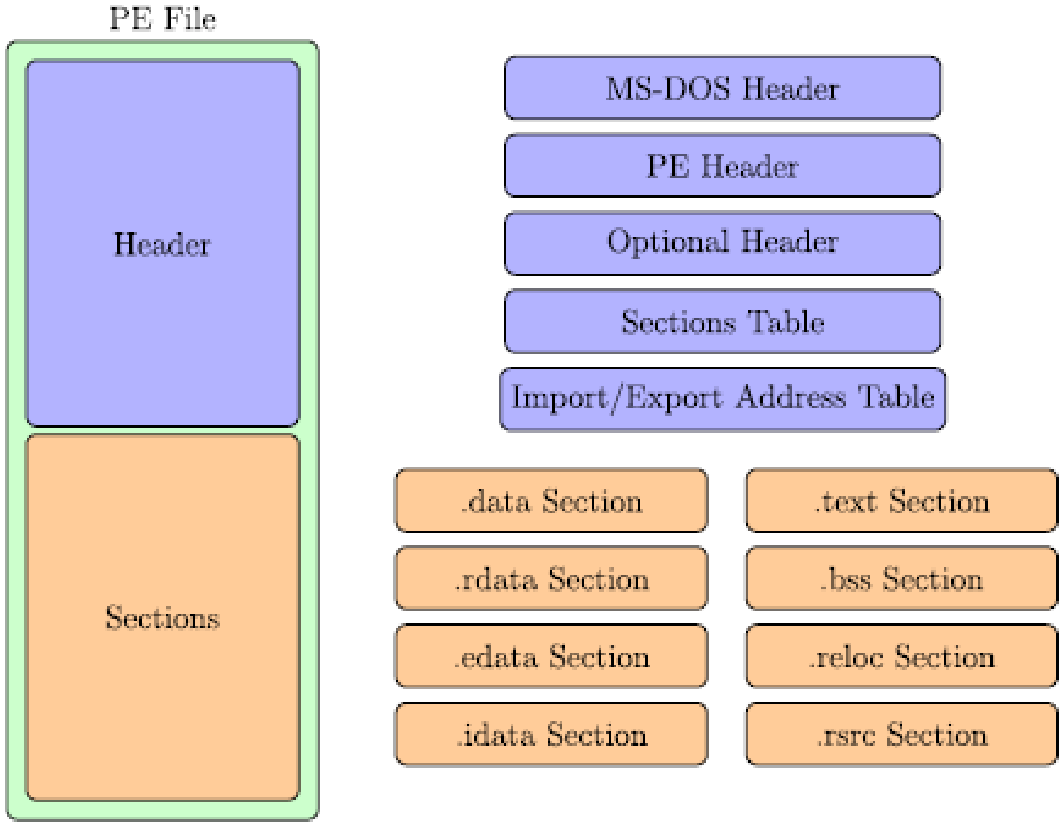
- **Static Analysis :** Examining the code or structure of the executable file without executing it. Produces signature and uses hashing (MD5 and SHA-1)
- **Common static analysis approaches are:**
  - **Finding sequence of characters or strings from binary file. (references to file paths of files modified , IP addresses, domain names attack command)**
  - **Gathering linked libraries and metadata about the file included in the headers**
  - **Analyze PE file headers and sections . PEView tool**
  - **Searching for packed and encrypted code**
  - **Disassembling the program : Translate the machine code into assembly language.**

Malware Analysis: Dynamic Analysis

- **Dynamic Analysis :** involves executing the program and monitoring its behaviour on the system.
- **The execution must be carried out in a safe environment:**
  - **Physical machines must be set up on air-gapped networks that is isolated from networks to prevent malware from spreading**
  - **Setup virtual machines to perform dynamic analysis *Vmware station, Oracle Virtualbox***
  - **All-in-one software products based on sandbox : *Cuckoo sandbox***
  - **Additional utilities : Process Monitor, Process Explorer, Regshot, NetCat and Wireshark**
- **Risks:** Some malware can detect when it running in a virtual machine or a sandbox

Portable Executable file format

- PE format is file formats for executables, object codes, DLLs for windows OS.
- **It encapsulates the information necessary for a windows OS to manage the executable code**



Traditional Machine Learning approach

- **Depending on the method of analysis, ML methods can be classified as :**
  - **Static methods : extract features from static analysis**
  - **Dynamic methods : extract features from dynamic analysis**
  - **Hybrid methods : combines both aspect of static and dynamic approach**

