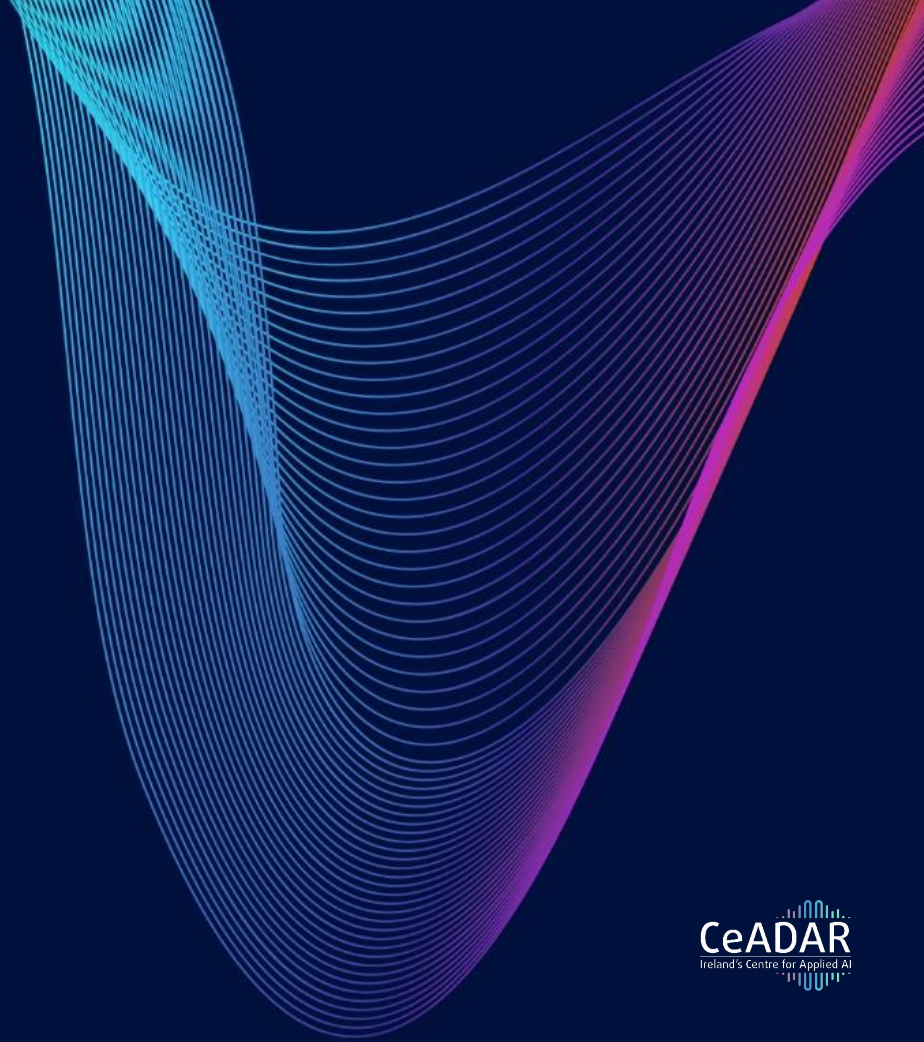
Machine Learning for Malware Detection and Classification

Dr. Daniel Gibert
CeADAR, University College Dublin



CeADAR
Ireland's Centre for Applied AI

Outline

- Components of a Malware Detection System
- Machine Learning for Malware Detection
- Current Work

