MAGIC:
Malware Analysis to Generate Important Capabilities

Sean Kilgallon

Problem Statement

Complex and numerous malware

- Require adaptive-based techniques

Scarce datasets

- Demand extensive knowledge engineering work

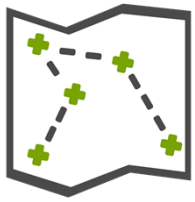
Dynamic Analysis

- Bottleneck for studying malware

MAGIC Model

Leverages **fast** static analysis

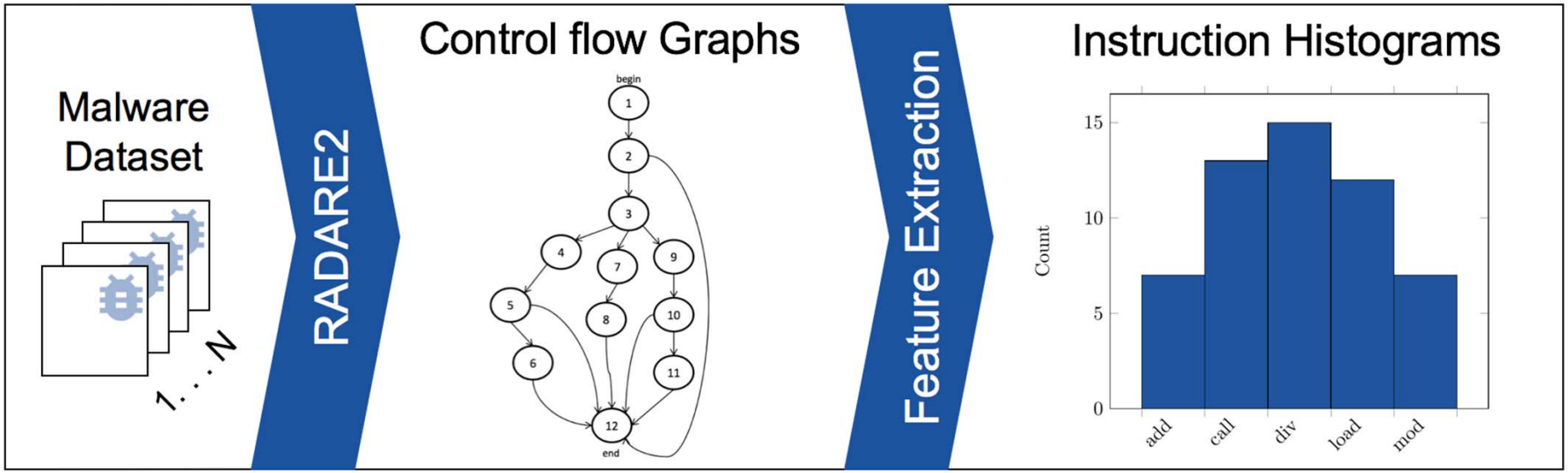
Decreases the need for **dynamic analysis**



