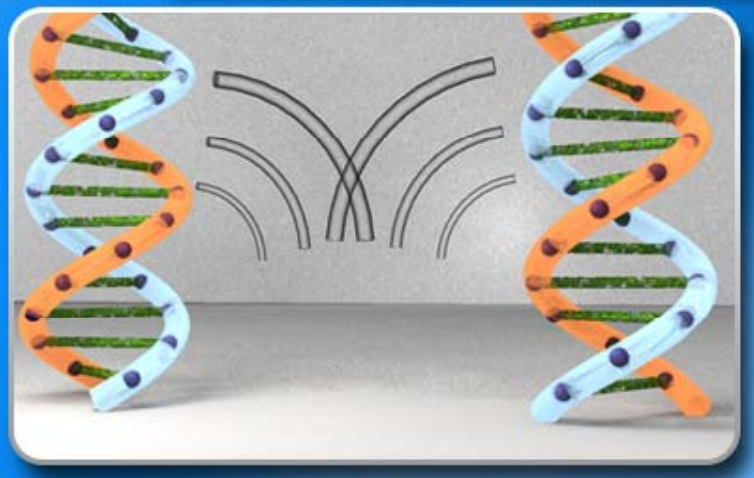
PE-Probe: Leveraging Packer Detection and Morphological Information to Detect Malicious Portable Executables



M. Zubair Shafiq, S. Momina Tabish, Muddassar Farooq

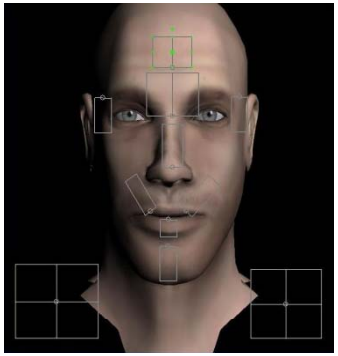
Next Generation Intelligent Networks Research Center (**nexGIN RC)**
National University of Computer and Emerging Sciences
Islamabad, Pakistan
http://www.nexginrc.org/

Agenda

- Projects' Introduction
- Motivation & Problem Statement
- Proposed Solution
- Results
- Q/A

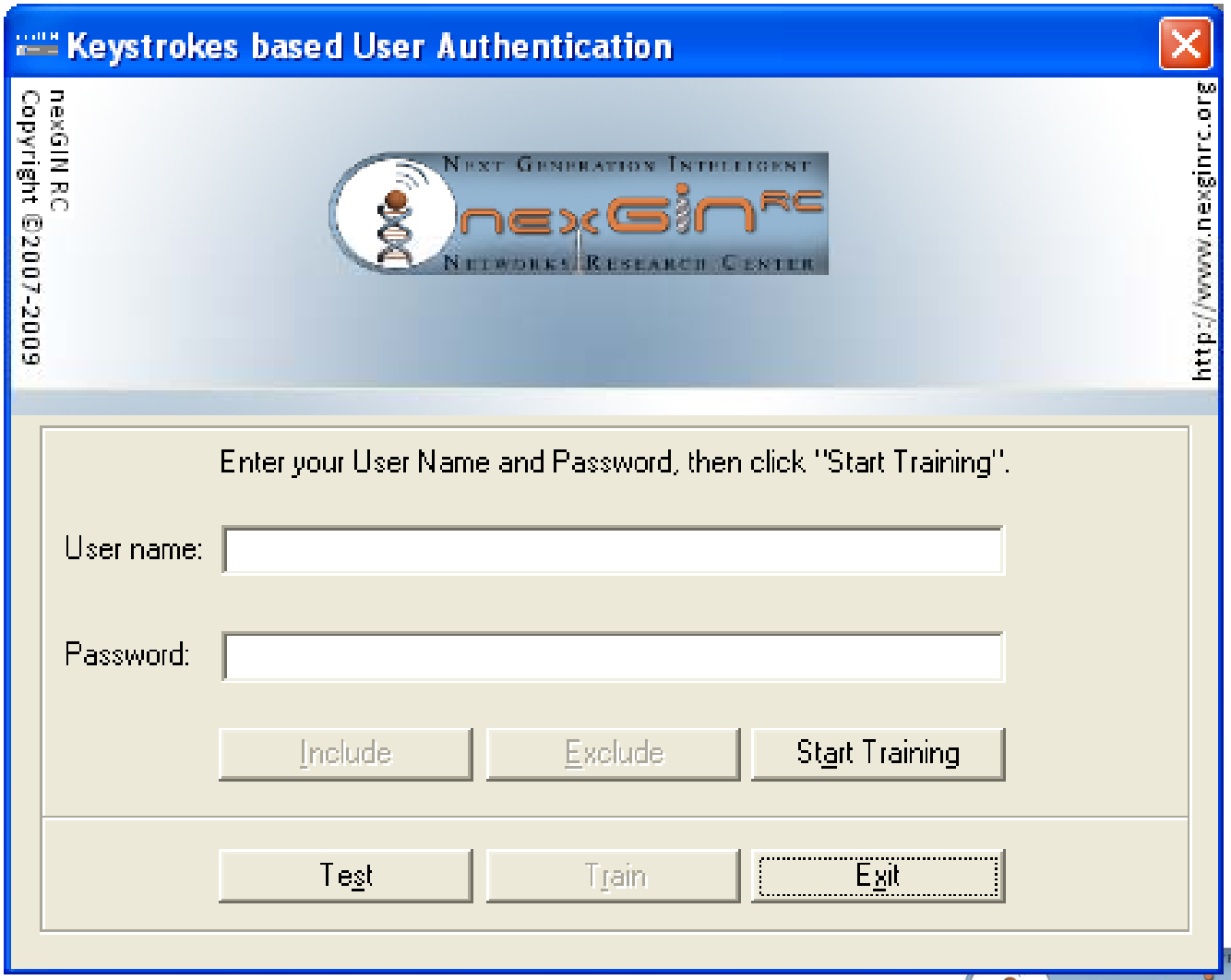
Its in your Hands, like its in your Eyes and Face

It is believed that keystrokes of people are distinct from each other just like their faces, finger prints, and eyes

