Friend or Foe: Discerning Benign vs Malicious Software and Malware Family

Aaron Walker*, Tapadhir Das*, Raj Mani Shukla† and Shamik Sengupta*
*Department of Computer Science and Engineering
University of Nevada, Reno, Reno, USA
Email: awalker@unr.edu, tapadhird@nevada.unr.edu, ssengupta@unr.edu
†Department of Computer Science
University of Bristol, Bristol, UK
Email: raj.shukla@bristol.ac.uk

***Abstract*—Malware remains one of the gravest threats to cybersecurity, second only to social engineering or a lack of user security awareness. As malware continues to evolve and frustrate legacy detection and prevention mechanisms, additional approaches are necessary to ensure security resilience. Machine learning offers many opportunities to better combat malware threats through the advantage of big datasets. Our research highlights how machine learning can be leveraged to identify malware threats with rapid results, enabling cybersecurity professionals to learn and adapt to these threats. The approach we present in this paper produces an efficient methodology to discern malware family and function through analysis of just the first 3,000 Windows system API function calls. We compare MLP, CNN, and SVM networks to determine the best performance in terms of accuracy and speed and find that MLP works the best with our dataset.**

***Index Terms*—Malware Detection, Malware Analysis, Malware Signature, Machine Learning**

I. Appendix

In our paper we show that it is possible to discern malware and benign software through learning the Windows system API function calls made by different classifications of software. The novelty of this approach can be found in how accurately it performs, given the sparse input of sequences of single API function call names. This produces a framework for quickly learning the differences between software to accurately predict not only if a given software is malicious or benign, but also to classify malicious software by family type. The accuracy of this approach is increased when including disparate software types and we believe that the overall, general accuracy will be increased by adding additional classes with a mix of benign and malicious software. This provides a quick tool for accurate malware analysis to provide greater insight and visibility into the nature of malware, one that may otherwise be unavailable for many cybersecurity professionals.

The following section illustrates the extensive results of our experimentation in the form of tables presenting Precision, Recall, and F1-score as a result of MLP networks trained on both malicious and benign software. Table I describes the software used in our experiments. Confusion matrices accompany the tables for each experiment in the next section.

TABLE I
Malware by Antivirus Signature Classification & Benign Software Set

ID	Signature	ID	Benign Executable
M1	Virus:VBS/Ramnit.gen!A	B1	7-Zip 32-bit
M2	Virus:VBS/Ramnit.gen!C	B2	7-Zip 64-bit
M3	PUA:Win32/Puamson.A!ml	B3	Avira Antivirus
M4	TrojanClicker:JS/Faceliker.M	B4	CCleaner
M5	Trojan:JS/Iframeinject	B5	Google Chrome
M6	Trojan:HTML/Redirector.CF	B6	Epson scanner software
M7	Trojan:Win32/Skeeyah.A!bit		
M8	Exploit:HTML/IframeRef.gen	B7	GifCam animated gif software
M9	Virus:VBS/Ramnit.B		
M10	TrojanClicker:JS/Faceliker.D	B8	GIMP image software
M11	TrojanClicker:JS/Faceliker.C		
M12	Trojan:JS/Redirector.QE	B9	OpenVPN
M13	Trojan:JS/BlacoleRef	B10	Ultrasurf proxy
M14	PUA:Win32/Presenoker	B11	Microsoft Visual Studio Code
M15	Trojan:HTML/Brocoiner.D!lib		
M16	TrojanClicker:JS/Faceliker!rfn		
M17	Trojan:Win32/Vibem.O		
M18	Trojan:HTML/Redirector.EP		
M19	Exploit:HTML/IframeRef		
M20	TrojanClicker:JS/Faceliker.A		
M21	Exploit:HTML/IframeRef.DM		
M22	Trojan:HTML/Phish		
M23	PUA:Win32/Kuaiba		

II. Tables

TABLE II
Statistics for Experiment 1.1

	Precision	Recall	F1-score	Support
B1	0.53	0.90	0.67	906
M12	0.59	0.22	0.32	880
M9	1.00	0.89	0.94	914
Accuracy			0.68	2700
Macro Avg.	0.71	0.67	0.64	2700
Weighted Avg.	0.71	0.68	0.65	2700