Daniel Gibert

CeADAR
Ireland's Centre for Applied AI

An Overview to the Components of a Malware Detection System

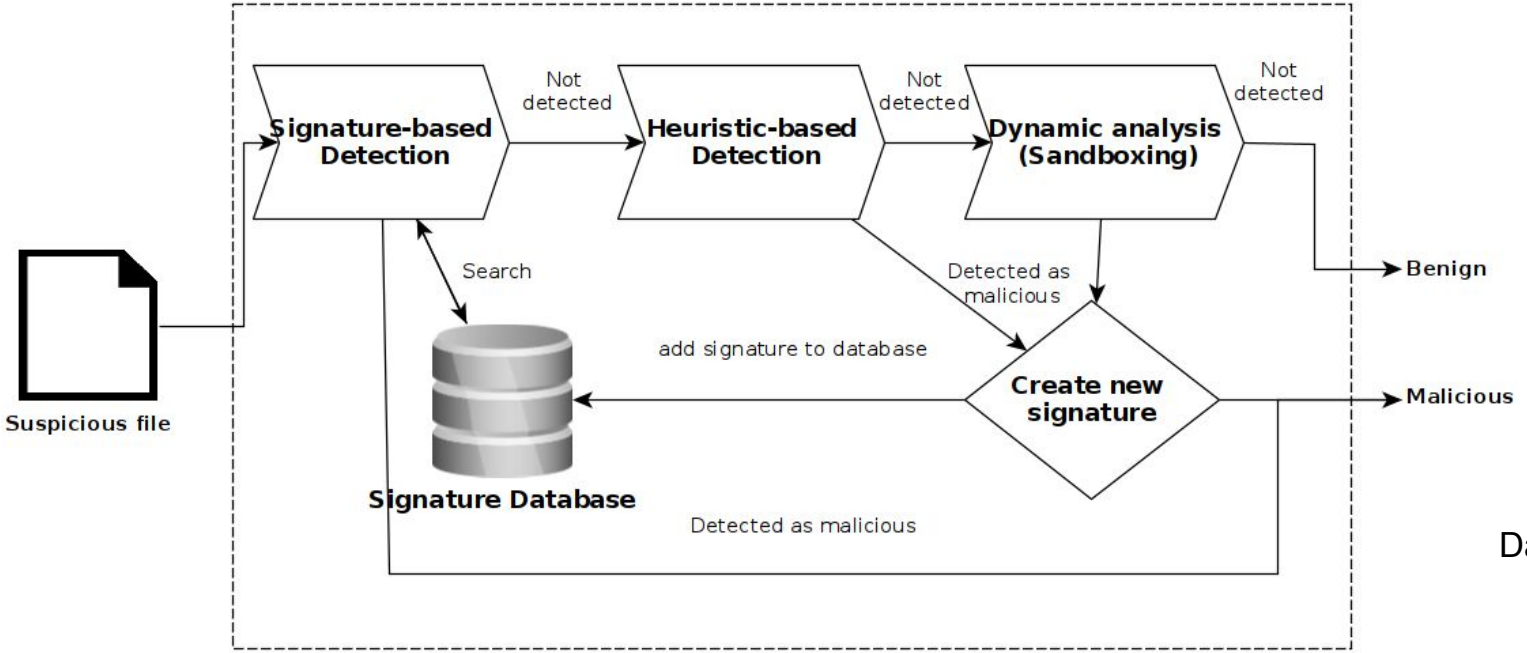


Fig 1. An illustration of a malware detection system

Daniel Gibert

CeADAR
Ireland's Centre for Applied AI