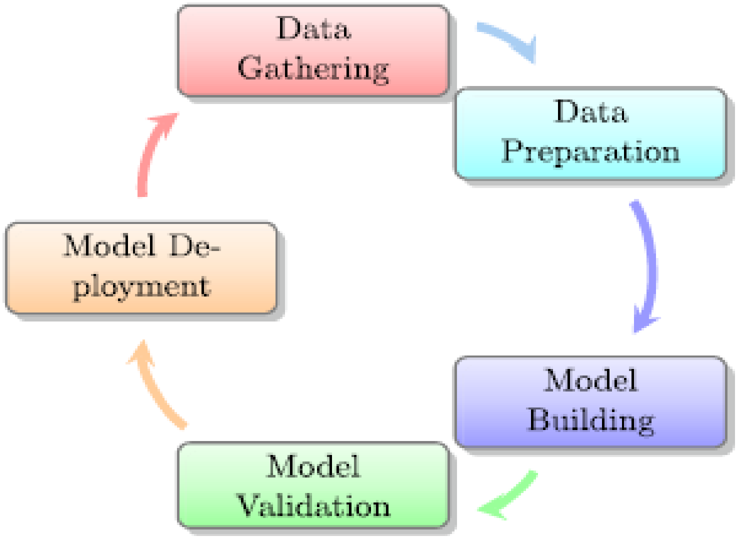


Fig. 2. Machine learning workflow.

Traditional Machine Learning approach

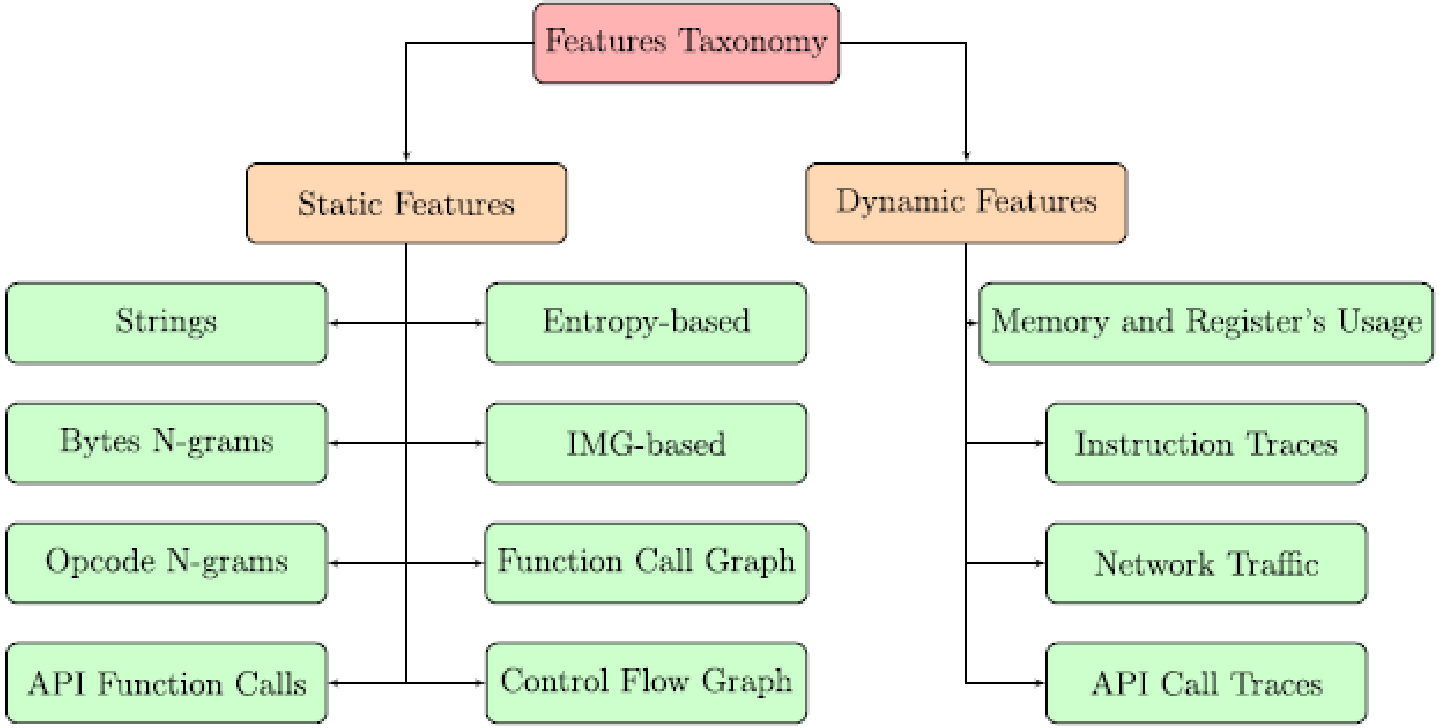


Fig. 3. Taxonomy of features used by traditional M.L. approaches.