TABLE III
CONFUSION MATRIX FOR EXPERIMENT 1.1

Predicted Class	Actual Class		
	B1	M12	M9
B1	818	88	0
M12	684	196	0
M9	49	47	818

TABLE IV
STATISTICS FOR EXPERIMENT 1.2

	Precision	Recall	F1-score	Support
M5	0.50	0.41	0.45	914
B2	1.00	0.89	0.94	880
M2	0.51	0.65	0.57	906
Accuracy			0.65	2700
Macro Avg.	0.67	0.65	0.65	2700
Weighted Avg.	0.66	0.65	0.65	2700

TABLE V
CONFUSION MATRIX FOR EXPERIMENT 1.2

Predicted Class	Actual Class		
	M5	B2	M2
M5	374	0	540
B2	64	783	33
M2	317	0	589

TABLE VI
STATISTICS FOR EXPERIMENT 1.3

	Precision	Recall	F1-score	Support
M17	0.98	0.78	0.87	880
B3	0.83	0.86	0.84	914
M9	0.85	1.00	0.92	906
Accuracy			0.88	2700
Macro Avg.	0.89	0.88	0.88	2700
Weighted Avg.	0.89	0.88	0.88	2700

TABLE VII
CONFUSION MATRIX FOR EXPERIMENT 1.3

Predicted Class	Actual Class		
	B1	B3	M9
M17	687	155	38
B3	16	782	116
M9	0	0	906

TABLE VIII
STATISTICS FOR EXPERIMENT 2.1

	Precision	Recall	F1-score	Support
B4	0.53	0.4	0.46	897
M10	1.00	0.99	1.00	914
M12	0.51	0.63	0.56	889
Accuracy			0.68	2700
Macro Avg.	0.68	0.68	0.67	2700
Weighted Avg.	0.68	0.68	0.68	2700

TABLE IX
CONFUSION MATRIX FOR EXPERIMENT 2.1

Predicted Class	Actual Class		
	B4	M10	M12
B4	363	0	534
M10	1	906	7
M12	326	0	563

TABLE X
STATISTICS FOR EXPERIMENT 2.2

	Precision	Recall	F1-score	Support
B5	0.50	0.98	0.66	897
M11	1.00	0.91	0.95	914
M7	0.57	0.06	0.11	889
Accuracy			0.65	2700
Macro Avg.	0.69	0.65	0.57	2700
Weighted Avg.	0.69	0.65	0.58	2700

TABLE XI
CONFUSION MATRIX FOR EXPERIMENT 2.2

Predicted Class	Actual Class		
	B5	M11	M7
B5	881	2	14
M11	54	835	25
M7	837	0	52

TABLE XII
STATISTICS FOR EXPERIMENT 2.3

	Precision	Recall	F1-score	Support
B6	0.48	1.00	0.65	880
M22	0.96	0.43	0.60	896
M18	0.89	0.46	0.61	924
Accuracy			0.63	2700
Macro Avg.	0.78	0.63	0.62	2700
Weighted Avg.	0.78	0.63	0.62	2700

TABLE XIII
CONFUSION MATRIX FOR EXPERIMENT 2.3

Predicted Class	Actual Class		
	B6	M22	M18
B6	878	0	2
M22	458	389	49
M18	482	18	424

TABLE XIV
STATISTICS FOR EXPERIMENT 3.1

	Precision	Recall	F1-score	Support
B9	0.86	0.41	0.55	906
B7	0.62	0.97	0.75	880
B8	0.96	0.93	0.95	914
Accuracy			0.77	2700
Macro Avg.	0.81	0.77	0.75	2700
Weighted Avg.	0.82	0.77	0.75	2700