Machine Learning for Malware Detection

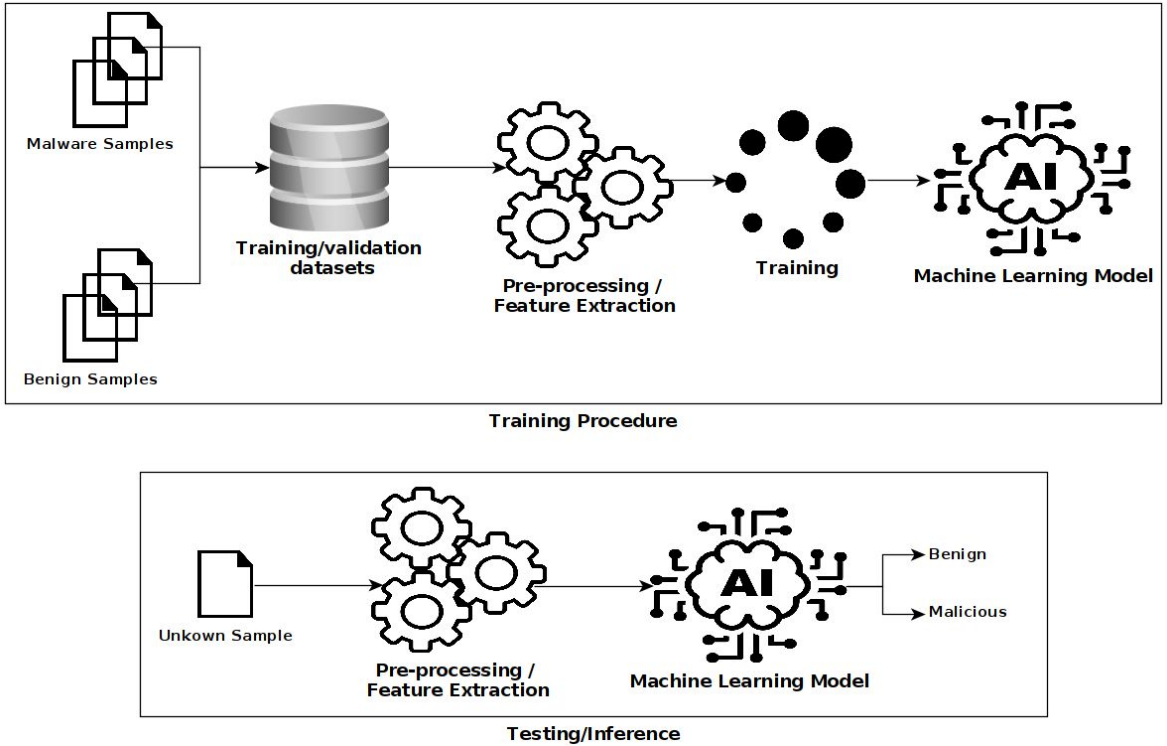


Fig 2. Training and inference process.

Daniel Gibert

CeADAR
Ireland's Centre for Applied AI

Feature-based Malware Detectors

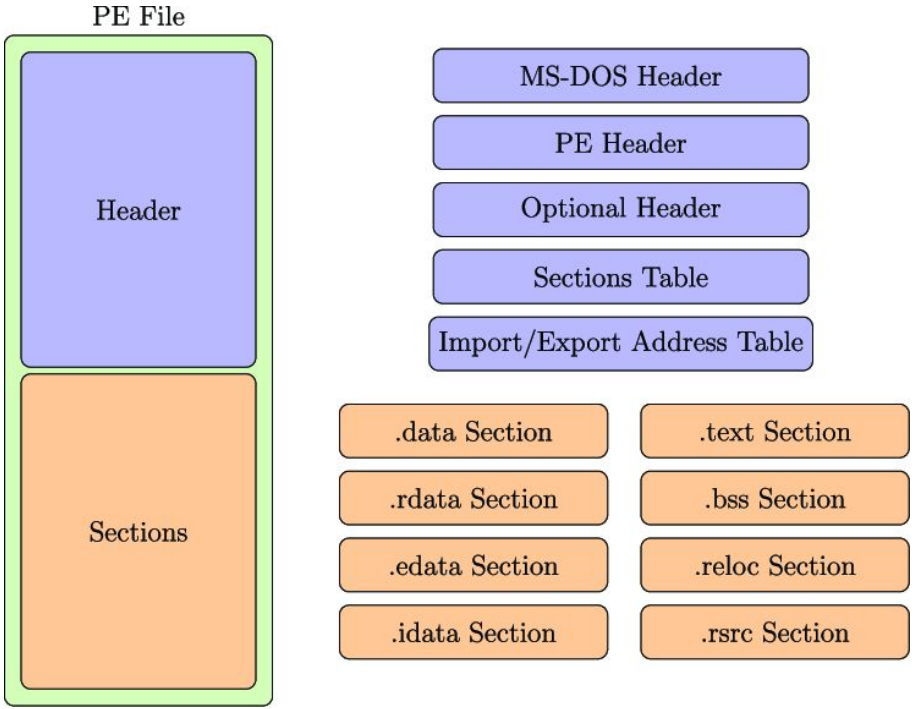


Fig 3. An overview of a Portable Executable file.