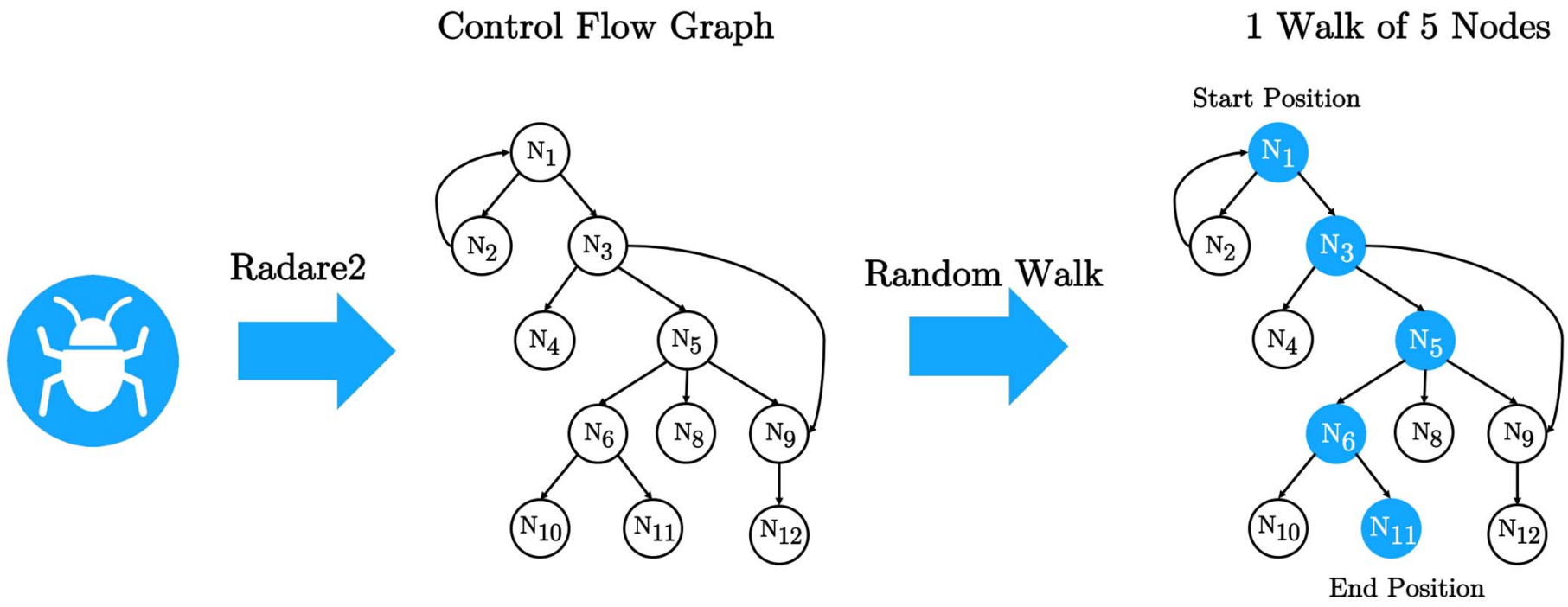
Maps **instructions** to specific **capabilities**

Highly **interpretable** and easy to **visualize**

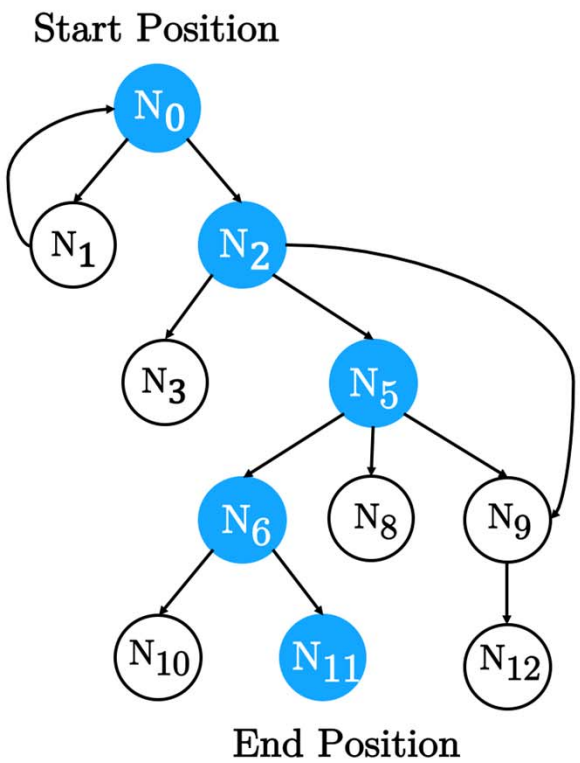
Static Characterization



Static Characterization (cont'd)