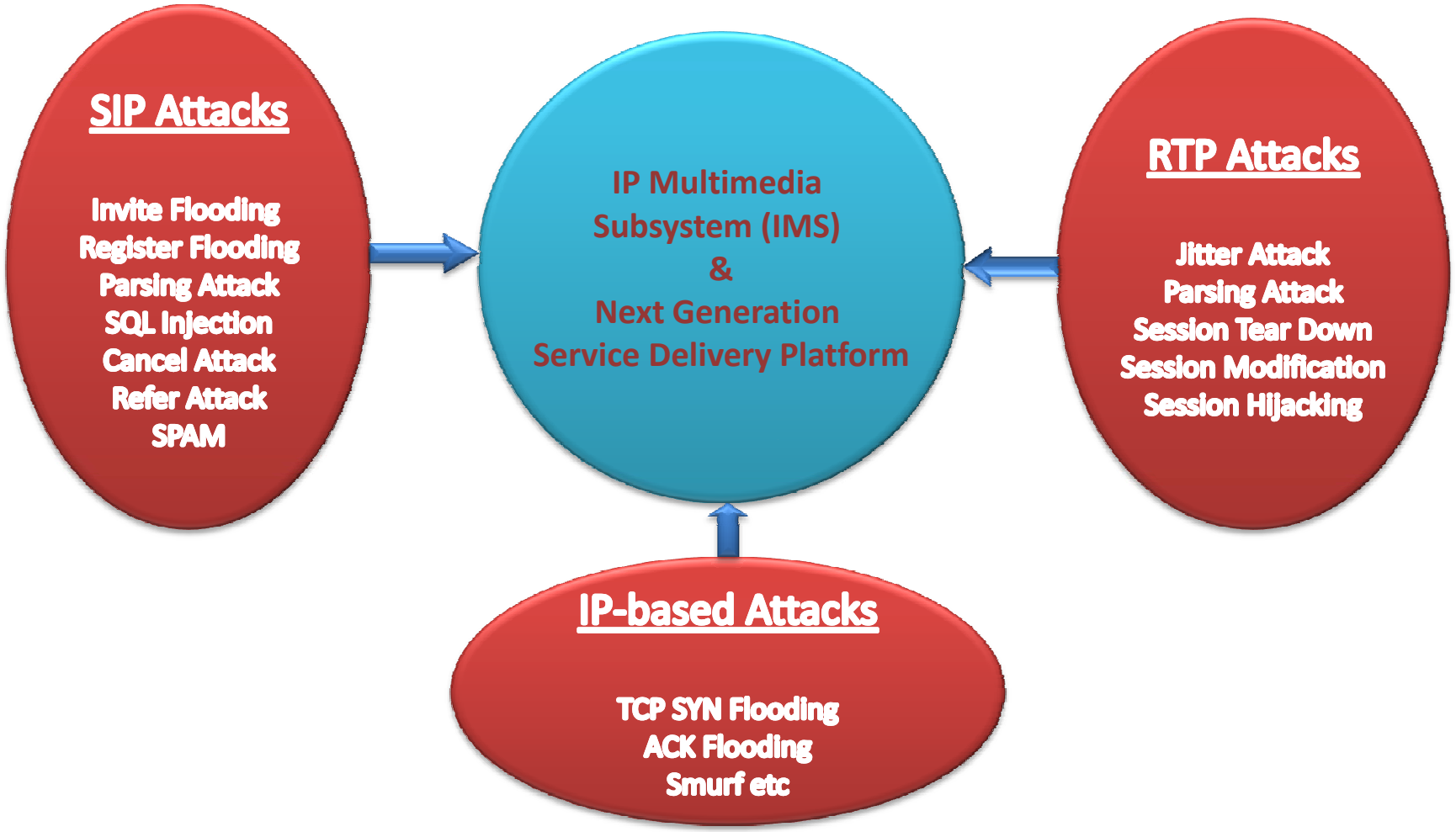


Doesn't require any extra hardware for identification

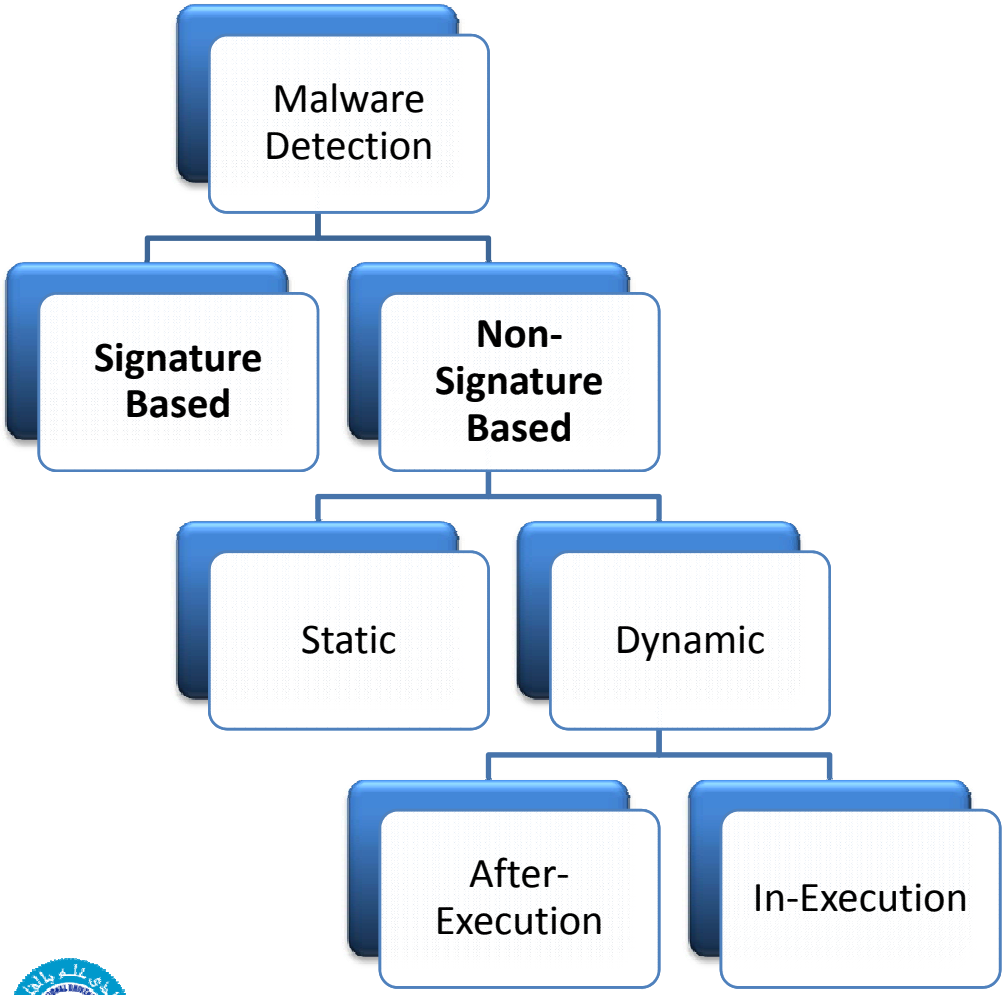
User Authentication System



IMS Security Challenges



Malware Detection



Signature Based:

· Detection on the basis of known byte sequences

· Unable to detect new malware

· Regular updates required

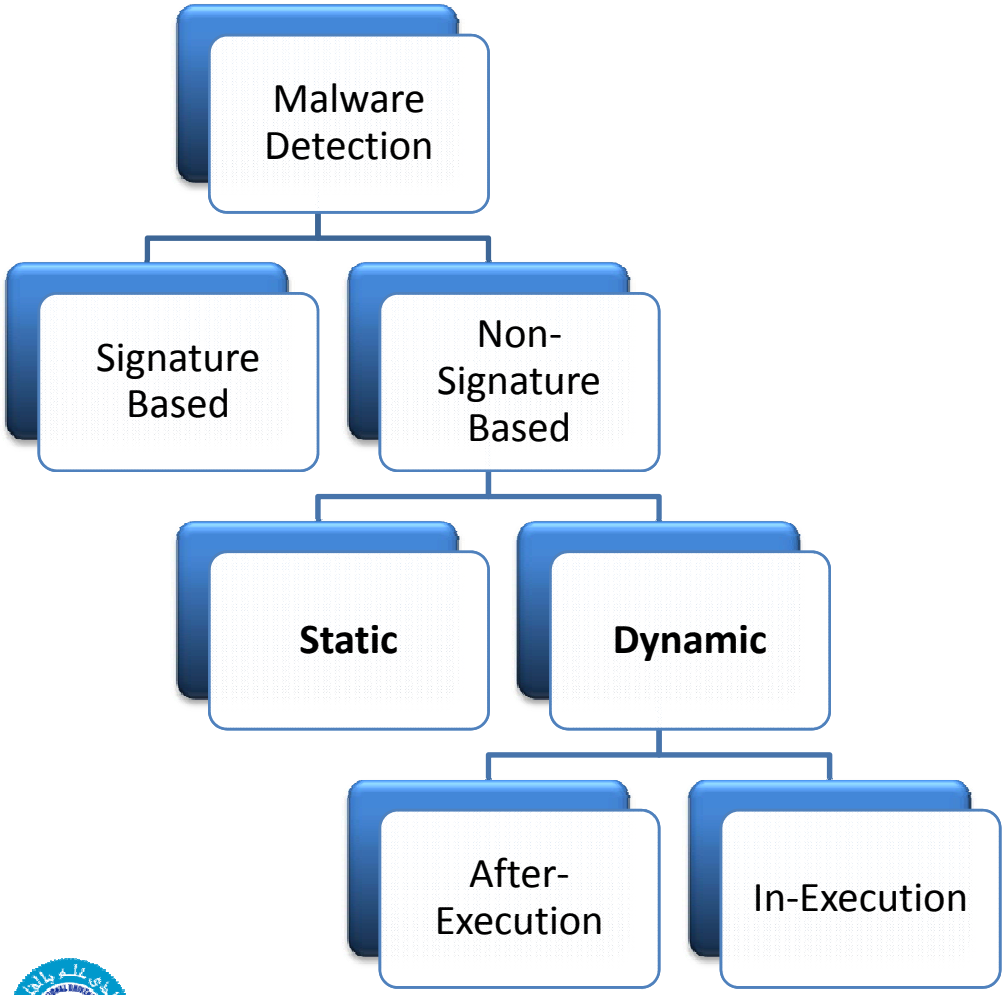
Non-Signature Based:

· Detection on the basis of smarter features

· Able to detect new malware

· Regular updates may not be necessary

Malware Detection



Static Detection:

· Detection on the basis of file as residing on secondary storage

· Prone to techniques such as code-obfuscation

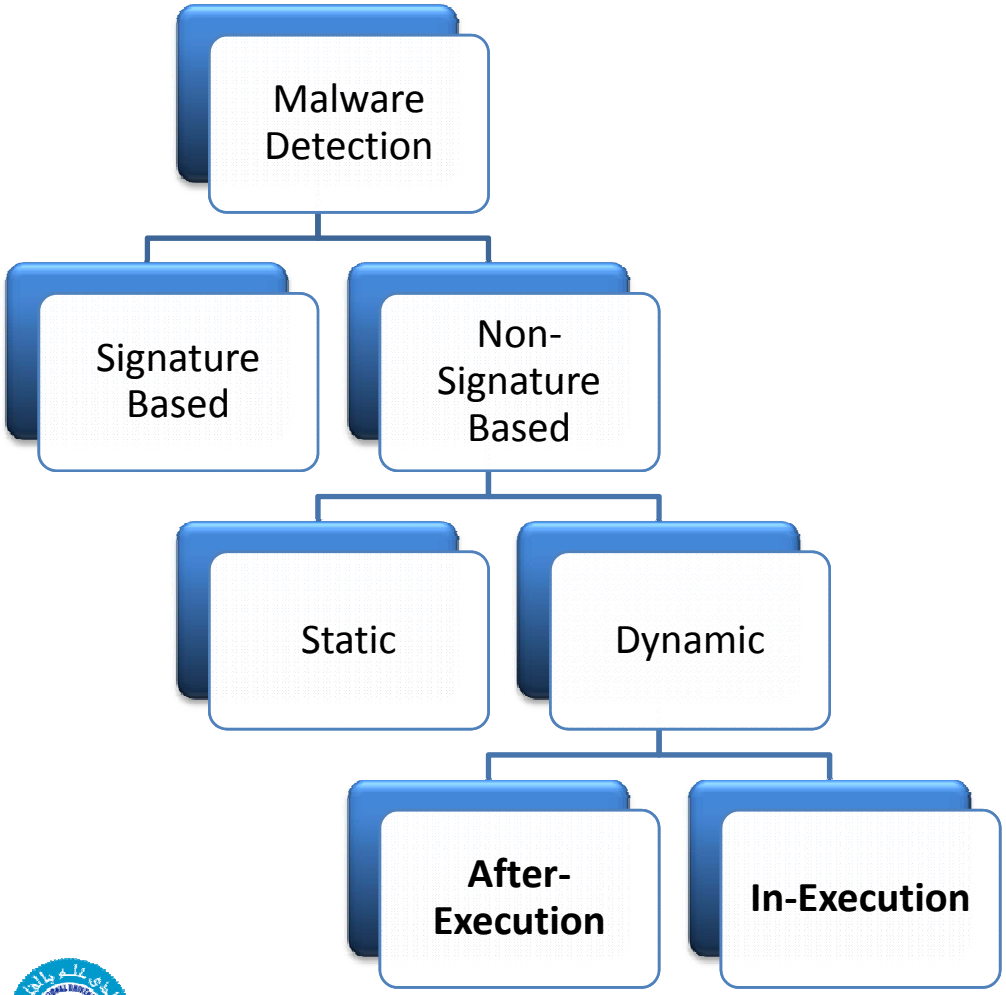
Dynamic Detection:

· Detection on the basis of run-time behavior (a more direct look)

· Resilient to techniques such as code obfuscation

· High processing overhead

Malware Detection



After-Execution Detection:

· Forensic Analysis

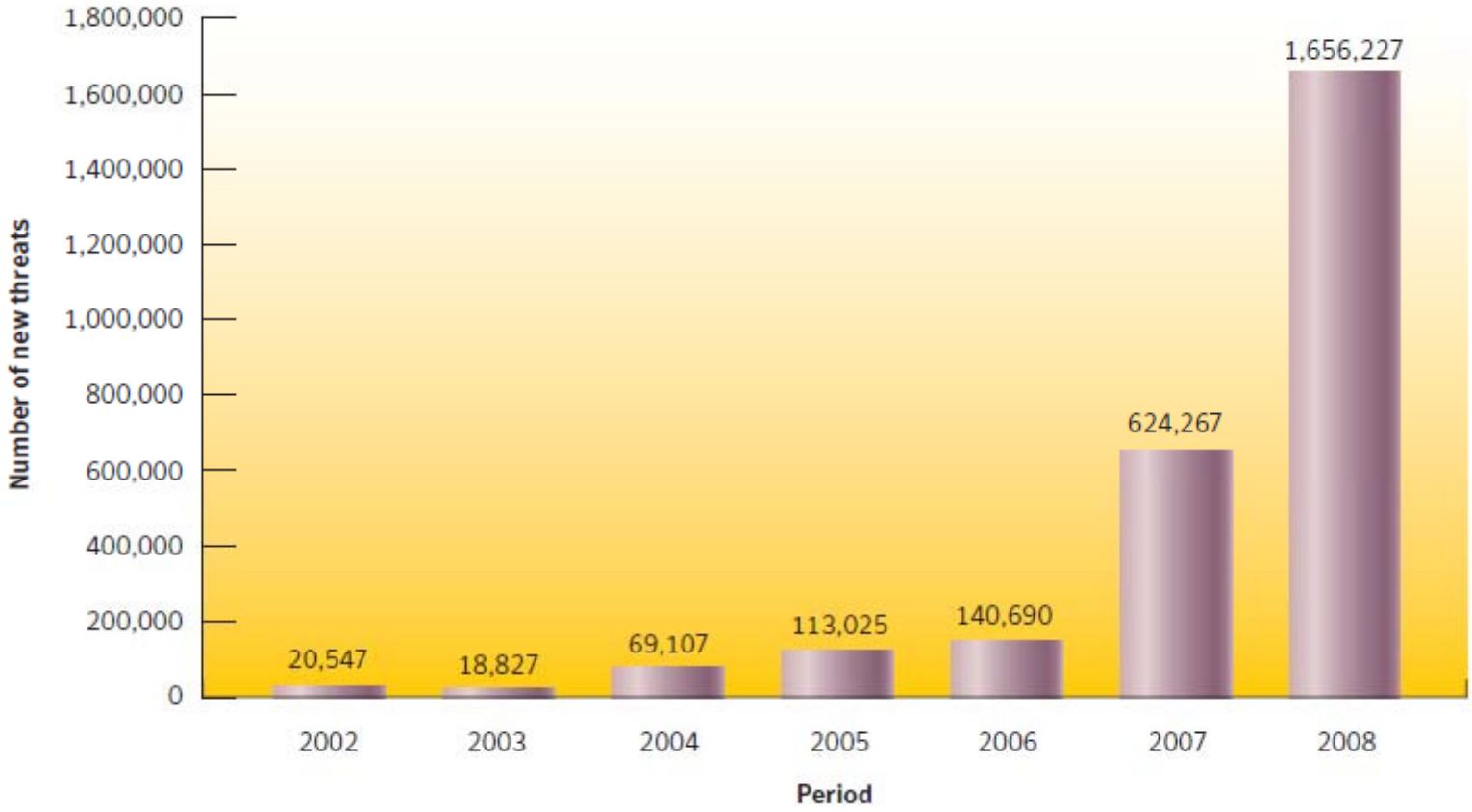
· Lower processing overhead

In-Execution Detection:

· End user tool

· High processing overhead

Motivation


1,800,000
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
Number of new threats
20,547
18,827
69,107
113,025
140,690
624,267
1,656,227
2002
2003
2004
2005
2006
2007
2008
Period

Motivation(2)

In Year 2008 Only [11]

- **5,491** new software vulnerabilities
- **1.6 million** new malware signatures
- **245 million** new attacks
- **1 Trillion** dollar in revenues

Motivation(3)

	Norton AV	Command AV	McAfee AV
Chernobyl-1.4	Not detected	Not detected	Not detected
F0sf0r0	Not detected	Not detected	Not detected
Hare	Not detected	Not detected	Not detected
Z0mbie-6.b	Not detected	Not detected	Not detected

Motivation(4)

Issues with Commercial Anti-virus software

- Cannot detect new malware
- Size of signature database cannot scale
- Signatures are evaded by code obfuscation techniques (such as packing)

Motivation(5)

Packing of Malware [12]

- **50%** new malware are simply re-packed versions of known malware

- **92%** malware use packing techniques

Motivation(6)

Non-signature based Malware Detection Schemes

- Machine-level code
- Disassembled code
- Static calls from disassembled code
- Run-time API calls

Motivation(7)

Issues with Non-signature based Schemes

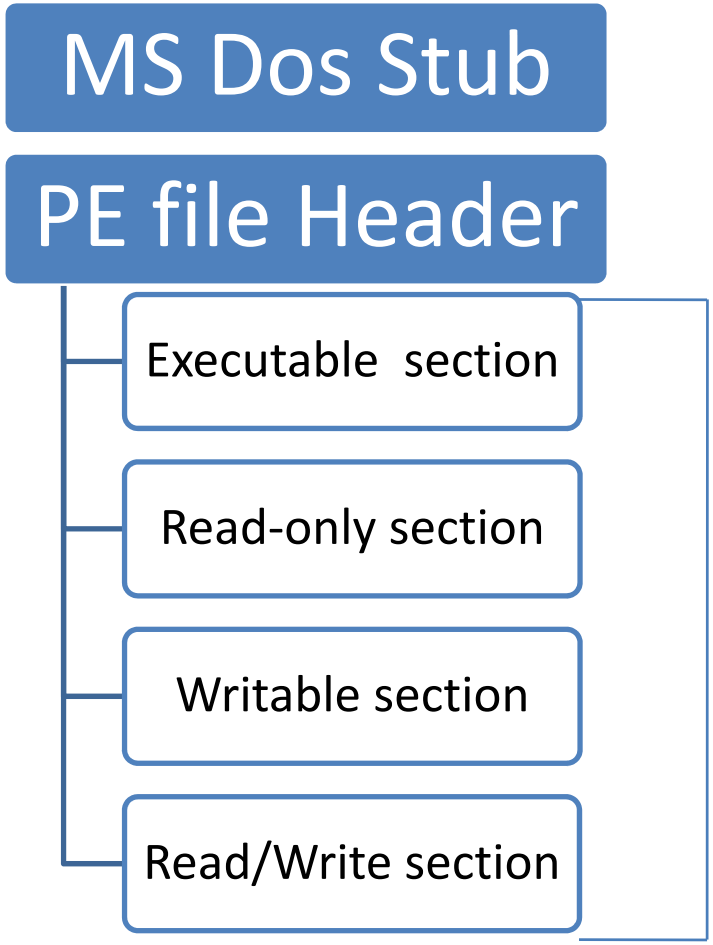
- High run-time computational complexity
- High false alarm rates
- Low reliability (e.g. crash, halt, evasion)

Problem Statement

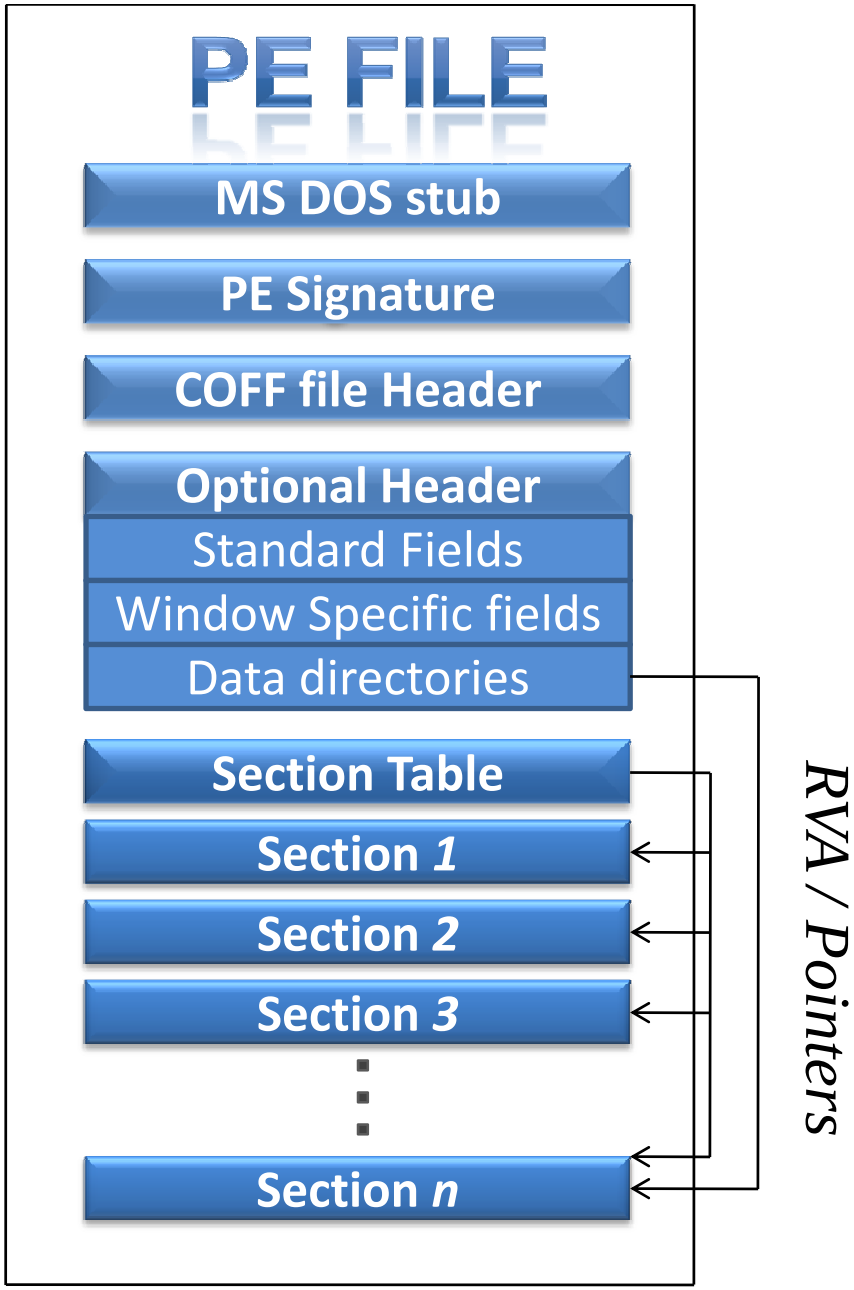
Problem Statement

- Non-signature based solution
- Low run-time complexity
- Low false alarms
- Robustness to Packing
- Must not use an unpacker for detection

PE File



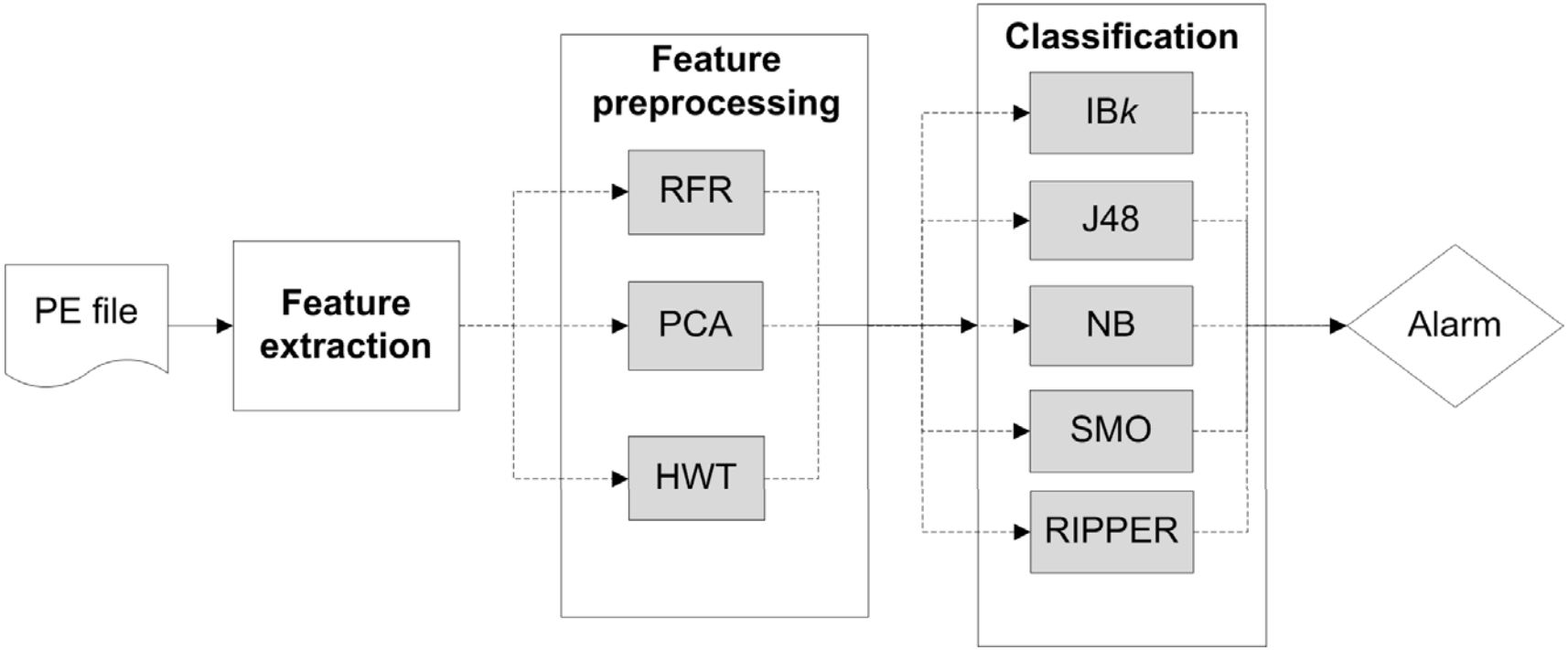
Existing non signature based Schemes are based on this area of PE file


