PE-header-based Malware Study and Detection

Yibin Liao

Problems

- Is it possible to detect malware by simply looking at the properties of the PE headers?

Related Work

- S. Alvarez et al, The death of AV defense in depth? – revisiting anti-virus software:
  - AV scanners need to parse a variety of formats
- Jana et al, Abusing File Processing in malware Detectors for Fun and Profit:
  - Simplest malware can evade from sophisticated AV.
- M. Zubair et al, PE-Miner: Mining Structural Information to Detect Malicious Executables in Realtime:
  - Extract features, but scan the whole PE file

Motivations

- More Simple and Faster Detection
  - Just looking at the PE headers
  - Less scanning time

Threat Model

- What type of characteristic or anomalies?
- Embedded icons? What types are prevalent?
- Any other elements may be used for effective detection?

PE Format Overview

1. Overview

Figure 1 illustrates the Microsoft PE executable format.

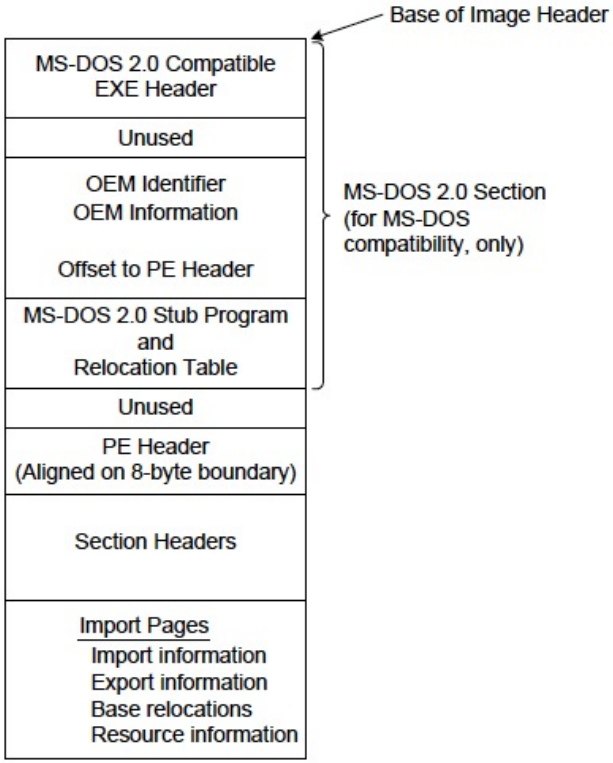
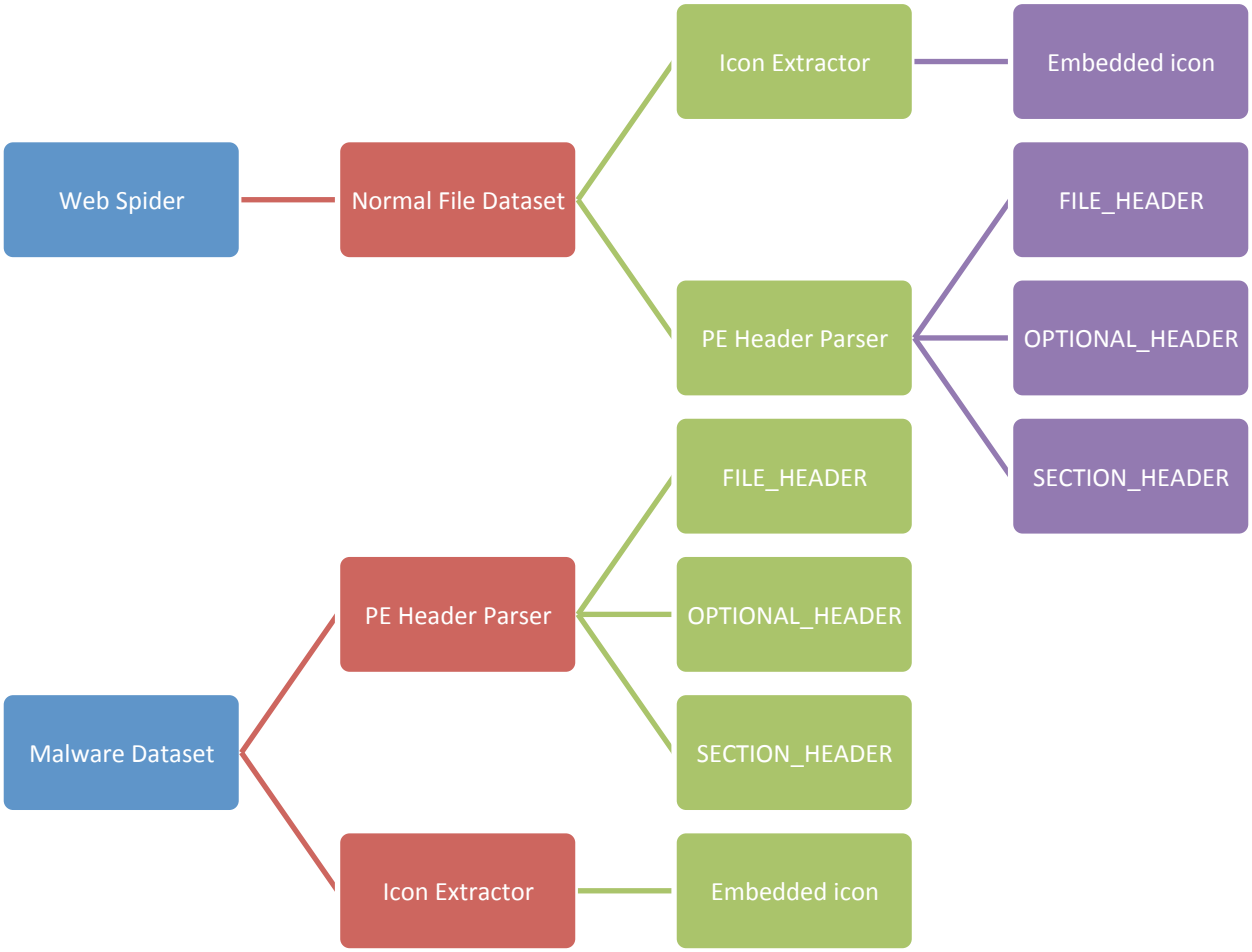