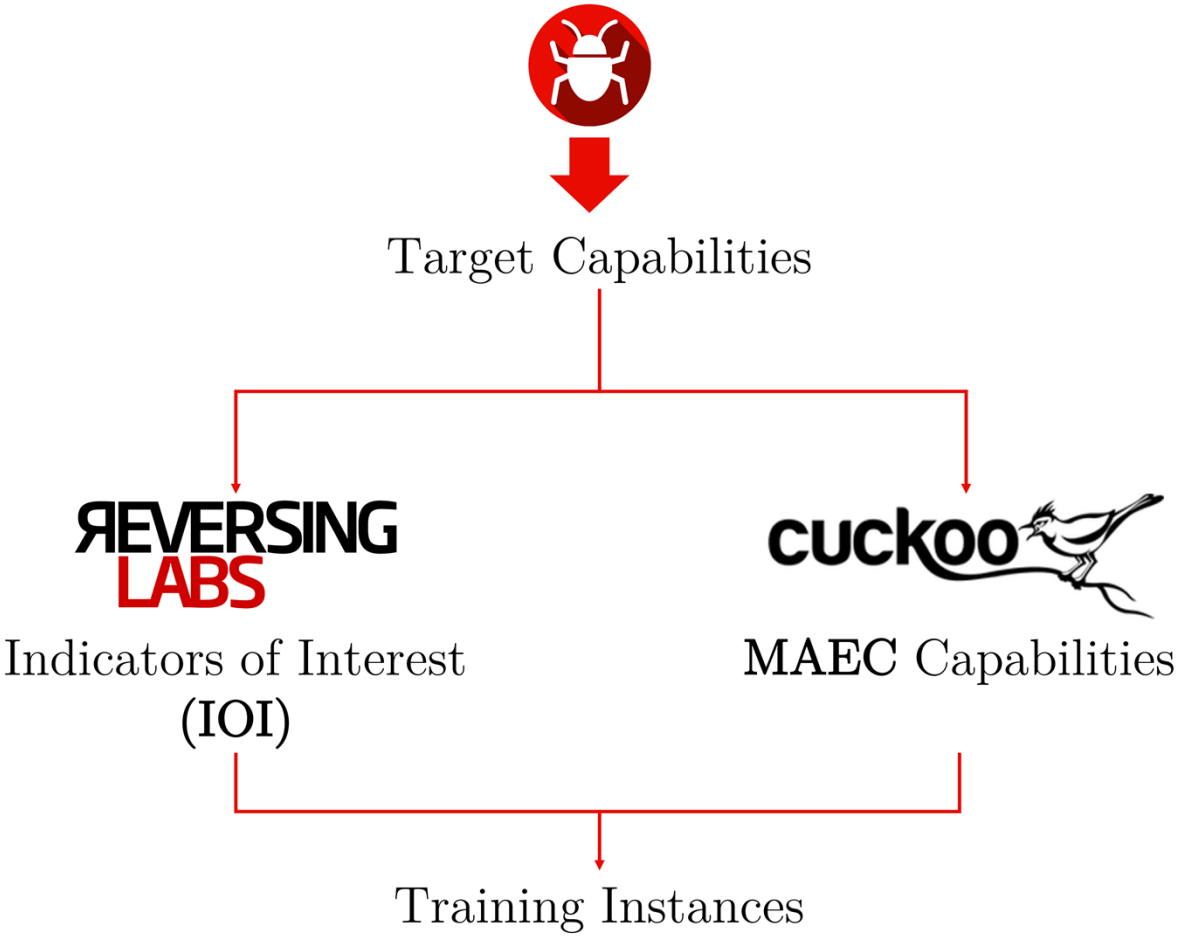
Static Characterization (cont'd)