PE FILE
MS DOS stub
PE Signature
COFF file Header
Optional Header
Standard Fields
Window Specific fields
Data directories
Section Table
Section 1
Section 2
Section 3
Section n
RVA / Pointers

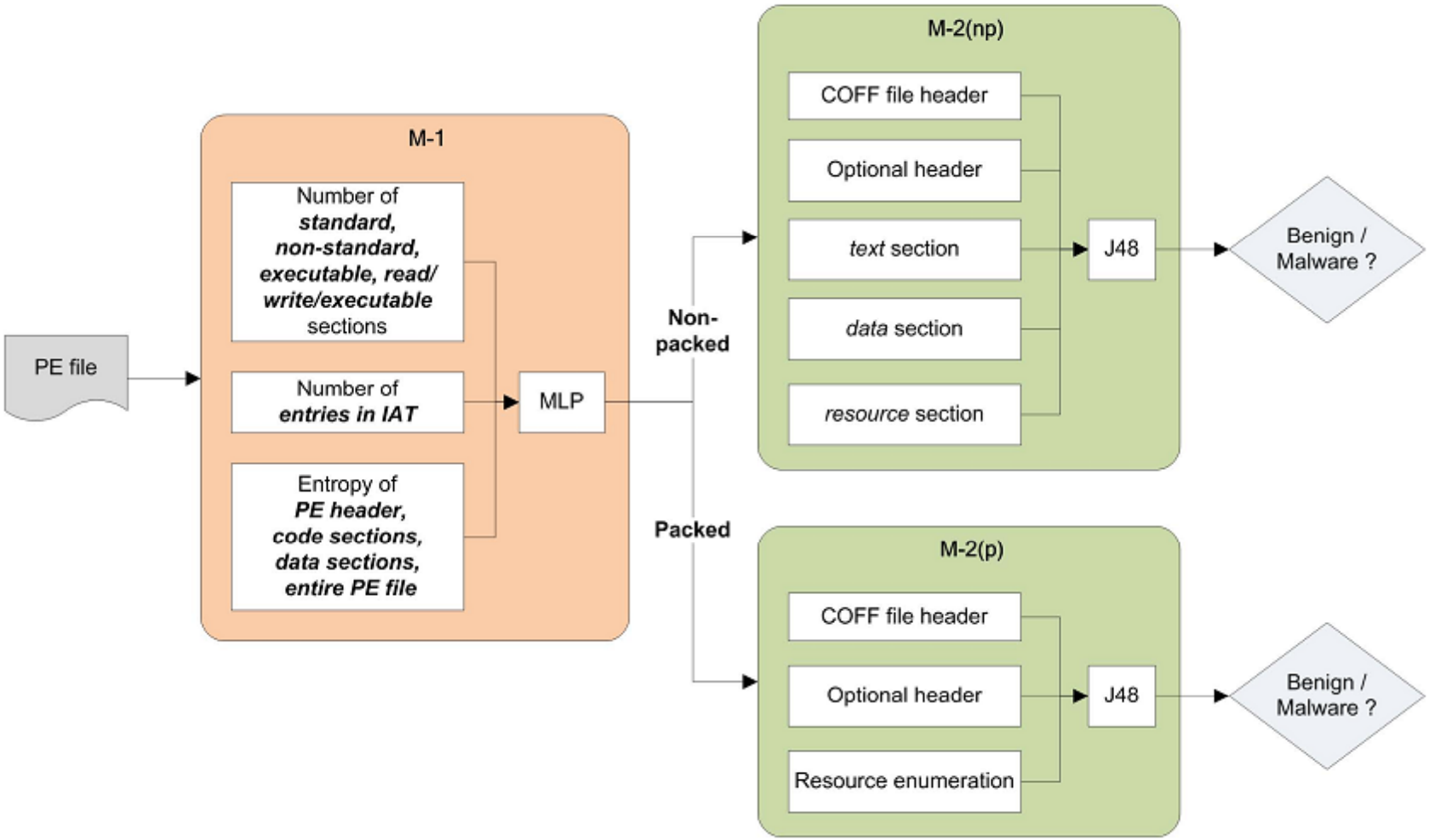
List of Features from PE file

Feature Description	Type	Quantity
DLLs referred	binary	73
COFF file header	integer	7
Optional header – standard fields	integer	9
Optional header – Windows specific fields	integer	22
Optional header – data directories	integer	30
.text section – header fields	integer	9
.data section – header fields	integer	9
.rsrc section – header fields	integer	9
Resource directory table & resources	integer	21
Total		189

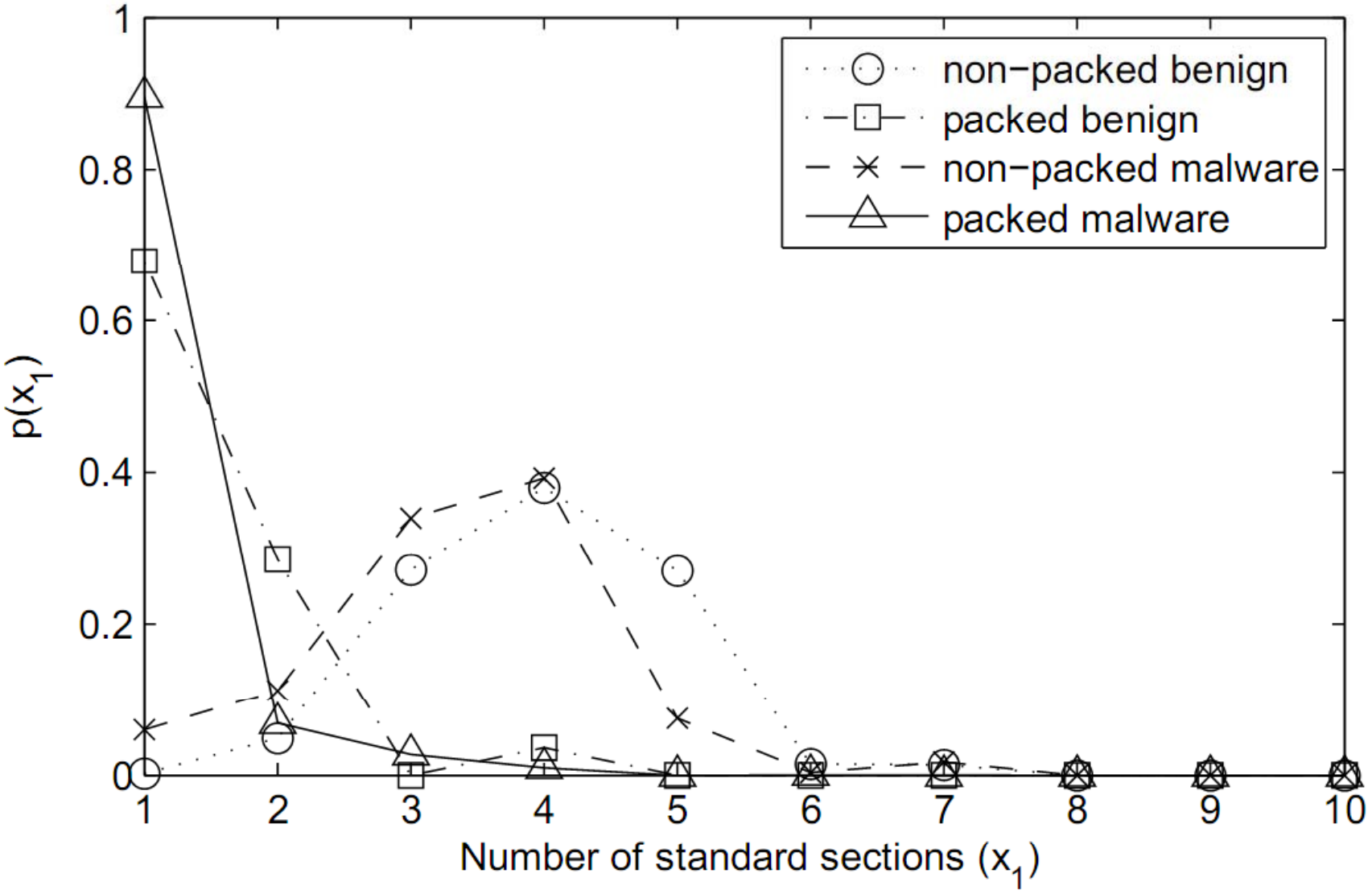
PE Miner


PE file
Feature extraction
Feature preprocessing
RFR
PCA
HWT
Classification
IBk
J48
NB
SMO
RIPPER
Alarm