TABLE XV
Confusion Matrix for Experiment 3.1

Predicted Class	Actual Class		
	B9	B7	B8
B9	367	509	30
B7	20	856	4
B8	38	23	853

TABLE XVI
Statistics for Experiment 3.2

	Precision	Recall	F1-score	Support
B10	0.89	0.71	0.79	840
B4	0.67	0.70	0.68	891
B11	0.56	0.64	0.60	916
Accuracy			0.68	2700
Macro Avg.	0.71	0.68	0.69	2700
Weighted Avg.	0.70	0.68	0.69	2700

TABLE XVII
Confusion Matrix for Experiment 3.2

Predicted Class	Actual Class		
	B10	B4	B11
B10	593	19	228
B4	36	623	232
B11	39	287	590

TABLE XVIII
Statistics for Experiment 3.3

	Precision	Recall	F1-score	Support
B1	0.50	0.26	0.35	914
B2	0.47	0.98	0.63	906
B3	0.49	0.17	0.25	880
Accuracy			0.47	2700
Macro Avg.	0.48	0.47	0.41	2700
Weighted Avg.	0.48	0.47	0.41	2700

TABLE XIX
Confusion Matrix for Experiment 3.3

Predicted Class	Actual Class		
	B1	B2	B3
B1	241	517	156
B2	14	890	2
B3	228	503	149

TABLE XX
Statistics for Experiment 4.1

	Precision	Recall	F1-score	Support
B8	0.87	0.57	0.69	914
B7	0.70	0.89	0.79	880
M12	0.93	1.00	0.96	906
Accuracy			0.82	2700
Macro Avg.	0.83	0.82	0.81	2700
Weighted Avg.	0.83	0.82	0.81	2700

TABLE XXI
Confusion Matrix for Experiment 4.1

Predicted Class	Actual Class		
	B8	B7	M12
B1	519	336	59
B7	80	786	14
M12	0	0	906

TABLE XXII
Statistics for Experiment 4.2

	Precision	Recall	F1-score	Support
B10	0.98	0.21	0.34	916
M9	0.61	0.49	0.54	839
B11	0.50	0.99	0.66	892
Accuracy			0.56	2647
Macro Avg.	0.69	0.56	0.52	2647
Weighted Avg.	0.70	0.56	0.51	2647

TABLE XXIII
Confusion Matrix for Experiment 4.2

Predicted Class	Actual Class		
	B10	M12	B11
B10	190	258	468
M9	3	410	426
B11	1	7	884

TABLE XXIV
Statistics for Experiment 4.3

	Precision	Recall	F1-score	Support
B1	0.50	0.81	0.62	914
M12	0.82	1.00	0.90	906
M9	0.38	0.05	0.10	880
Accuracy			0.63	2700
Macro Avg.	0.57	0.62	0.54	2700
Weighted Avg.	0.57	0.63	0.54	2700

TABLE XXV
Confusion Matrix for Experiment 4.3

Predicted Class	Actual Class		
	B1	M12	M9
B1	740	96	78
M12	0	906	0
M9	735	97	48

TABLE XXVI
Statistics for Experiment 5.1

	Precision	Recall	F1-score	Support
M9	0.40	0.74	0.52	913
M12	0.35	0.20	0.25	882
M3	0.44	0.26	0.33	905
Accuracy			0.40	2700
Macro Avg.	0.40	0.40	0.37	2700
Weighted Avg.	0.40	0.40	0.37	2700

TABLE XXVII
Confusion Matrix for Experiment 5.1

Predicted Class	Actual Class		
	M9	M12	M3
M9	672	129	112
M12	521	175	186
M3	469	199	237

TABLE XXVIII
Statistics for Experiment 5.2

	Precision	Recall	F1-score	Support
M5	0.60	0.19	0.29	913
M23	1.00	0.99	1.00	882
M2	0.51	0.87	0.65	905
Accuracy			0.68	2700
Macro Avg.	0.71	0.69	0.64	2700
Weighted Avg.	0.70	0.68	0.64	2700