Feature Generation

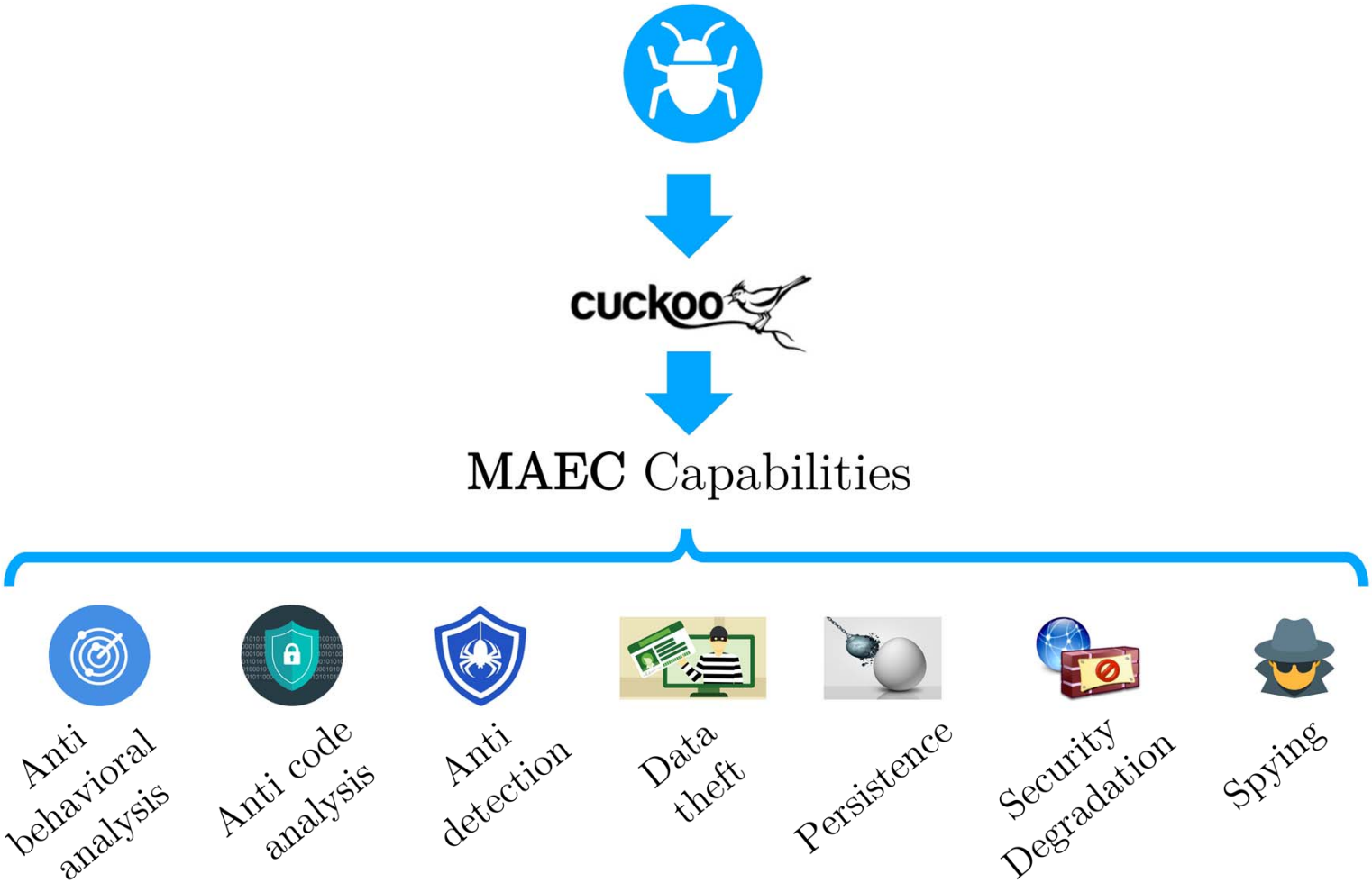
		Instruction Histogram				
		i_1	i_2	$\ldots$	i_{52}	i_{53}
Visited Nodes	n_0	$n_0 i_1$	$n_0 i_2$	...	$n_0 i_{52}$	$n_0 i_{53}$
	n_2	$n_2 i_1$	$n_2 i_2$	$\ldots$	$n_2 i_{52}$	$n_2 i_{53}$
	n_4	$n_4 i_1$	$n_4 i_2$	$\ldots$	$n_4 i_{52}$	$n_4 i_{53}$
	n_5	$n_5 i_1$	$n_5 i_2$	$\ldots$	$n_5 i_{52}$	$n_5 i_{53}$
	n_9	$n_9 i_1$	$n_9 i_2$	$\ldots$	$n_9 i_{52}$	$n_9 i_{53}$