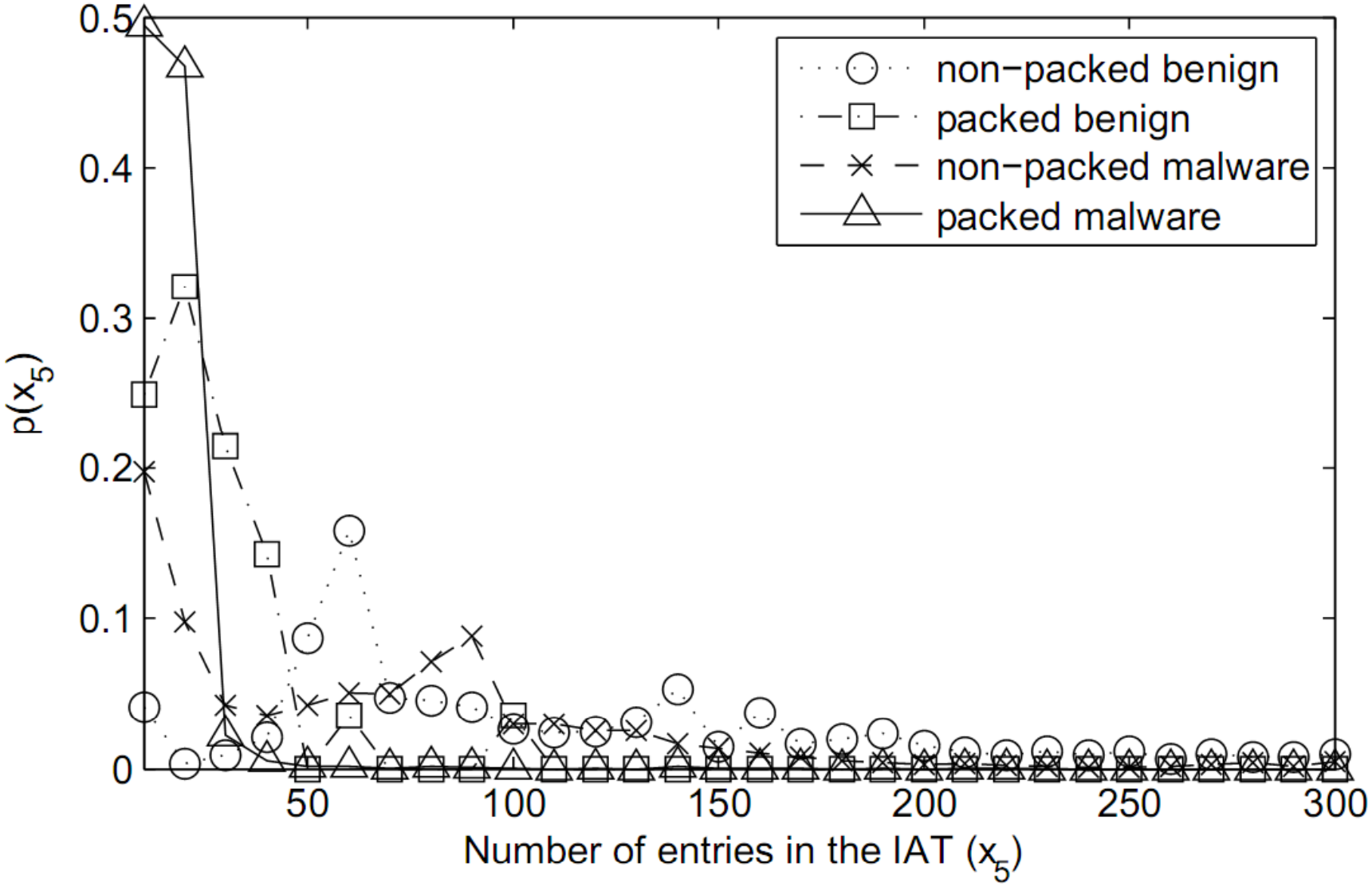
Architecture of PE-Probe



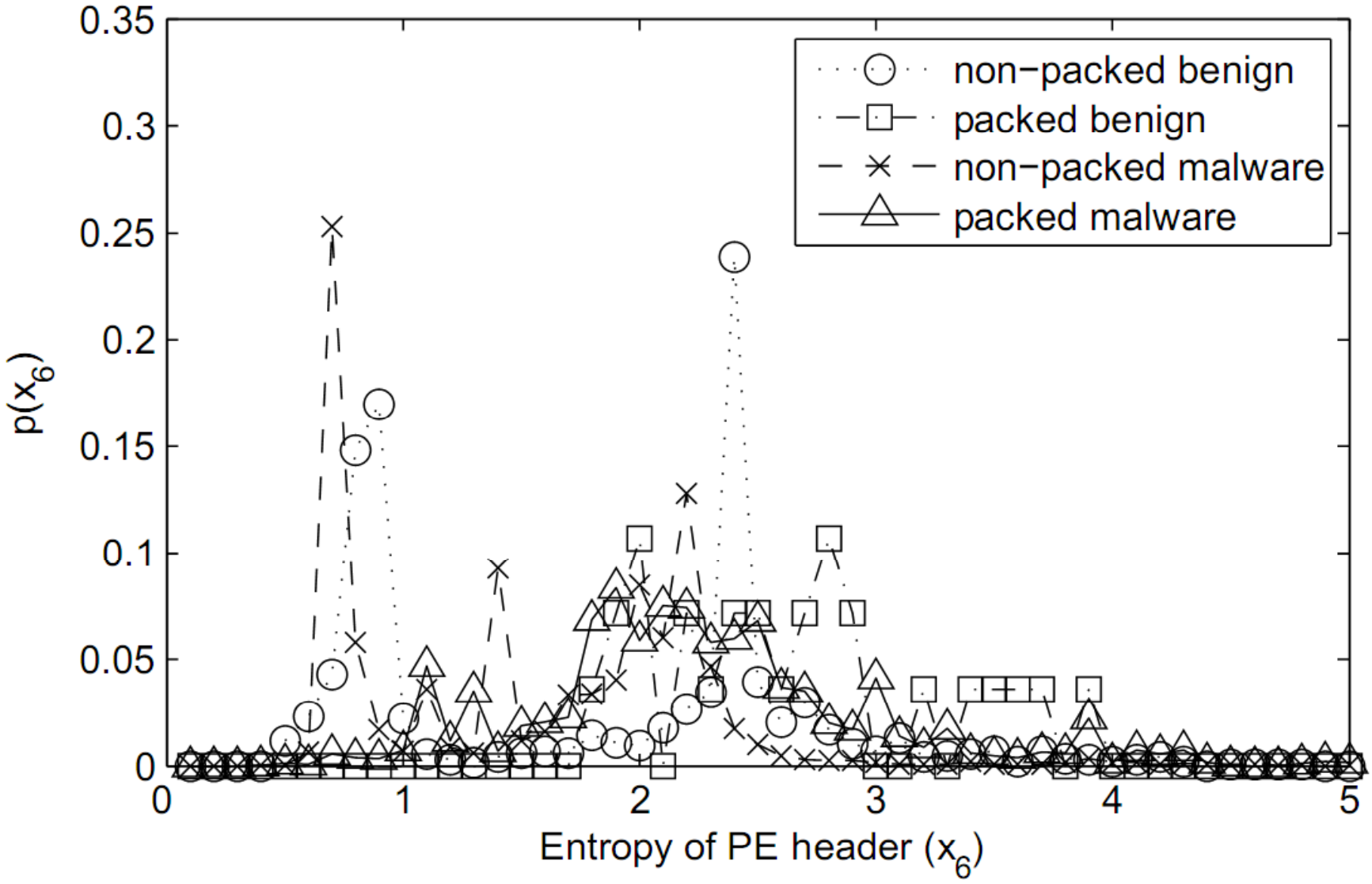
Distribution of Number of standard sections



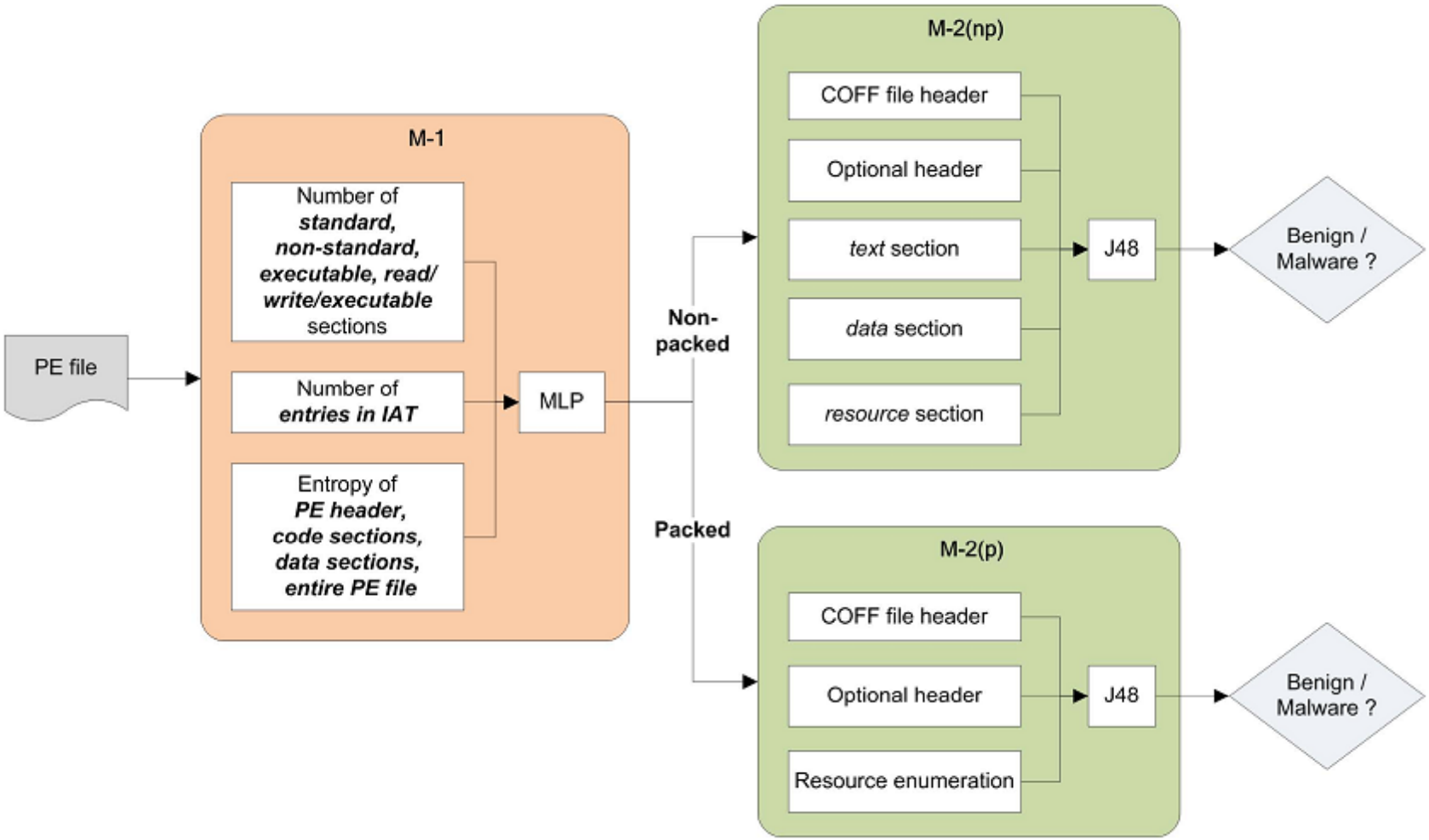
Distribution of Number of entries in Import Address Table