Figure 1. Typical Portable EXE File Layout

Approach Overview

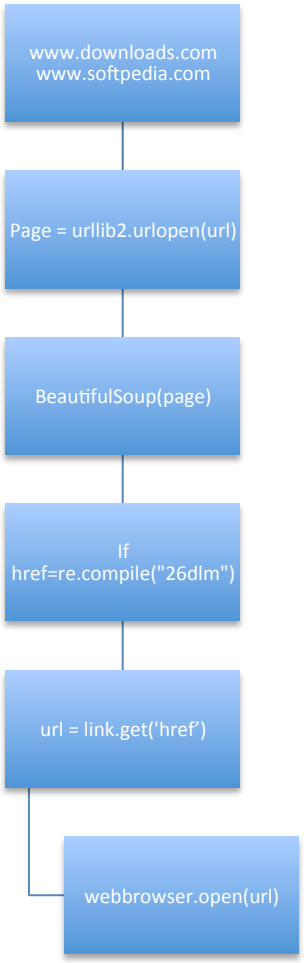


Approach Details

- Web Spider
  - Used to collect a dataset of normal files
- PE header Parser
  - Used to extract the features of each head field
- Icon Extractor
  - Used to extract the icon from the PE

Web Spider

(Python + Beautiful Soup + Appscript)



PE Header Parser

(Python + pefile)

- import pefile
- pe = pefile.PE('/path/simple.exe')
- print pe.FILE_HEADER
- print pe.OPTIONAL_HEADER
- print pe.dump_info()
  - Get the section header information

print pe.NT_HEADERS
print pe.FILE_HEADER

Normal .exe

```
----------NT_HEADERS----------

[IMAGE_NT_HEADERS]
0xE8       0x0   Signature:                    0x4550

----------FILE_HEADER----------

[IMAGE_FILE_HEADER]
0xEC       0x0   Machine:                      0x14C
0xEE       0x2   NumberOfSections:             0x4
0xF0       0x4   TimeDateStamp:                0x5002969A
 [Sun Jul 15 10:08:26 2012 UTC]
0xF4       0x8   PointerToSymbolTable:         0x0
0xF8       0xC   NumberOfSymbols:              0x0
0xFC       0x10  SizeOfOptionalHeader:         0xE0
0xFE       0x12  Characteristics:              0x103
Flags: IMAGE_FILE_32BIT_MACHINE, IMAGE_FILE_EXECUTABLE_IMA
GE, IMAGE_FILE_RELOCS_STRIPPED
```