Dynamic Characterization

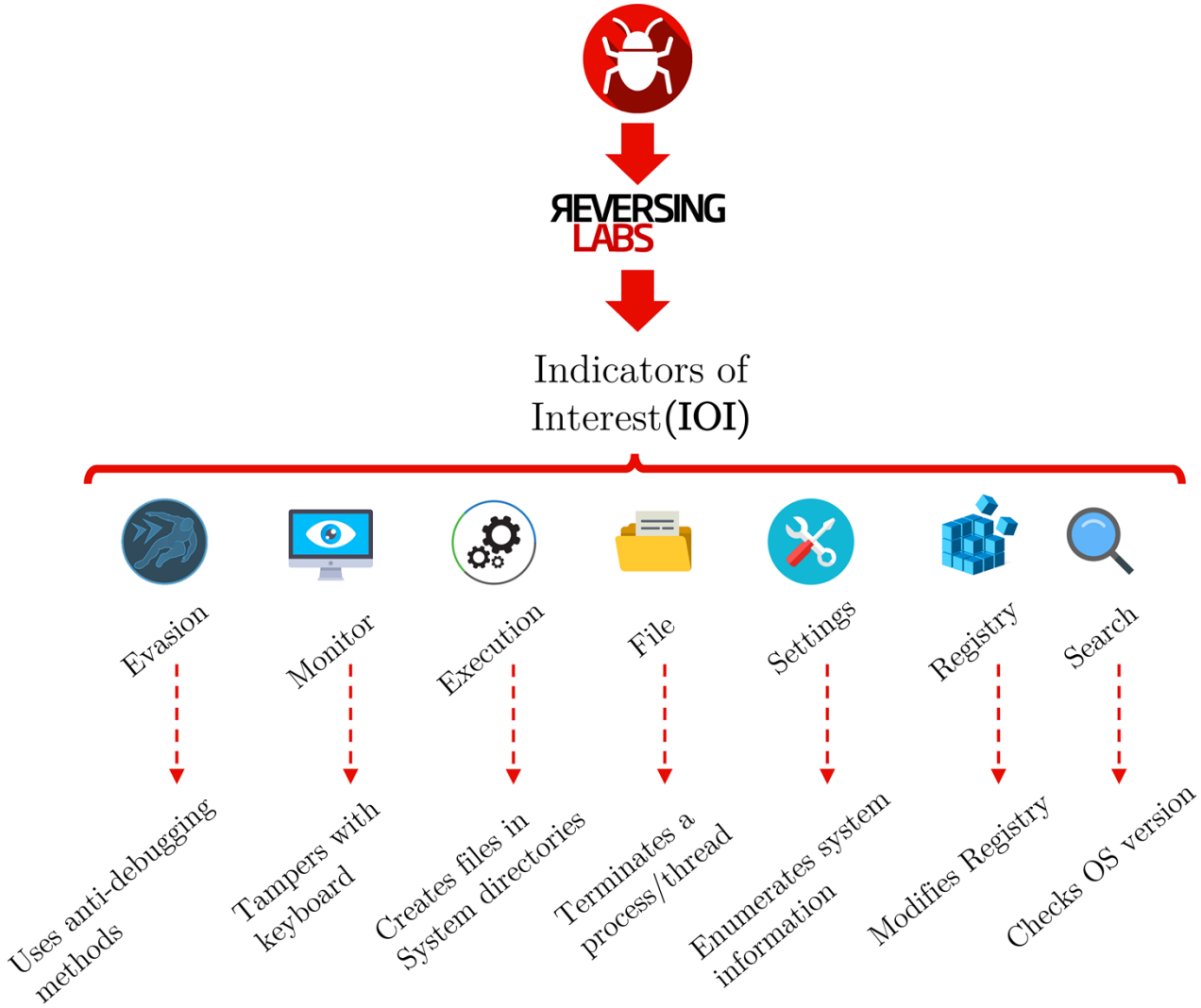


Malware	Cap_1	Cap_2	...	Cap_{n-1}	Cap_n
56ed8788fo...b5ba8c8r28	1	0	...	0	1

Dynamic Characterization (cont'd)



Dynamic Characterization (cont'd)

