Cloud Malware

Domhnall Carlin, Dr. P. O'Kane & Prof. S. Sezer

Centre for Secure Information Technologies, Queen's University, Belfast, N. Ireland.

Research Objective

Develop a strategy for the detection of malware, which is immune to modern obfuscation methods, and which is applicable at the hypervisor level.

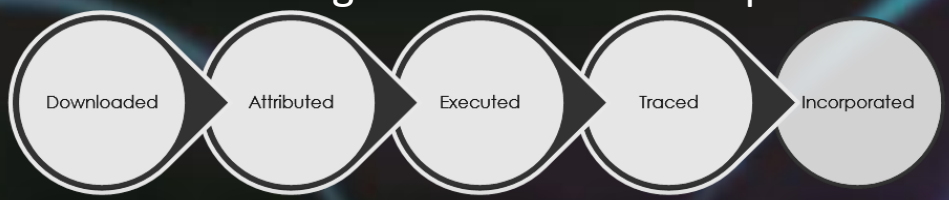
Signature-detection is the most widely used approach in commercial malware detection[1].
New malware must be captured & analysed for a signature, which is deployed to users.
Obfuscation techniques compound the issue- 1000s of variations generated per day.
By definition, we are constantly behind the curve in the arms race against malware.

- Static analysis of opcodes (assembly instructions) can detect malware.
- Bypasses issues inherent in signature-based methods (i.e. the instance can be novel).
- However, packed or encrypted malware cannot generally be investigated.

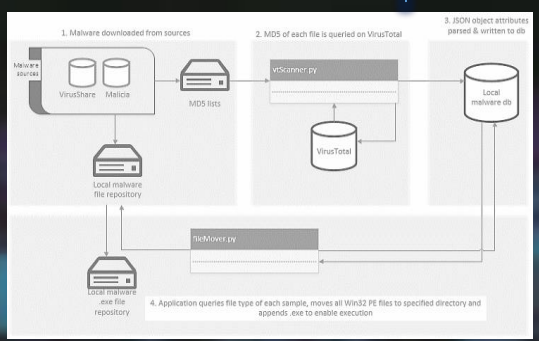
- Dynamic analysis of opcodes allows malware to reveal itself (decryption, unpacking etc).
- This allows the detection and analysis of obfuscated malware.
- Datasets in the literature are typically small and poorly sampled.

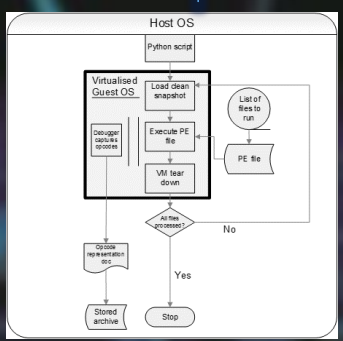
Work in progress- creation of a large dynamically-generated runtrace dataset

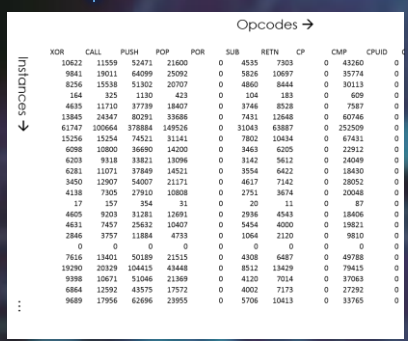
Processing each malicious sample



VirusShare_01bc9247fd095cec46507 0441150aaa1	Trojan Win32 PE	01bc9247fd095ce c465070441150aa a1.exe	PUSH XOR ADC CLI MOV MOV PUSH SUB POP JE BT RETN	PUSH SUB RETN 343004748 49333 566749398 733







XOR	CALL	PUSH	POP	POR	SUB	RETN	CP	CMP	CPUID
10622	11559	52471	21600	0	4535	7303	0	43260	0
9841	19011	64099	25092	0	5826	10697	0	35774	0
8256	15538	51302	20707	0	4860	8444	0	30113	0
164	325	1130	423	0	104	183	0	609	0
4635	11710	37739	18407	0	3746	8528	0	7587	0
13845	24347	80291	33686	0	7431	12648	0	60746	0
61747	100664	378884	149526	0	31043	63887	0	252509	0
15256	15254	74521	31141	0	7802	10434	0	67431	0
6098	10800	36690	14200	0	3463	6205	0	22912	0
6203	9318	33821	13096	0	3142	5612	0	24049	0
6281	11071	37849	14521	0	3954	6422	0	18430	0
3450	12907	54007	21171	0	4617	7142	0	28052	0
4138	7305	27910	10808	0	2751	3674	0	20048	0
17	157	354	31	0	20	11	0	87	0
4605	9203	31281	12691	0	2936	4543	0	18406	0
4631	7457	25632	10407	0	5454	4000	0	19821	0
2846	3757	11884	4733	0	1064	2120	0	9810	0
0	0	0	0	0	0	0	0	0	0
7616	13401	50189	21515	0	4308	6487	0	49788	0
19290	20329	104415	43448	0	8512	13429	0	79415	0
9398	10671	51046	21369	0	4120	7014	0	37063	0
6864	12592	43575	17572	0	4002	7173	0	27292	0
9689	17956	62696	23955	0	5706	10413	0	33765	0

Aim is to create dataset of sufficient depth & breadth to allow significant sub-analyses, while controlling for class imbalances, malicious file percentage and sample size, and share within research community.

Future work facilitated by new dataset:

One-class learning	Novel classification	Algorithm comparison	Deep learning
Novel feature reduction	Novel Features	Uninvestigated malware	Graph-based approach