TABLE XXIX
Confusion Matrix for Experiment 5.2

Predicted Class	Actual Class		
	M5	M23	M2
M5	174	231	0
M23	437	444	1
M2	5	94	806

TABLE XXX
Statistics for Experiment 5.3

	Precision	Recall	F1-score	Support
M9	0.61	0.75	0.67	913
M14	0.58	0.50	0.54	882
M17	1.00	0.89	0.94	905
Accuracy			0.72	2700
Macro Avg.	0.73	0.71	0.72	2700
Weighted Avg.	0.73	0.72	0.72	2700

TABLE XXXI
Confusion Matrix for Experiment 5.3

Predicted Class	Actual Class		
	M9	M14	M17
M9	682	231	0
M14	437	444	1
M17	5	94	806

TABLE XXXII
Statistics for Experiment 6.1

	Precision	Recall	F1-score	Support
M6	0.36	0.52	0.42	913
M10	0.37	0.40	0.39	882
M12	0.33	0.15	0.21	905
Accuracy			0.36	2700
Macro Avg.	0.35	0.36	0.34	2700
Weighted Avg.	0.35	0.36	0.34	2700

TABLE XXXIII
Confusion Matrix for Experiment 6.1

Predicted Class	Actual Class		
	M6	M10	M12
M6	473	304	136
M10	386	353	143
M12	472	293	140

TABLE XXXIV
Statistics for Experiment 6.2

	Precision	Recall	F1-score	Support
M7	0.38	0.03	0.06	913
M11	0.41	0.50	0.45	882
M15	0.37	0.63	0.47	905
Accuracy			0.39	2700
Macro Avg.	0.39	0.39	0.33	2700
Weighted Avg.	0.39	0.39	0.32	2700

TABLE XXXV
Confusion Matrix for Experiment 6.2

Predicted Class	Actual Class		
	M7	M11	M15
M7	30	311	572
M11	37	441	404
M15	13	319	573

TABLE XXXVI
Statistics for Experiment 6.3

	Precision	Recall	F1-score	Support
M22	0.36	0.43	0.40	913
M13	0.33	0.61	0.43	882
M18	0.00	0.00	0.00	905
Accuracy			0.34	2700
Macro Avg.	0.23	0.35	0.27	2700
Weighted Avg.	0.23	0.34	0.27	2700

TABLE XXXVII
Confusion Matrix for Experiment 6.3

Predicted Class	Actual Class		
	M22	M13	M18
M22	394	519	0
M13	348	534	0
M18	338	567	0

TABLE XXXVIII
Statistics for Experiment 7.1

	Precision	Recall	F1-score	Support
M1	1.00	0.96	0.98	913
M9	0.55	0.74	0.63	882
M2	0.61	0.42	0.50	905
Accuracy			0.71	2700
Macro Avg.	0.72	0.71	0.70	2700
Weighted Avg.	0.72	0.71	0.70	2700

TABLE XXXIX
CONFUSION MATRIX FOR EXPERIMENT 7.1

Predicted Class	Actual Class		
	M1	M9	M2
M1	876	18	19
M9	0	655	227
M2	0	521	384

TABLE XL
STATISTICS FOR EXPERIMENT 7.2

	Precision	Recall	F1-score	Support
M4	0.34	0.29	0.31	913
M16	0.32	0.42	0.36	882
M20	0.32	0.27	0.29	905
Accuracy			0.32	2700
Macro Avg.	0.33	0.32	0.32	2700
Weighted Avg.	0.33	0.32	0.32	2700

TABLE XLI
CONFUSION MATRIX FOR EXPERIMENT 7.2

Predicted Class	Actual Class		
	M4	M16	M20
M4	262	397	254
M16	247	371	264
M20	256	408	241

TABLE XLII
STATISTICS FOR EXPERIMENT 7.3

	Precision	Recall	F1-score	Support
M21	0.34	0.98	0.51	913
M8	0.35	0.03	0.06	882
M19	0.00	0.00	0.00	905
Accuracy			0.34	2700
Macro Avg.	0.23	0.34	0.19	2700
Weighted Avg.	0.23	0.34	0.19	2700

TABLE XLIII
CONFUSION MATRIX FOR EXPERIMENT 7.3

Predicted Class	Actual Class		
	M21	M8	M19
M21	894	19	0
M8	853	29	0
M19	871	34	0