SCIENCE CHINA
Information Sciences

• Supplementary File •

Honey Insight: The Interpretable Reasoning Engine for Dynamic TTP Inference

Yinghai Zhou[1,2,3 †] , Ziyu Wang[1 †] , Rui Wang[1], Yuyu He[1 *], Hui Lu[1,2,3] & Zhihong Tian[1,2,3 *]

[1]*Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou 510006, China*
[2]*Guangdong Key Laboratory of Industrial Control System Security, Guangzhou 510006, China*
[3]*Huangpu Research School of Guangzhou University, Guangzhou 510006, China*

Appendix A

Table A1 Dataset System Behavior Atom Categories and Frequency Statistics

Category	System Behavior Atom	Total Frequency	Avg. Freq./Sample	Operation Semantic Example
File Operations	files_opened(b_1)	1284128	159.54	`cat /etc/passwd`
	files_written(b_2)	256694	36.55	`echo "mal" > /tmp/mal.txt`
	files_deleted (b_3)	167230	29.95	`rm /tmp/mal.txt`
	files_attribute_changed(b_4)	44021	19.51	`chmod 777 /tmp/mal.sh`
	files_copied(b_5)	24934	10.51	`cp /tmp/mal.sh /usr/bin/`
	files_dropped(b_6)	171994	29.14	`curl http://mal.com/shell.php`
Process Operations	processes_terminated(b_7)	60962	7.60	`kill 1234`
	processes_killed(b_8)	461	1.98	`kill -9 1234`
	processes_injected(b_9)	4136	2.09	Memory backdoor
Service Operations	services_opened(b_{10})	14714	5.43	`ssh user@ip`
	services_created(b_{11})	1179	2.45	`sc create MalService`
	services_started(b_{12})	1542	1.69	`systemctl start mal`
	services_stopped(b_{13})	97	1.26	`systemctl stop mal`
	services_deleted(b_{14})	49	1.32	`sc delete MalService`
	services_bound(b_{15})	84	2.80	`nc -l 8080`
Window Operations	windows_searched(b_{16})	9118	6.56	`tasklist`
	windows_hidden(b_{17})	4341	1.39	`powershell -w hidden`
Mutex Operations	mutexes_opened(b_{18})	36427	7.10	Open system mutex
	mutexes_created(b_{19})	126635	19.18	Create malicious mutex
Registry Operations	registry_keys_opened(b_{20})	1976118	235.22	`reg query HKLM/Software/Mal`
	registry_keys_set(b_{21})	599769	88.20	`reg add HKLM/Run /v Mal`
	registry_keys_deleted(b_{22})	65538	16.79	`reg delete HKLM/Software/Mal`
Network Communication	tls(b_{23})	3450	3.30	`https://mal.com`
	dns_lookups(b_{24})	47013	9.00	`nslookup mal.com`
	http_conversations(b_{25})	61721	18.64	`curl http://mal.com`
	ip_traffic(b_{26})	120178	19.67	TCP traffic to remote IP:Port

Appendix B

* Corresponding author (email: heyy@gzhu.edu.cn, tianzhihong@gzhu.edu.cn)
† These authors contributed equally to this work.

Table B1 Statistical Distribution of the Malware Dataset (N=10,126*)

(a) Malware family distribution

Family	**Count**	**Percentage (%)**
bedep	217	3.05
agentb	175	2.46
zusy	151	2.12
w97m	143	2.01
doina	123	1.73
graftor	104	1.46
msil	100	1.40
bulz	95	1.33
winlnk	91	1.28
ursu	83	1.17
Subtotal (Top-10)	**1282**	**18.01**
Other Families (>100 families)	5835	81.99
Total	**7117***	**100**

(b) Platform/OS distribution

Platform / OS	**Count**	**Percentage (%)**
Windows	8817	99.18
Linux	71	0.80
Unix	2	0.02
Total	**8890**	**100**

(c) Tactic distribution

Tactic ID	**Tactic Name**	**Frequency**	**Percentage (%)**
TA0007	Discovery	42413	32.13
TA0005	Defense Evasion	35930	27.22
TA0002	Execution	12860	9.74
TA0004	Privilege Escalation	12523	9.49
TA0003	Persistence	9407	7.13
TA0009	Collection	6250	4.73
TA0011	Command and Control	6217	4.71
TA0006	Credential Access	4227	3.20
TA0040	Impact	1072	0.81
TA0008	Lateral Movement	553	0.42
TA0001	Initial Access	456	0.35
TA0010	Exfiltration	42	0.03
TA0042	Resource Development	40	0.03
TA0043	Reconnaissance	25	0.02
Total		**132015**	**100**

Note: One Malware may involve multiple tactic stages. Additionally, not all malware samples have families (total dataset N=10126, but only 7117 have assigned families).*.