Malware

```
----------NT_HEADERS----------

[IMAGE_NT_HEADERS]
0xC0       0x0   Signature:                    0x4550

----------FILE_HEADER----------

[IMAGE_FILE_HEADER]
0xC4       0x0   Machine:                      0x14C
0xC6       0x2   NumberOfSections:             0x4
0xC8       0x4   TimeDateStamp:                0x45CFC4C0
 [Mon Feb 12 01:37:04 2007 UTC]
0xCC       0x8   PointerToSymbolTable:         0x0
0xD0       0xC   NumberOfSymbols:              0x0
0xD4       0x10  SizeOfOptionalHeader:         0xE0
0xD6       0x12  Characteristics:              0x10E
Flags: IMAGE_FILE_LOCAL_SYMS_STRIPPED, IMAGE_FILE_32BIT_MA
CHINE, IMAGE_FILE_EXECUTABLE_IMAGE, IMAGE_FILE_LINE_NUMS_S
TRIPPED
```

print pe.OPTIONAL_HEADER

Normal .exe

```
----------OPTIONAL_HEADER----------

[IMAGE_OPTIONAL_HEADER]
0x100      0x0   Magic:                          0x10B
0x102      0x2   MajorLinkerVersion:             0xA
0x103      0x3   MinorLinkerVersion:             0x0
0x104      0x4   SizeOfCode:                     0x1E00
0x108      0x8   SizeOfInitializedData:          0x3800
0x10C      0xC   SizeOfUninitializedData:        0x0
0x110      0x10  AddressOfEntryPoint:            0x1078
0x114      0x14  BaseOfCode:                     0x1000
0x118      0x18  BaseOfData:                     0x3000
0x11C      0x1C  ImageBase:                      0x400000
0x120      0x20  SectionAlignment:               0x1000
0x124      0x24  FileAlignment:                  0x200
0x128      0x28  MajorOperatingSystemVersion:    0x5
0x12A      0x2A  MinorOperatingSystemVersion:    0x0
0x12C      0x2C  MajorImageVersion:              0x0
0x12E      0x2E  MinorImageVersion:              0x0
0x130      0x30  MajorSubsystemVersion:          0x5
0x132      0x32  MinorSubsystemVersion:          0x0
0x134      0x34  Reserved1:                      0x0
0x138      0x38  SizeOfImage:                    0x8000
0x13C      0x3C  SizeOfHeaders:                  0x400
0x140      0x40  CheckSum:                       0xFDAB
0x144      0x44  Subsystem:                      0x2
0x146      0x46  DllCharacteristics:             0x8500
0x148      0x48  SizeOfStackReserve:             0x100000
0x14C      0x4C  SizeOfStackCommit:              0x1000
0x150      0x50  SizeOfHeapReserve:              0x100000
0x154      0x54  SizeOfHeapCommit:               0x1000
0x158      0x58  LoaderFlags:                    0x0
0x15C      0x5C  NumberOfRvaAndSizes:            0x10
DllCharacteristics: IMAGE_DLL_CHARACTERISTICS_TERMINAL_SER
VER_AWARE, IMAGE_DLL_CHARACTERISTICS_NX_COMPAT, IMAGE_DLL_
CHARACTERISTICS_NO_SEH
```

Malware

```
----------OPTIONAL_HEADER----------

[IMAGE_OPTIONAL_HEADER]
0xD8       0x0   Magic:                          0x10B
0xDA       0x2   MajorLinkerVersion:             0x5
0xDB       0x3   MinorLinkerVersion:             0xC
0xDC       0x4   SizeOfCode:                     0x0
0xE0       0x8   SizeOfInitializedData:          0x0
0xE4       0xC   SizeOfUninitializedData:        0x0
0xE8       0x10  AddressOfEntryPoint:            0x31630
0xEC       0x14  BaseOfCode:                     0x2E000
0xF0       0x18  BaseOfData:                     0x0
0xF4       0x1C  ImageBase:                      0x400000
0xF8       0x20  SectionAlignment:               0x1000
0xFC       0x24  FileAlignment:                  0x200
0x100      0x28  MajorOperatingSystemVersion:    0x4
0x102      0x2A  MinorOperatingSystemVersion:    0x0
0x104      0x2C  MajorImageVersion:              0x0
0x106      0x2E  MinorImageVersion:              0x0
0x108      0x30  MajorSubsystemVersion:          0x4
0x10A      0x32  MinorSubsystemVersion:          0x0
0x10C      0x34  Reserved1:                      0x0
0x110      0x38  SizeOfImage:                    0x44000
0x114      0x3C  SizeOfHeaders:                  0x1000
0x118      0x40  CheckSum:                       0x0
0x11C      0x44  Subsystem:                      0x2
0x11E      0x46  DllCharacteristics:             0x0
0x120      0x48  SizeOfStackReserve:             0x100000
0x124      0x4C  SizeOfStackCommit:              0x1000
0x128      0x50  SizeOfHeapReserve:              0x100000
0x12C      0x54  SizeOfHeapCommit:               0x1000
0x130      0x58  LoaderFlags:                    0x0
0x134      0x5C  NumberOfRvaAndSizes:            0x10
DllCharacteristics:
```