Daniel Gibert

CeADAR
Ireland's Centre for Applied AI

Deep Learning-based Detectors

```
00401000 56 8D 44 24 08 50 8B F1 E8 1C 1B 00 00 C7 06 08
00401010 BB 42 00 8B C6 5E C2 04 00 CC CC CC CC CC CC CC
00401020 C7 01 08 BB 42 00 E9 26 1C 00 00 CC CC CC CC CC
00401030 56 8B F1 C7 06 08 BB 42 00 E8 13 1C 00 00 F6 44
00401040 24 08 01 74 09 56 E8 6C 1E 00 00 83 C4 04 8B C6
00401050 5E C2 04 00 CC CC CC CC CC CC CC CC CC CC CC CC
00401060 8B 44 24 08 8A 08 8B 54 24 04 88 0A C3 CC CC CC
00401070 8B 44 24 04 8D 50 01 8A 08 40 84 C9 75 F9 2B C2
00401080 C3 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
00401090 8B 44 24 10 8B 4C 24 0C 8B 54 24 08 56 8B 74 24
004010A0 08 50 51 52 56 E8 18 1E 00 00 83 C4 10 8B C6 5E
004010B0 C3 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
004010C0 8B 44 24 10 8B 4C 24 0C 8B 54 24 08 56 8B 74 24
004010D0 08 50 51 52 56 E8 65 1E 00 00 83 C4 10 8B C6 5E
004010E0 C3 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
004010F0 33 C0 C2 10 00 CC CC CC CC CC CC CC CC CC CC CC
00401100 B8 08 00 00 00 C2 04 00 CC CC CC CC CC CC CC CC
00401110 B8 03 00 00 00 C3 CC CC CC CC CC CC CC CC CC CC
00401120 B8 08 00 00 00 C3 CC CC CC CC CC CC CC CC CC CC
00401130 8B 44 24 04 A3 AC 49 52 00 B8 FE FF FF FF C2 04
00401140 00 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
00401150 A1 AC 49 52 00 85 C0 74 16 8B 4C 24 08 8B 54 24
00401160 04 51 52 FF D0 C7 05 AC 49 52 00 00 00 00 00 B8
00401170 FB FF FF FF C2 08 00 CC CC CC CC CC CC CC CC CC
00401180 6A 04 68 00 10 00 00 68 68 BE 1C 00 6A 00 FF 15
00401190 9C 63 52 00 50 FF 15 C8 63 52 00 8B 4C 24 04 6A
```

Figure 5. Hexadecimal view.

```
.text:00401081 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC     align 10h
.text:00401090 8B 44 24 10                                     mov     eax, [esp+10h]
.text:00401094 8B 4C 24 0C                                     mov     ecx, [esp+0Ch]
.text:00401098 8B 54 24 08                                     mov     edx, [esp+8]
.text:0040109C 56                                              push    esi
.text:0040109D 8B 74 24 08                                     mov     esi, [esp+8]
.text:004010A1 50                                              push    eax
.text:004010A2 51                                              push    ecx
.text:004010A3 52                                              push    edx
.text:004010A4 56                                              push    esi
.text:004010A5 E8 18 1E 00 00                                  call    _memcpy_s
.text:004010AA 83 C4 10                                        add     esp, 10h
.text:004010AD 8B C6                                           mov     eax, esi
.text:004010AF 5E                                              pop     esi
.text:004010B0 C3                                              retn
.text:004010B0                                                ; ---------------------
.text:004010B1 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC     align 10h
.text:004010C0 8B 44 24 10                                     mov     eax, [esp+10h]
.text:004010C4 8B 4C 24 0C                                     mov     ecx, [esp+0Ch]
.text:004010C8 8B 54 24 08                                     mov     edx, [esp+8]
.text:004010CC 56                                              push    esi
.text:004010CD 8B 74 24 08                                     mov     esi, [esp+8]
.text:004010D1 50                                              push    eax
.text:004010D2 51                                              push    ecx
.text:004010D3 52                                              push    edx
.text:004010D4 56                                              push    esi
.text:004010D5 E8 65 1E 00 00                                  call    _memmove_s
.text:004010DA 83 C4 10                                        add     esp, 10h
.text:004010DD 8B C6                                           mov     eax, esi
.text:004010DF 5E                                              pop     esi
.text:004010E0 C3                                              retn
.text:004010E0                                                ; ---------------------
.text:004010E1 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC     align 10h
.text:004010F0 33 C0                                           xor     eax, eax
.text:004010F2 C2 10 00                                        retn    10h
.text:004010F2                                                ; ---------------------
```

Figure 6. Assembly view.