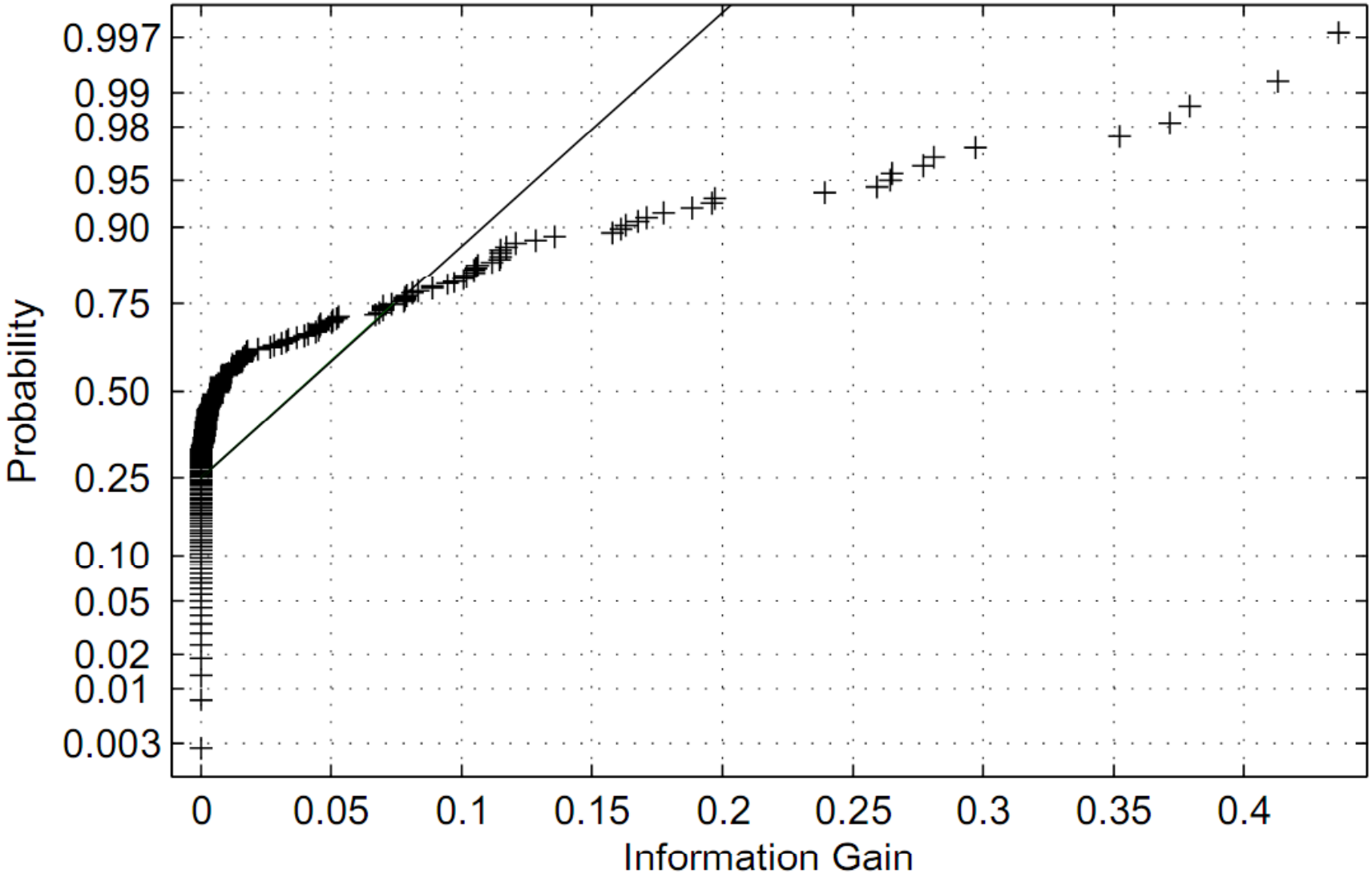
Distribution of Entropy of PE Header

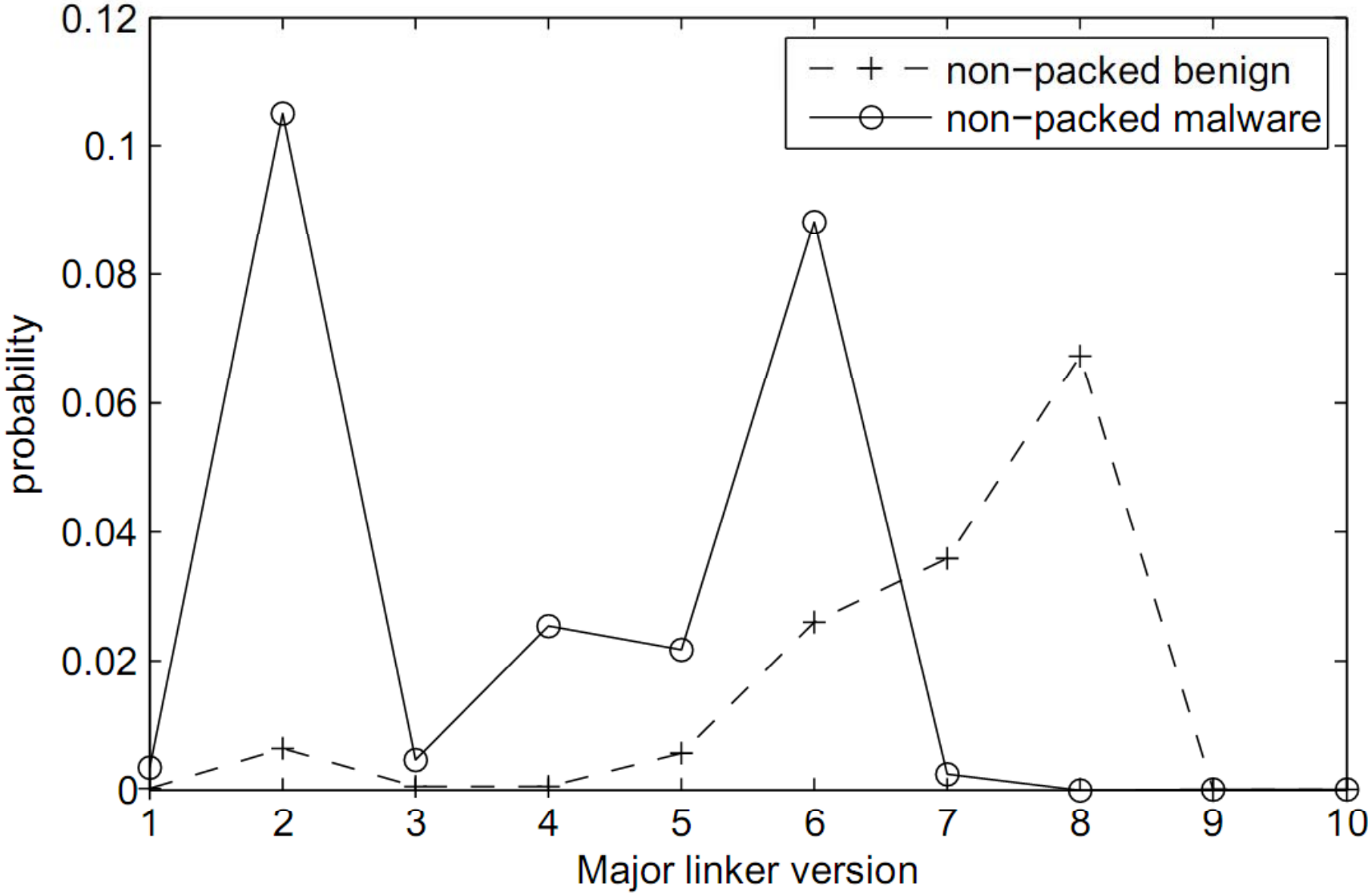


Architecture of PE-Probe



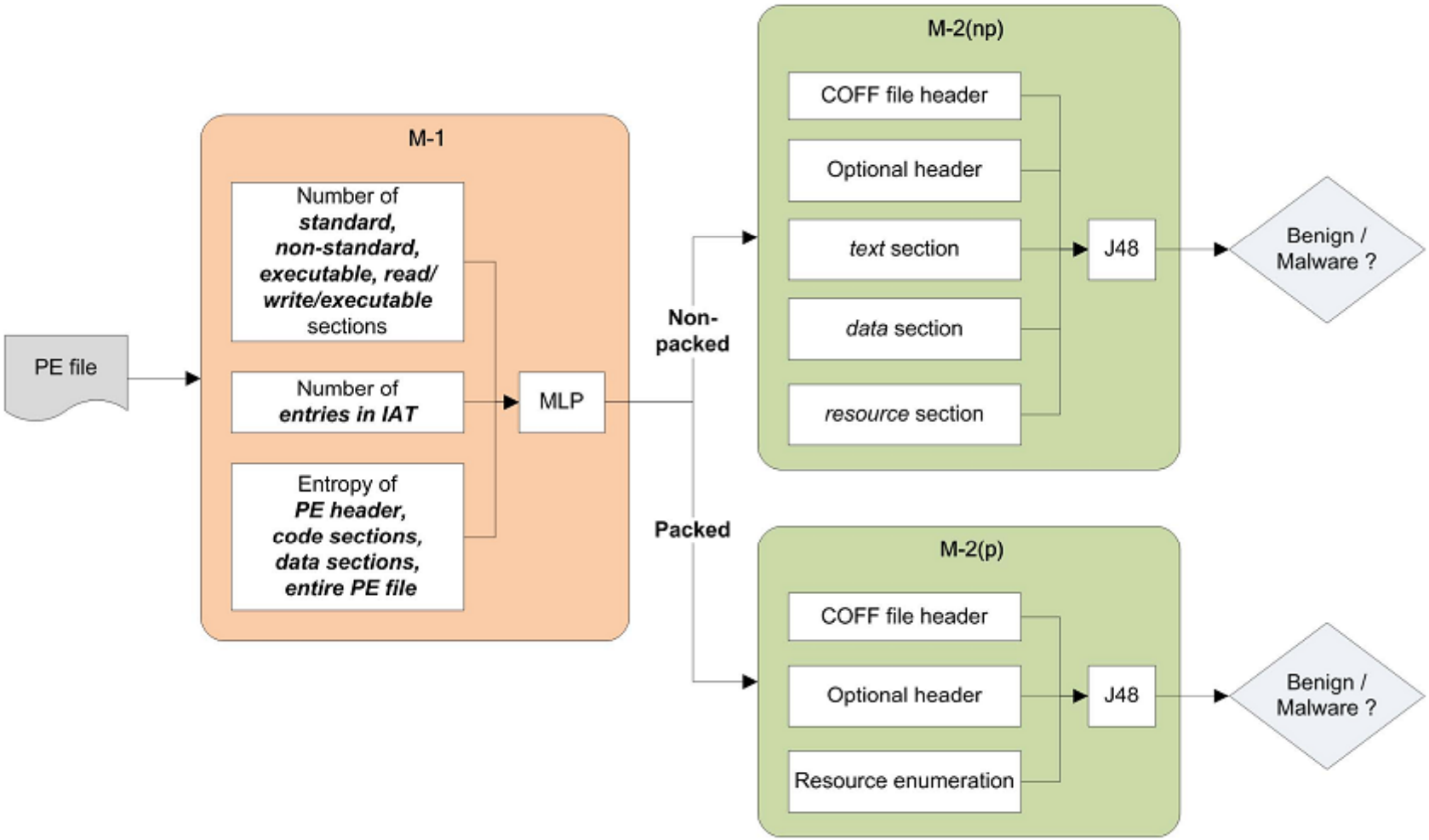
Structural features for non-packed PE files