SECTION_HEADER

Normal .exe

```
----------PE Sections----------

[IMAGE_SECTION_HEADER]
0x178      0x0   Name:                          .text
0x180      0x8   Misc:                          0x9BC94
0x180      0x8   Misc_PhysicalAddress:          0x9BC94
0x180      0x8   Misc_VirtualSize:              0x9BC94
0x184      0xC   VirtualAddress:                0x2000
0x188      0x10  SizeOfRawData:                 0x9BE00
0x18C      0x14  PointerToRawData:              0x400
0x190      0x18  PointerToRelocations:          0x0
0x194      0x1C  PointerToLinenumbers:          0x0
0x198      0x20  NumberOfRelocations:           0x0
0x19A      0x22  NumberOfLinenumbers:           0x0
0x19C      0x24  Characteristics:               0x60000020
Flags: IMAGE_SCN_CNT_CODE, IMAGE_SCN_MEM_EXECUTE, IMAGE_SCN_MEM_READ
Entropy: 6.531569 (Min=0.0, Max=8.0)
MD5     hash: 45d6b0c6698ea6249cfb72381ebeb0c7
SHA-1   hash: 285ca6b24b4c265f79ea1b6fc8c20ba94e70adc0
SHA-256 hash: 14d2a257ed7f9b45561668242ed987bf92b3e7114dc1e5f2147a59888f229afe
SHA-512 hash: b2ab50dd6c60527a11364ef0dc53692673e304cae7ff3cbcb65690d6b7fe3b3de6
725336a8a232da894525004a513f73361609cc24d08ffb845bdf11a2cadfbc

[IMAGE_SECTION_HEADER]
0x1A0      0x0   Name:                          .sdata
0x1A8      0x8   Misc:                          0xB1
0x1A8      0x8   Misc_PhysicalAddress:          0xB1
0x1A8      0x8   Misc_VirtualSize:              0xB1
0x1AC      0xC   VirtualAddress:                0x9E000
0x1B0      0x10  SizeOfRawData:                 0x200
0x1B4      0x14  PointerToRawData:              0x9C200
0x1B8      0x18  PointerToRelocations:          0x0
0x1BC      0x1C  PointerToLinenumbers:          0x0
0x1C0      0x20  NumberOfRelocations:           0x0
0x1C2      0x22  NumberOfLinenumbers:           0x0
0x1C4      0x24  Characteristics:               0xC0000040
Flags: IMAGE_SCN_MEM_WRITE, IMAGE_SCN_CNT_INITIALIZED_DATA, IMAGE_SCN_MEM_READ
Entropy: 2.183433 (Min=0.0, Max=8.0)
MD5     hash: c233166dbd9f5656be8d1ebec0c92687
SHA-1   hash: bc35aecc8d4fcf69b5d35d345e8401feb2f32ac5
SHA-256 hash: fa4bdff654589006dfd5623b450b3e78c6d5520a0be7c174cee743d38a6d9f52
SHA-512 hash: 38b82129d81c3fe43aa27c64913aa079bf007fe8cda73fa88db8248971f1716ece
d728d4223f504493f1af66f0ed38b4660861a926776affc5ad4a8da5ea136b
```