Daniel Gibert

CeADAR
Ireland's Centre for Applied AI

Grayscale Image Representation

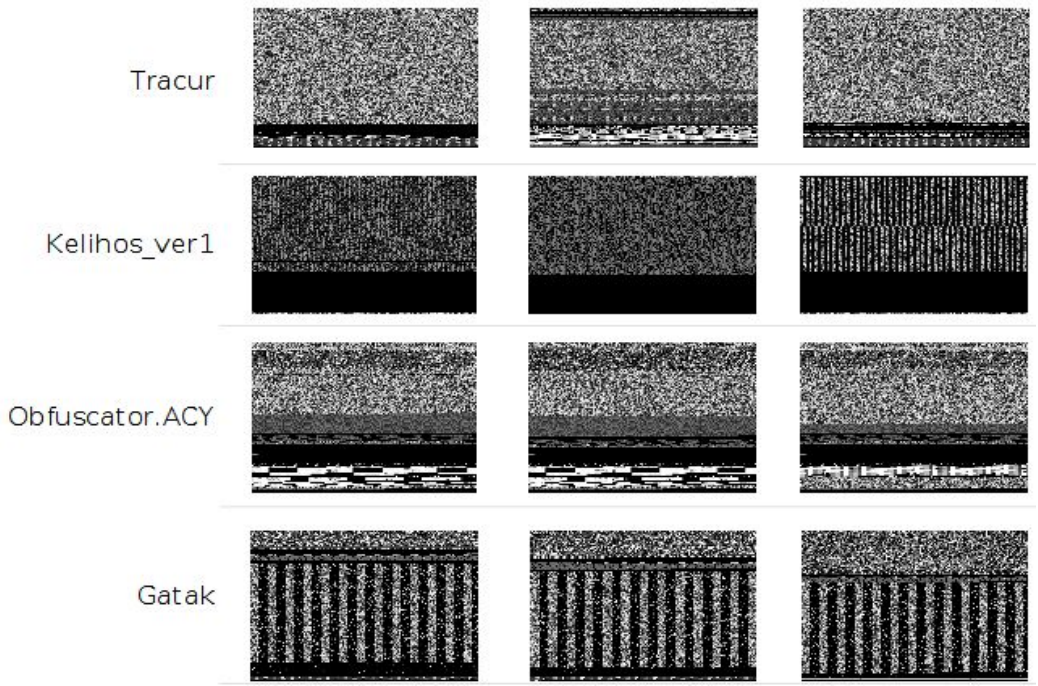
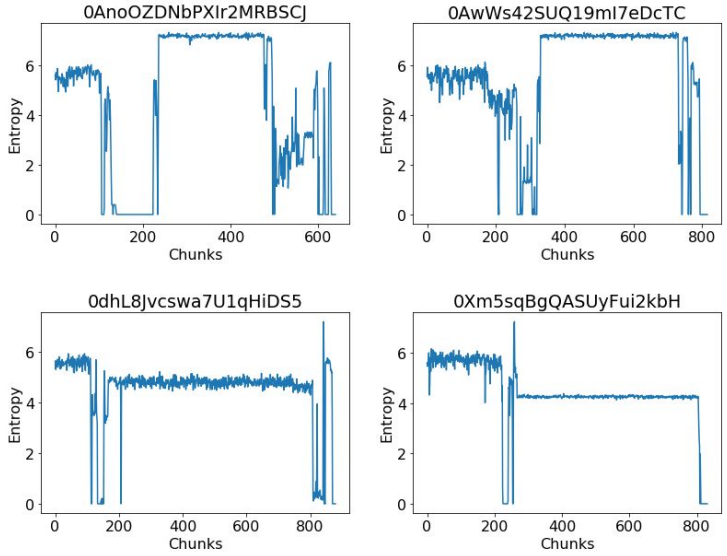


Figure 9. Grayscale image representation of malware

Daniel Gibert

CeADAR
Ireland's Centre for Applied AI

Structural Entropy Representation



Ramnit family

Gatak family

Figure 11. Entropy time series

Daniel Gibert

CeADAR
Ireland's Centre for Applied AI

Current Work at CeADAR & IBM

- Development of Evasion Attacks against ML-based Malware Detectors.
- Development of Defenses Against Evasion Attacks

Daniel Gibert

CeADAR
Ireland's Centre for Applied AI

Thank You

For Your Attention

Daniel Gibert

CeADAR
Ireland's Centre for Applied AI