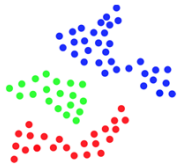


Malware Dataset

Name	Type	Count	
		Cuckoo/MAEC	A1000
Andromeda	virus	3516	4458
Banker	spyware	2730	8456
Banload	downloader	4358	8582
Cutwail	downloader	3018	4073
Inject	trojan	3181	8193
Injector	trojan	2498	8522
Ramnit	trojan	3525	7169
Shifu	spyware	3497	4629
Zbot	downloader	3751	5545

Decision Trees

Highly **interpretable** and **easy** to visualize

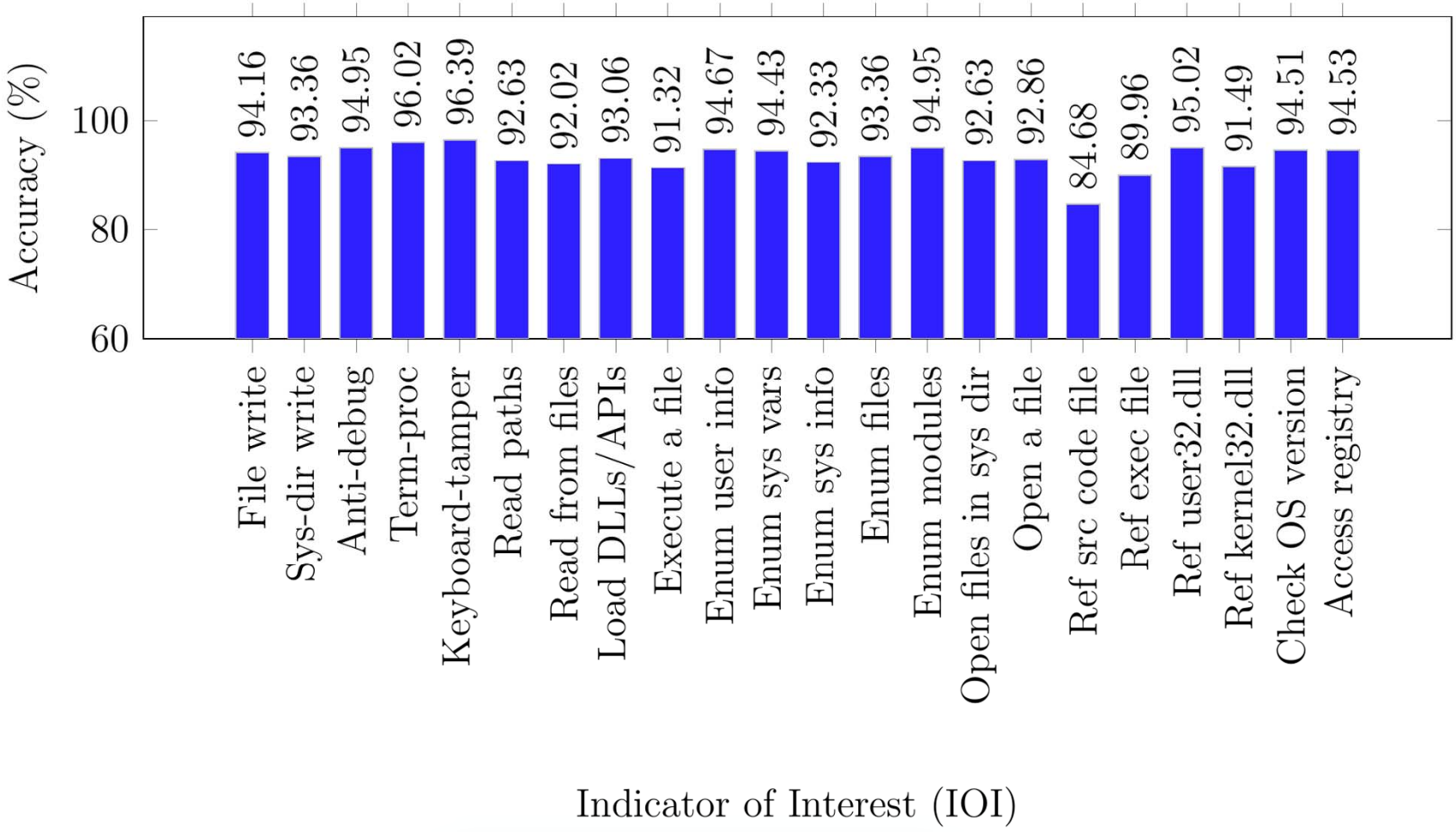


Inherently perform **feature importance** and **selection**

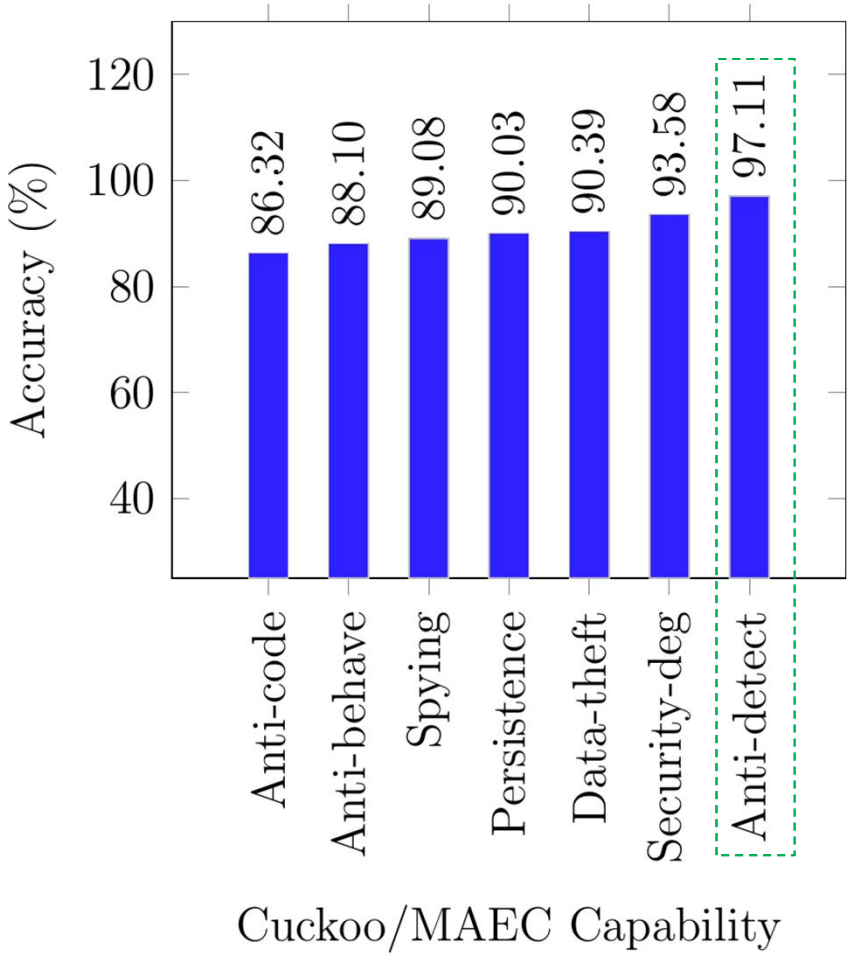
Low computational cost

Identify **instructions** that lead to specific **capabilities**

Results


Accuracy (%)
100
80
60
94.16
93.36
94.95
96.02
96.39
92.63
92.02
93.06
91.32
94.67
94.43
92.33
93.36
94.95
92.63
92.86
84.68
89.96
95.02
91.49
94.51
94.53
File write
Sys-dir write
Anti-debug
Term-proc
Keyboard-tamper
Read paths
Read from files
Load DLLs/APIs
Execute a file
Enum user info
Enum sys vars
Enum sys info
Enum files
Enum modules
Open files in sys dir
Open a file
Ref src code file
Ref exec file
Ref user32.dll
Ref kernel32.dll
Check OS version
Access registry
Indicator of Interest (IOI)