Malware

```
----------PE Sections----------

[IMAGE_SECTION_HEADER]
0x1B8      0x0   Name:                          .packed
0x1C0      0x8   Misc:                          0x21000
0x1C0      0x8   Misc_PhysicalAddress:          0x21000
0x1C0      0x8   Misc_VirtualSize:              0x21000
0x1C4      0xC   VirtualAddress:                0x1000
0x1C8      0x10  SizeOfRawData:                 0x0
0x1CC      0x14  PointerToRawData:              0x400
0x1D0      0x18  PointerToRelocations:          0x0
0x1D4      0x1C  PointerToLinenumbers:          0x0
0x1D8      0x20  NumberOfRelocations:           0x0
0x1DA      0x22  NumberOfLinenumbers:           0x0
0x1DC      0x24  Characteristics:               0xE0000020
Flags: IMAGE_SCN_MEM_WRITE, IMAGE_SCN_CNT_CODE, IMAGE_SCN_MEM_EXECUTE, IMAGE_SCN
_MEM_READ
Entropy: 0.000000 (Min=0.0, Max=8.0)
MD5     hash: d41d8cd98f00b204e9800998ecf8427e
SHA-1   hash: da39a3ee5e6b4b0d3255bfef95601890afd80709
SHA-256 hash: e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855
SHA-512 hash: cf83e1357eefb8bdf1542850d66d8007d620e4050b5715dc83f4a921d36ce9ce47
d0d13c5d85f2b0ff8318d2877eec2f63b931bd47417a81a538327af927da3e

[IMAGE_SECTION_HEADER]
0x1E0      0x0   Name:                          .RLPack
0x1E8      0x8   Misc:                          0xC000
0x1E8      0x8   Misc_PhysicalAddress:          0xC000
0x1E8      0x8   Misc_VirtualSize:              0xC000
0x1EC      0xC   VirtualAddress:                0x22000
0x1F0      0x10  SizeOfRawData:                 0xBAA9
0x1F4      0x14  PointerToRawData:              0x400
0x1F8      0x18  PointerToRelocations:          0x0
0x1FC      0x1C  PointerToLinenumbers:          0x0
0x200      0x20  NumberOfRelocations:           0x0
0x202      0x22  NumberOfLinenumbers:           0x0
0x204      0x24  Characteristics:               0xE0000020
Flags: IMAGE_SCN_MEM_WRITE, IMAGE_SCN_CNT_CODE, IMAGE_SCN_MEM_EXECUTE, IMAGE_SCN
_MEM_READ
Entropy: 7.974447 (Min=0.0, Max=8.0)
MD5     hash: 6f1aaff493ef2907bdaf9ff003f3fd55
SHA-1   hash: 9c43c581d30deaffc01f25784e002b88739f40ed
SHA-256 hash: a437bdaef11db963cb39fcead79eeac5e727b20bb251d4c8b6b1b80cb5132d6a
SHA-512 hash: 40b5db7d388463481bf18f18647ff177d796ab84c3f42958b5db02fd942860c898
64f0db25b29f0f822960abd692a3eee1359d3baa686cae00a4b283216c572f
```

SECTION_HEADER(cont.)

Normal .exe

```
[IMAGE_SECTION_HEADER]
0x1C8      0x0   Name:                          .rsrc
0x1D0      0x8   Misc:                          0x7314
0x1D0      0x8   Misc_PhysicalAddress:          0x7314
0x1D0      0x8   Misc_VirtualSize:              0x7314
0x1D4      0xC   VirtualAddress:                0xA0000
0x1D8      0x10  SizeOfRawData:                 0x7400
0x1DC      0x14  PointerToRawData:              0x9C400
0x1E0      0x18  PointerToRelocations:          0x0
0x1E4      0x1C  PointerToLinenumbers:          0x0
0x1E8      0x20  NumberOfRelocations:           0x0
0x1EA      0x22  NumberOfLinenumbers:           0x0
0x1EC      0x24  Characteristics:               0x40000040
Flags: IMAGE_SCN_CNT_INITIALIZED_DATA, IMAGE_SCN_MEM_READ
Entropy: 5.255001 (Min=0.0, Max=8.0)
MD5     hash: f4f0ed226fae17857cb0f4d7c19c2128
SHA-1   hash: 8d2e7d82a1e6d9684c4d49e0b76f7cc8d4f16ccb
SHA-256 hash: c1d70d4249f8183433bb99cb5614f65b7d4c675602a0cce73c0b0ee8745d3282
SHA-512 hash: 9e0a06420ede42b2b280dd1847549f862a6ecfc1be6d31d1ebdbdc0cc75642ab1e
a60c74f7268e24324e9453ec5238c014a779ebb822b5b91a66ed0c090cae88

[IMAGE_SECTION_HEADER]
0x1F0      0x0   Name:                          .reloc
0x1F8      0x8   Misc:                          0xC
0x1F8      0x8   Misc_PhysicalAddress:          0xC
0x1F8      0x8   Misc_VirtualSize:              0xC
0x1FC      0xC   VirtualAddress:                0xA8000
0x200      0x10  SizeOfRawData:                 0x200
0x204      0x14  PointerToRawData:              0xA3800
0x208      0x18  PointerToRelocations:          0x0
0x20C      0x1C  PointerToLinenumbers:          0x0
0x210      0x20  NumberOfRelocations:           0x0
0x212      0x22  NumberOfLinenumbers:           0x0
0x214      0x24  Characteristics:               0x42000040
Flags: IMAGE_SCN_CNT_INITIALIZED_DATA, IMAGE_SCN_MEM_DISCARDABLE, IMAGE_SCN_MEM_
READ
Entropy: 0.101910 (Min=0.0, Max=8.0)
MD5     hash: 45576c8a9555c8cc5a9c2cba3c078790
SHA-1   hash: 93a69b6a8a8c139932ffaef1c10c84ca68597a59
SHA-256 hash: 2714041b6dc9e50894f669562abfce6a0201d9fdb24cae4148d1ebe80b75b398
SHA-512 hash: dd570aa96d8362f71bef59abd1169dc4096e757eac02f9bc45d3cbc01e24a7539c
8cc90cd0942b9cd4f1c33e51ae6d388c4d3bf97cd70c85ed3889e0b0203d9b
```