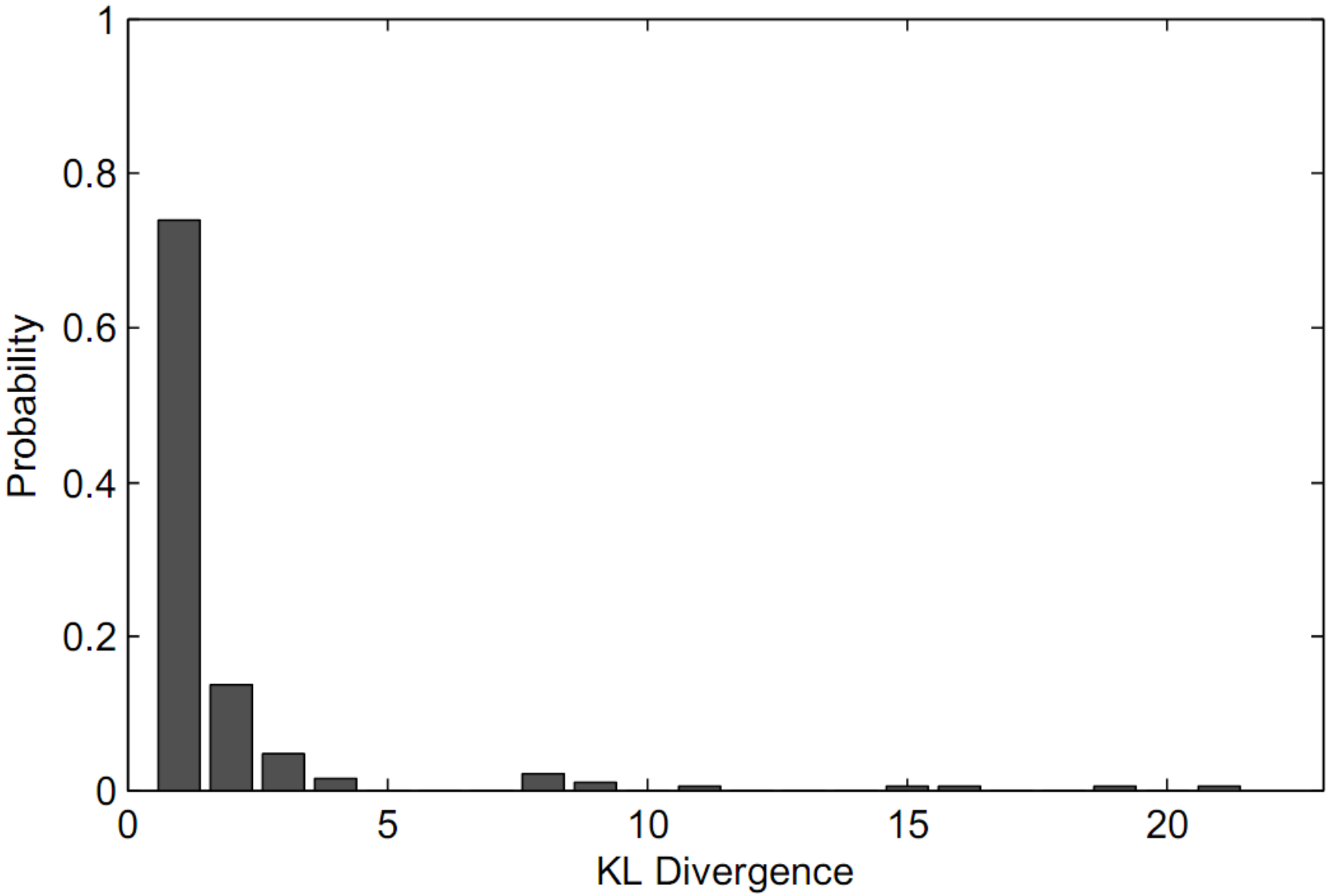


Distribution plot for “major linker version” feature

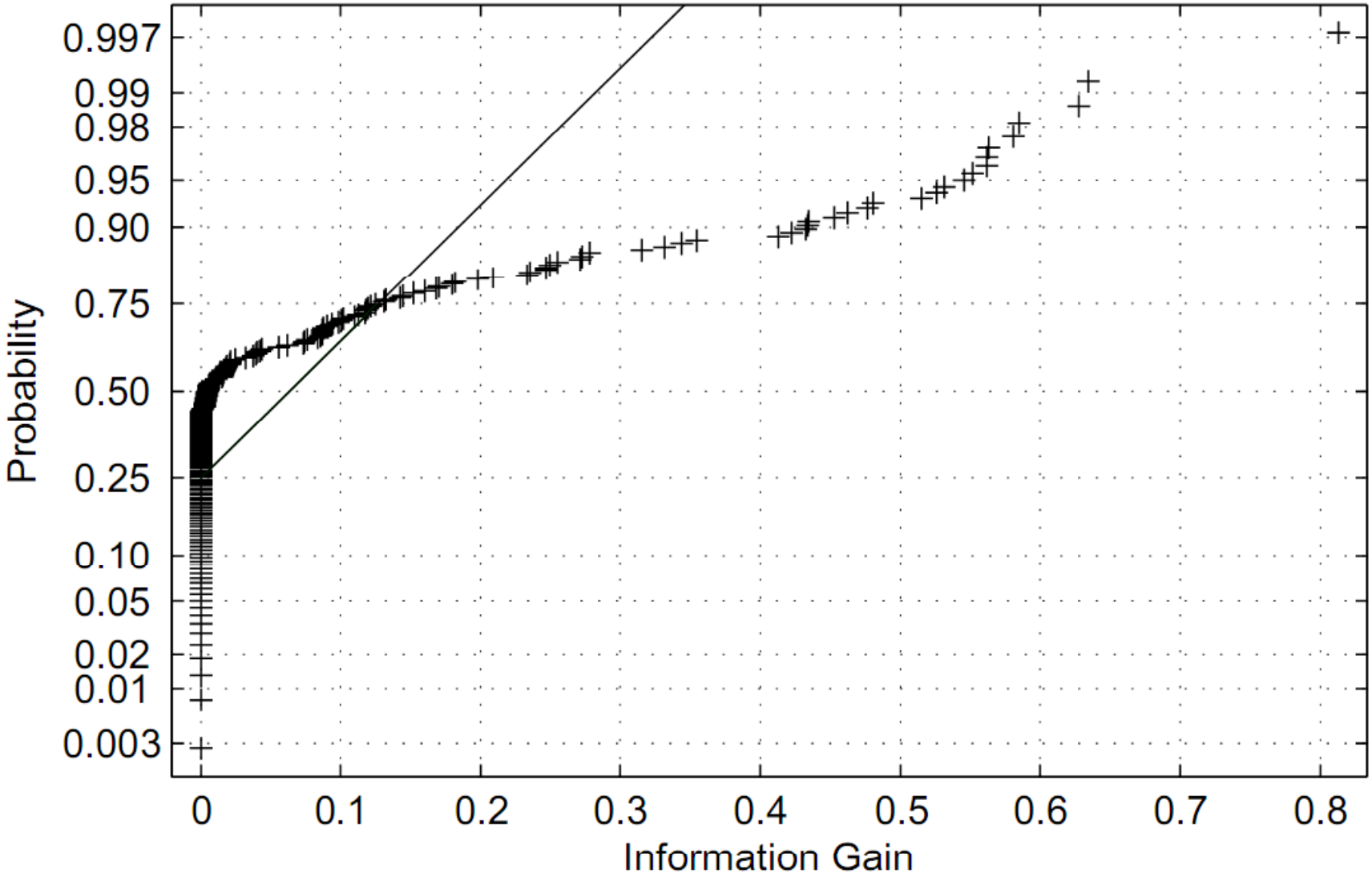
Architecture of PE-Probe



KL Divergence of features of packed/non-packed PE files



Structural features for packed PE files



Results

Dataset – Offensive Computing

Malware Category	Minimum Size (B)	Maximum Size (MB)	Average Size (KB)	Packed (%)
Benign	817	107.1	371.6	1.9
Backdoor	43	25.7	224.9	56.3
Constructor	62	7.23	265.4	50.0
DoS	610	1.33	131.0	50.6
Email Flooder	894	15.0	294.5	48.3
Email Worm	28	45.3	50.3	62.9
Exploit	57	4.6	133.6	43.6
Flooder	248	1.6	169.8	49.9
Hoax	25	8.7	87.2	47.2
AdWare	68	24.2	548.7	46.9
FraudTool	36	8.1	170.4	56.9
Porn	7008	0.2	103.2	59.1
Rootkit	75	2.7	85.1	57.3
Virtool	31	2.9	84.0	51.0
Worm	28	10.4	394.8	53.2
Virus	10	12.5	54.5	35.8
Trojan	8	46.8	252.6	39.2

Classification Metrics

$$\text{TP rate} = \frac{TP}{TP+FN}$$

$$\text{FP rate} = \frac{FP}{FP+TN}$$

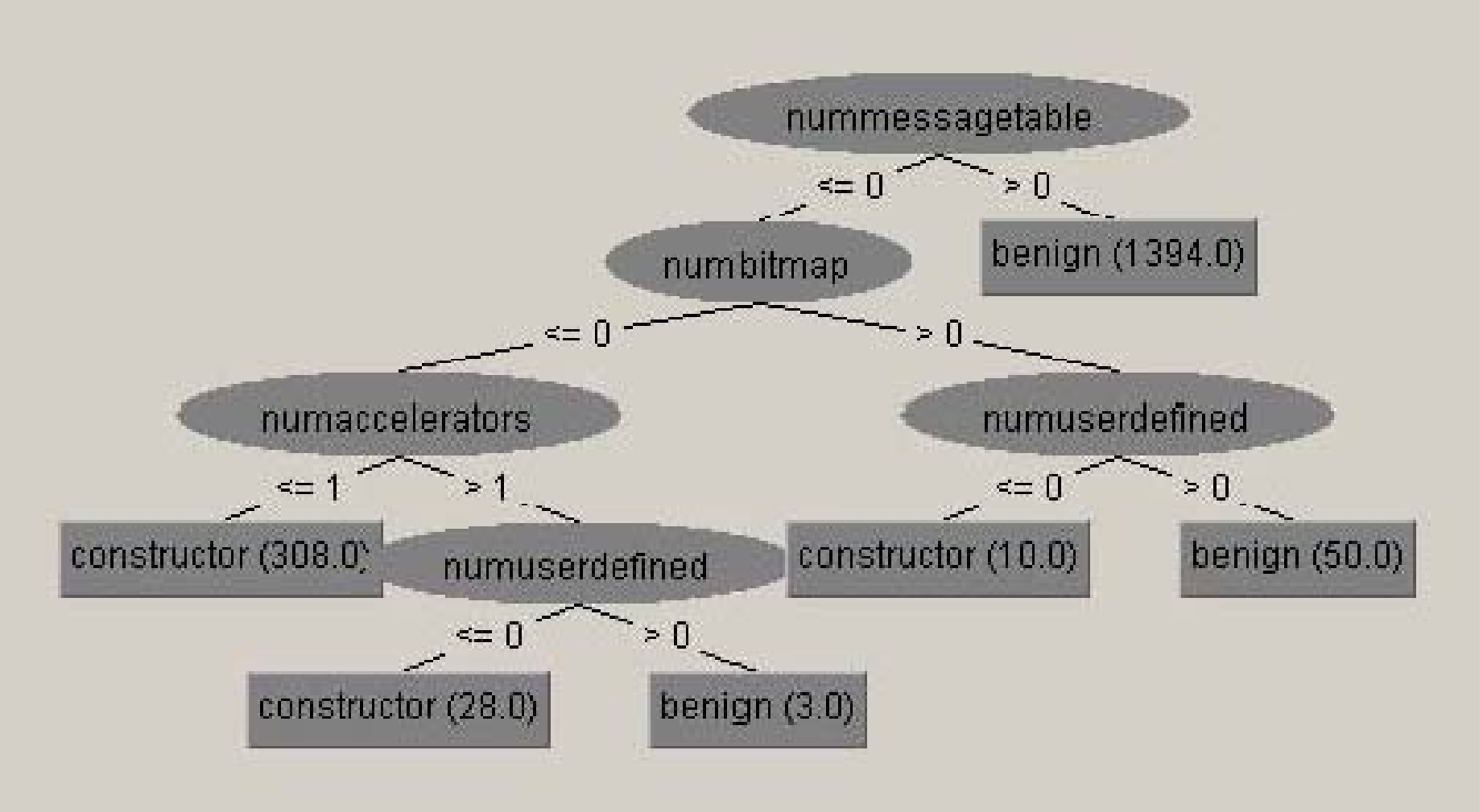
Accuracy of PE-Probe

Malware Category	Detection Accuracy			
-	Packed		Non-Packed	
-	Detection Rate	False Alarm Rate	Detection Rate	False Alarm Rate
Backdoor	0.999	0.001	0.998	0.002
Constructor	0.997	0.003	0.995	0.005
DoS	0.997	0.003	0.999	0.001
Email Flooder	0.995	0.005	0.996	0.004
Email Worm	0.995	0.005	0.999	0.001
Exploit	0.996	0.004	0.996	0.04
Flooder	0.997	0.003	1.000	0.000
Hoax	0.998	0.002	0.990	0.010
AdWare	0.989	0.011	0.978	0.022
FraudTool	0.998	0.002	0.981	0.019
Porn	1.000	0.000	1.000	0.000
Rootkit	0.998	0.002	1.000	0.000
Virtool	0.996	0.004	1.000	0.000
Worm	0.988	0.012	0.980	0.020
Virus	0.998	0.002	0.996	0.004
Trojan	1.000	0.000	1.000	0.000
Average	0.996	0.003	0.994	0.008