Results

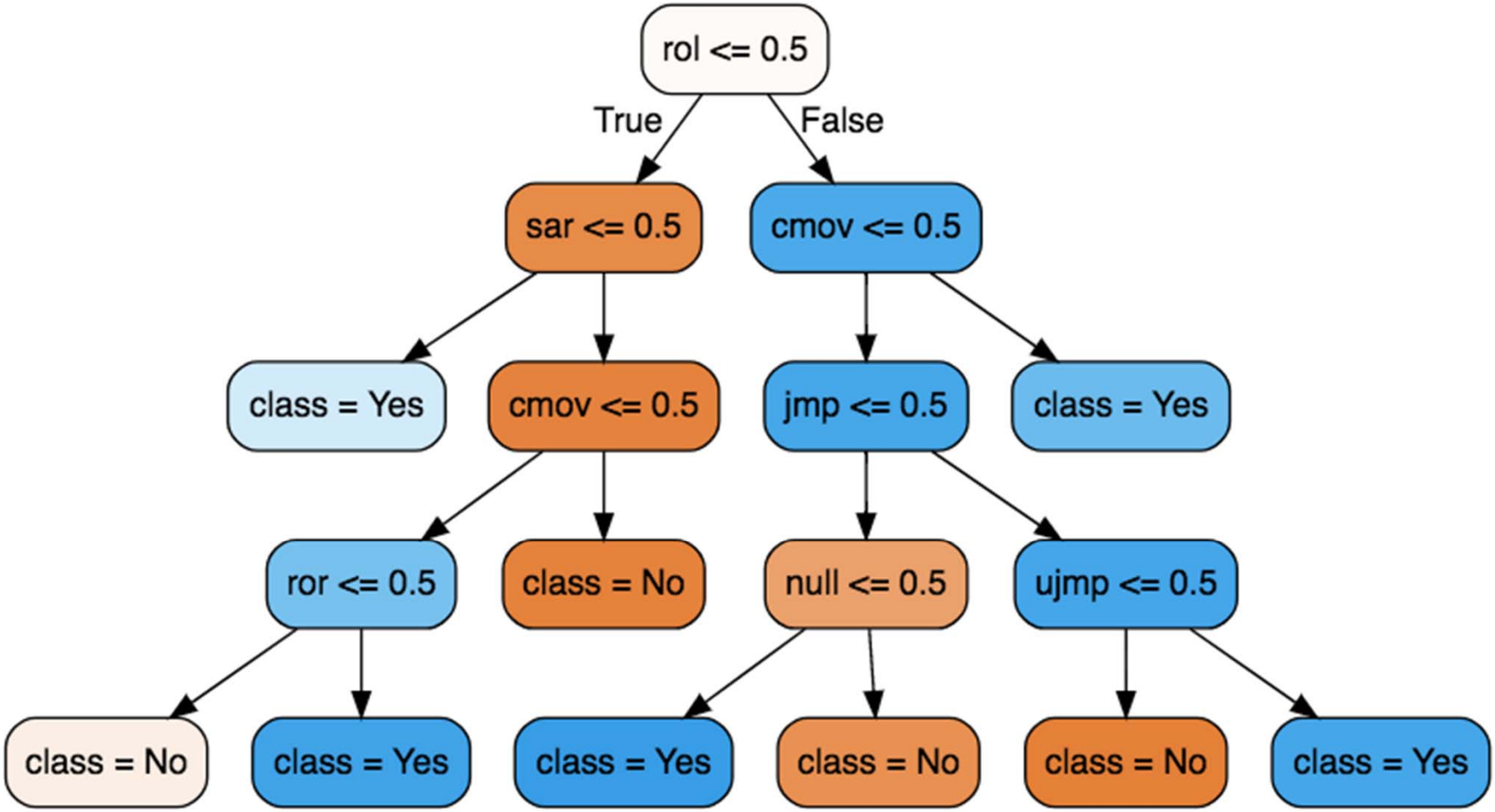


Fast static analysis to predict important capabilities

Identify **anti-sandbox** malware

For **bare-metal** systems

Spot **anti-detection** malware

Requires **more expensive** characterizations

Decision Tree Model