Malware

```
[IMAGE_SECTION_HEADER]
0x208      0x0   Name:                          s6rwe081
0x210      0x8   Misc:                          0x5000
0x210      0x8   Misc_PhysicalAddress:          0x5000
0x210      0x8   Misc_VirtualSize:              0x5000
0x214      0xC   VirtualAddress:                0x2E000
0x218      0x10  SizeOfRawData:                 0x4CB3
0x21C      0x14  PointerToRawData:              0xC000
0x220      0x18  PointerToRelocations:          0x0
0x224      0x1C  PointerToLinenumbers:          0x0
0x228      0x20  NumberOfRelocations:           0x0
0x22A      0x22  NumberOfLinenumbers:           0x0
0x22C      0x24  Characteristics:               0xE0000020
Flags: IMAGE_SCN_MEM_WRITE, IMAGE_SCN_CNT_CODE, IMAGE_SCN_MEM_EXECUTE, IMAGE_SCN
_MEM_READ
Entropy: 6.159646 (Min=0.0, Max=8.0)
MD5     hash: 0470b4e659dacb6ab6c2bbf6b2e5a6da
SHA-1   hash: 81ba1b644d684e34f2dabdefa3a5e84965cdd796
SHA-256 hash: fbbb0bab6452e8bd150535d30e34aa0f6a4d1e076e8080d4e0d97684a1ddf492
SHA-512 hash: 5c49017a4ad26c186ec3ed3473302ee2a4c2c666e736780efb3c40a173fffba956
7d3b96bd02dc76835edffec544460876f95040a943631831578d01e049cabc

[IMAGE_SECTION_HEADER]
0x230      0x0   Name:                          6i2eu9uw
0x238      0x8   Misc:                          0x8000
0x238      0x8   Misc_PhysicalAddress:          0x8000
0x238      0x8   Misc_VirtualSize:              0x8000
0x23C      0xC   VirtualAddress:                0x33000
0x240      0x10  SizeOfRawData:                 0x2C00
0x244      0x14  PointerToRawData:              0x10E00
0x248      0x18  PointerToRelocations:          0x0
0x24C      0x1C  PointerToLinenumbers:          0x0
0x250      0x20  NumberOfRelocations:           0x0
0x252      0x22  NumberOfLinenumbers:           0x0
0x254      0x24  Characteristics:               0xE0000060
Flags: IMAGE_SCN_MEM_WRITE, IMAGE_SCN_CNT_CODE, IMAGE_SCN_MEM_EXECUTE, IMAGE_SCN
_CNT_INITIALIZED_DATA, IMAGE_SCN_MEM_READ
Entropy: 6.291828 (Min=0.0, Max=8.0)
MD5     hash: 11c7bc4823706bbeecd1837dba2dcc78
SHA-1   hash: 3a8968a4e201980b3ef738277a915517a81e6530
SHA-256 hash: e18dcf3c5c174070b0f70aa06bca719ee5e7dd27b2467d1fdfe75470c99d371e
SHA-512 hash: 8271a70c32dda2069185ba68229dec4b88e7c776de3611541e9b4e005a5fff96f7
4acf0a0bc3d9c499ffa2431be7b66d03d41e4929515540a44c740355ac489e
```

Section Names

- Normal:

  .text, .bss, .rdata, .data, .rsrc, .edata, .idata, .pdata, and .debug

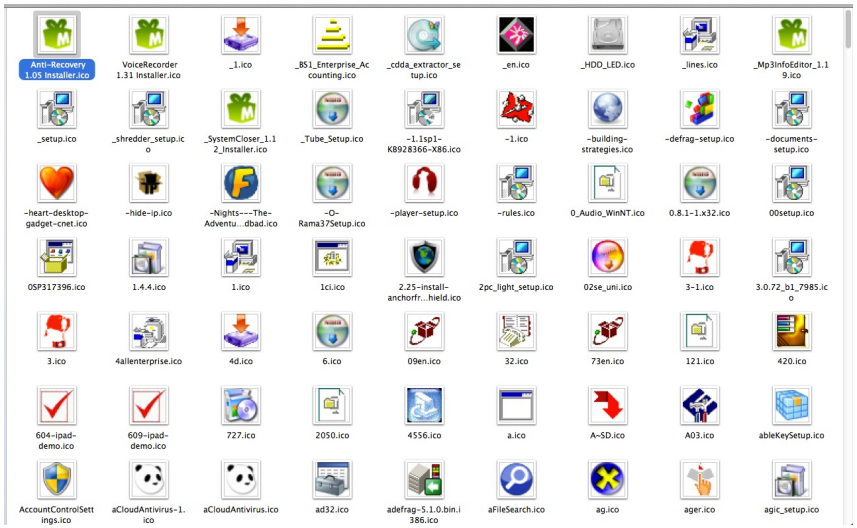
  .DATA, .CODE, .BSS, .sxdata, .itext, .adata, .ndata, .reloc, . _winzip_, .tls,
  UPX1, .UPX0, .boom, .seau, .code, .Shared, .gentee, .asp, .CRT, .PAGE, INIT, . tsu

- Malware:
  - Empty section name.
  - . 6dnn4fh4, . Bga1m3ar, . l0u15g4l, . 4ixl8myu, ...

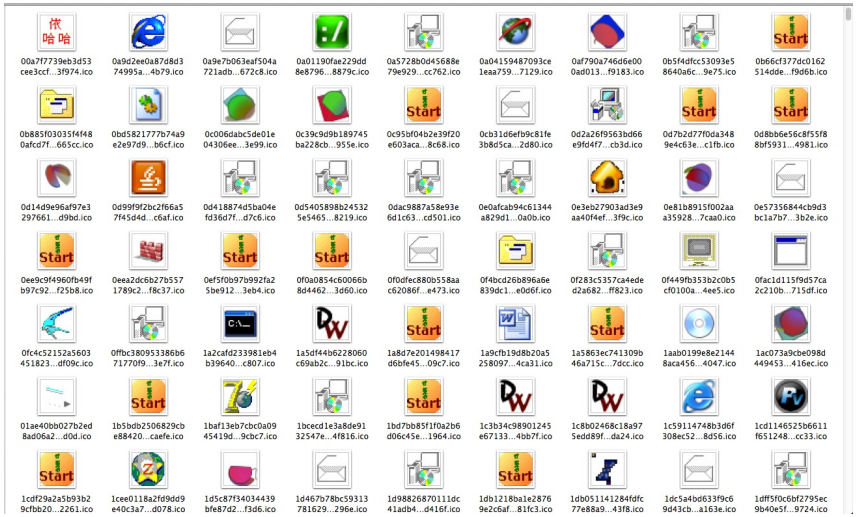
Icon Extractor

(Python + pywin32 + PyQt4)

win32gui.ExtractIconEx (“normal .exe”)
QtGui.QPixmap.fromWinHBITMAP(“.ico”)



win32gui.ExtractIconEx (“malware .exe”)
QtGui.QPixmap.fromWinHBITMAP(“.ico”)