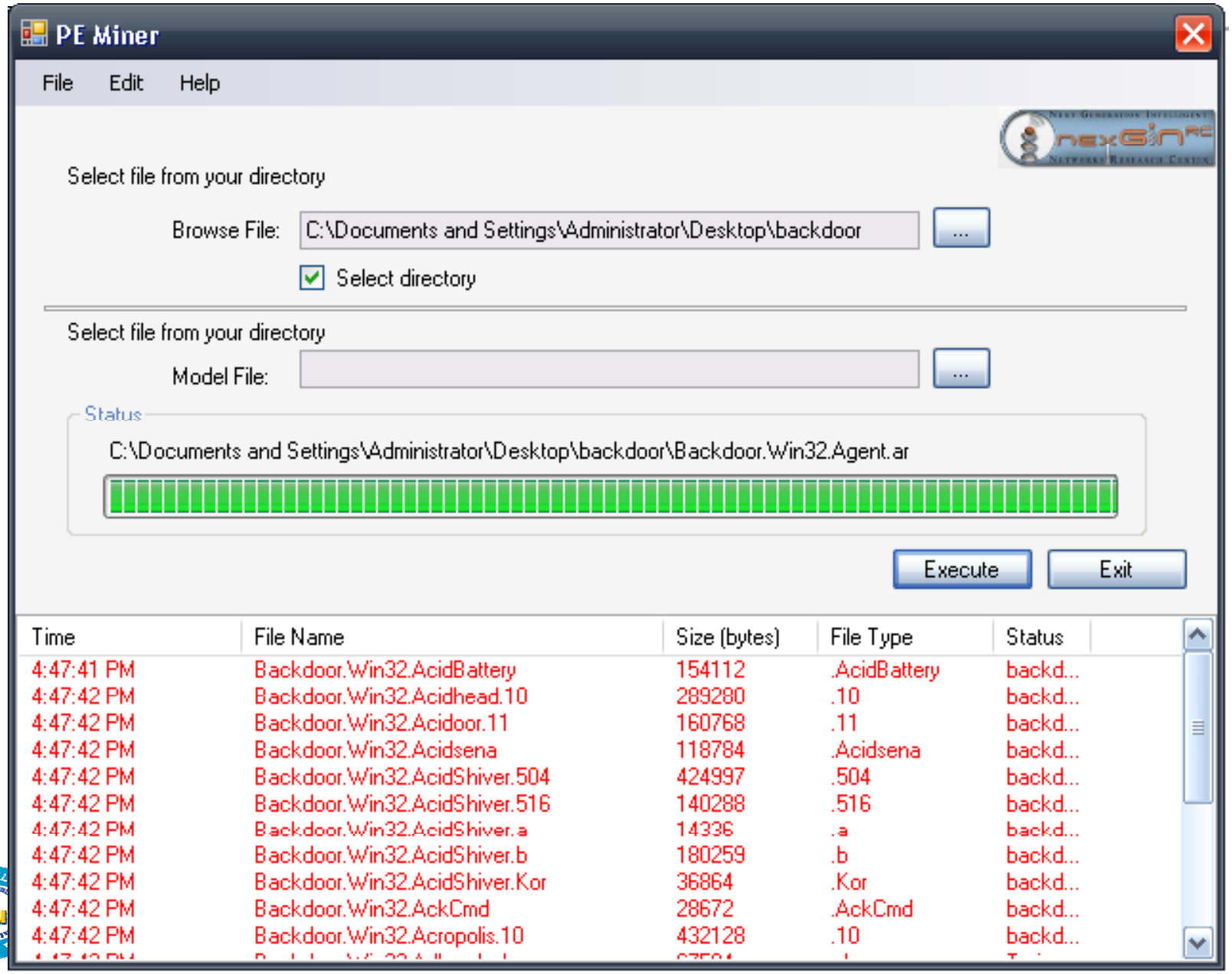
The processing overheads in (seconds/file)

		J48	NB	RIPPER	SMO	IBK	J48	NB	RIPPER	SMO
		TRAINING				TESTING				
PE-Miner(RFR)	-	0.0008	0.001	0.269	0.199	0.032	0.001	0.002	0.002	0.002
PE-Miner(PCA)	-	0.007	0.001	0.264	0.179	0.035	0.001	0.001	0.001	0.002
PE-Miner(HWT)	-	0.007	0.001	0.252	0.147	0.032	0.001	0.002	0.001	0.002
McBoost	-	0.021	0.004	1.305	1.122	0.218	0.010	0.007	0.005	0.022
Strings	-	0.009	0.002	0.799	0.838	0.163	0.003	0.003	0.002	0.003

Forensic Information

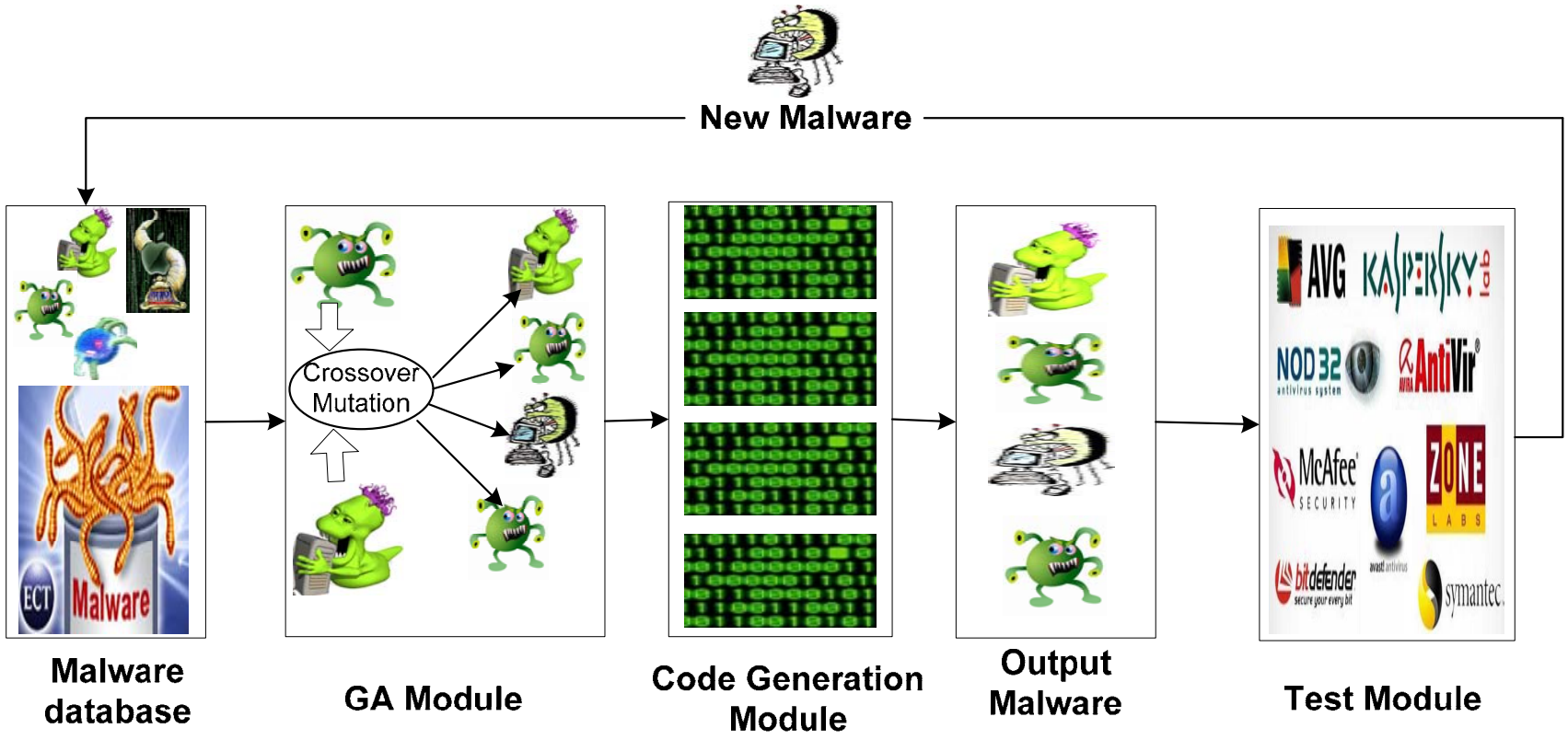


PE-Miner


PE Miner
File
Edit
Help
Select file from your directory
Browse File:
C:\Documents and Settings\Administrator\Desktop\backdoor
...
Select directory
Select file from your directory
Model File:
...
Status
C:\Documents and Settings\Administrator\Desktop\backdoor\Backdoor.Win32.Agent.ar
Execute
Exit
Time
File Name
Size (bytes)
File Type
Status
4:47:41 PM Backdoor.Win32.AcidBattery 154112 .AcidBattery backd...
4:47:42 PM Backdoor.Win32.Acidhead.10 289280 .10 backd...
4:47:42 PM Backdoor.Win32.Acidoor.11 160768 .11 backd...
4:47:42 PM Backdoor.Win32.Acidsena 118784 .Acidsena backd...
4:47:42 PM Backdoor.Win32.AcidShiver.504 424997 .504 backd...
4:47:42 PM Backdoor.Win32.AcidShiver.516 140288 .516 backd...
4:47:42 PM Backdoor.Win32.AcidShiver.a 14336 .a backd...
4:47:42 PM Backdoor.Win32.AcidShiver.b 180259 .b backd...
4:47:42 PM Backdoor.Win32.AcidShiver.Kor 36864 .Kor backd...
4:47:42 PM Backdoor.Win32.AckCmd 28672 .AckCmd backd...
4:47:42 PM Backdoor.Win32.Acropolis.10 432128 .10 backd...

Conficker Detected as a Backdoor

Evolvable Malware Framework



Conclusion

- PE Structural Information can be leveraged to detect malware
- Packing Robustness
- Machine Learning Classifiers can learn packed and non-packed models
- Robustness and Evasion analysis in accompanying PE-Miner paper in RAID 2009.
- Zero day detection of Conficker

ACKNOWLEDGEMENT

- Special thanks to National ICT R&D for funding this project.

QUESTIONS

For further information and research papers, visit http://www.**nexginrc**.org