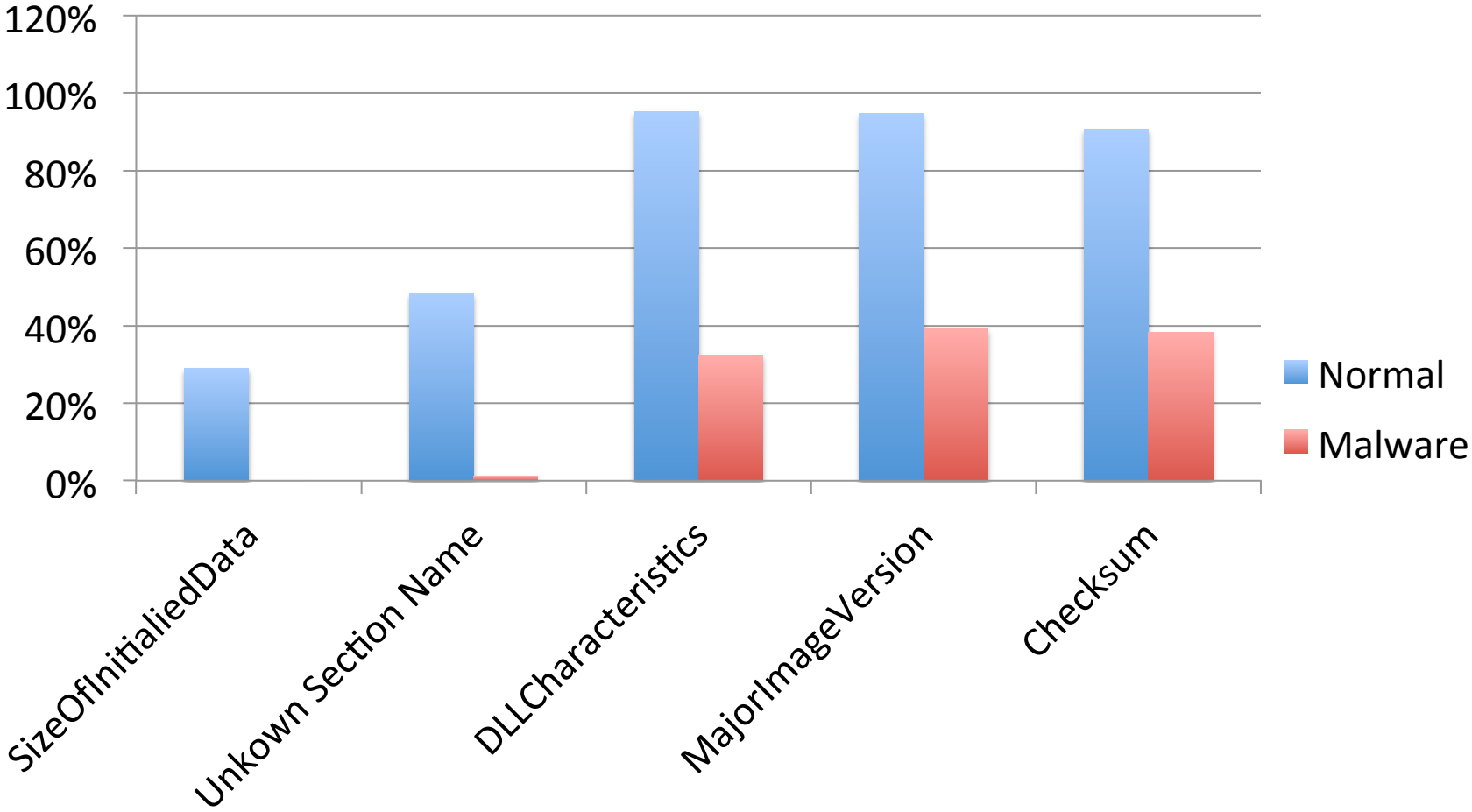
Experimental Setup

- Legitimate windows .exe dataset
  - Contains 1237 samples (less than 10MB)
  - www.download.com
  - www.softpedia.com
- Malware dataset
  - Contains 5598 samples
  - Provided by Dr. Perdisci

Result (Headers)

Index	Key Features	Malware (5598)	Normal (1237)	Difference
1	Size Of Initialized Data == 0	1626 (29%)	0 (0%)	29%
2	Unknown Section Name	2709 (48.4%)	16 (1.3%)	47.1%
3	DLL Characteristics == 0	5335 (95.3%)	401 (32.4%)	62.9%
4	Major Image Version == 0	5305 (94.8%)	486 (39.3%)	55.5%
5	Checksum == 0	5084 (90.8%)	474 (38.3%)	52.5%

Result (Headers cont.)



Result (Headers cont.)

Features combined	Malware detected	True positive	Normal detected	False positive
1,2	2811	50.2%	16	1.3%
1,2,3	5428	97.0%	10	0.8%
1,2,4	5402	96.7%	9	0.7%
1,2,5	5374	96.0%	3	0.2%
1,2,3,4	5506	98.4%	5	0.4%
1,2,3,5	5537	98.9%	3	0.2%
1,2,4,5	5482	97.9%	2	0.16%
1,2,3,4,5	5568	99.5%	2	0.16%

Result (Embedded Icons)

- Normal (813 of 1237)
- Malware (2165 of 5598)
  - the most prevalent icons which are seldom seen in legitimate PE files

    : 366  : 338  : 181
  - Misleading Icons

Conclusion

- It's possible to identify the malware by looking at some key fields in headers such as Checksum, Section Name, Initialized Data Size, DLL Characteristics and Major Image Version
- Looking at the PE header is faster. (Less than 20min)
- We can identify the malware by extracting the embedded icons such as the prevalent icons or misleading icons as described below.

Thank you!

Questions?