

STARCRAFT																Average Recall Across Attack				
WB Random Attack	25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense	0.93244	0.06756	0.03569	0.96431		0.88548	0.11452	0.06809	0.93191		0.76541	0.23459	0.15941	0.84059			0.25171	0.74829		87.13
Binary LSTM	0.93412	0.06588	0.04158	0.95842		0.90832	0.09168	0.05840	0.94160		0.74719	0.25281	0.18799	0.81201			0.28558	0.71442		85.66
Random Forest	0.92226	0.07774	0.04568	0.95432		0.89596	0.10404	0.09459	0.90541		0.80094	0.19906	0.20242	0.79758			0.29765	0.70235		83.99
SVM Classifier	0.77701	0.22299	0.12922	0.87078		0.75519	0.24481	0.17248	0.82752		0.66506	0.33494	0.23345	0.76655			0.16642	0.83358		82.46
Knearest Neighbors	0.93350	0.06650	0.05506	0.94494		0.89964	0.10036	0.09782	0.90218		0.79623	0.20377	0.20527	0.79473			0.29842	0.70158		83.59
Ensemble			0.02240	0.97760	0.44030			0.03693	0.96307	0.64646			0.07384	0.92616	0.82122		0.09940	0.90060	1.00000	94.19
WB Strategic-Timed Attack	25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense	0.94614	0.05386	0.00846	0.99154		0.99921	0.00079	0.00051	0.99949		0.90966	0.09034	0.03989	0.96011			0.00153	0.99847		98.74
Binary LSTM	0.91125	0.08875	0.02397	0.97603		0.99935	0.00065	0.00033	0.99967		0.83655	0.16345	0.02241	0.97759			0.00177	0.99823		98.79
Random Forest																				100.00
SVM Classifier																				100.00
Knearest Neighbors																				100.00
Ensemble			0.01003	0.98997	0.51396			0.00003	0.99997	0.73236			0.00217	0.99783	0.91823		0.00016	0.99984	1.00000	99.69
Bbox prediction	25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense	0.92710	0.07290	0.05493	0.94507		0.98790	0.01210	0.02443	0.97557		0.98714	0.01286	0.02944	0.97056			0.06882	0.93118		95.56
Binary LSTM	0.89614	0.10386	0.03706	0.96294		0.98241	0.01759	0.01657	0.98343		0.98434	0.01566	0.02908	0.97092			0.07303	0.92697		96.11
Random Forest																				100.00
SVM Classifier																				100.00
Knearest Neighbors																				100.00
Ensemble			0.01743	0.98257	0.43224			0.00608	0.99392	0.65921			0.008450	0.99154960	0.7817056		0.01609	0.98391	1.00000	98.80
Bbox random	25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense	0.98239	0.01761	0.03583	0.96417		0.99028	0.00972	0.02270	0.97730		0.98611	0.01389	0.02343	0.97657			0.02515	0.97485		97.32
Binary LSTM	0.97485	0.02515	0.02777	0.97223		0.98728	0.01272	0.01698	0.98302		0.98629	0.01371	0.01865	0.98135			0.03322	0.96678		97.58
Random Forest																				100.00
SVM Classifier																				100.00
Knearest Neighbors																				100.00
Ensemble			0.00946	0.99054	0.44304			0.00492	0.99508	0.56421			0.00312	0.99688	0.69092		0.00128	0.99872	1.00000	99.53
Bbox timed	25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense	0.92600	0.07400	0.00200	0.99800		0.99597	0.00403	0.00232	0.99768		0.98710	0.01290	0.01632	0.98368			0.07356	0.92644		97.65
Binary LSTM	0.93372	0.06628	0.00492	0.99508		0.99550	0.00450	0.00180	0.99820		0.98671	0.01329	0.01550	0.98450			0.07302	0.92698		97.62
Random Forest																				100.00
SVM Classifier																				100.00
Knearest Neighbors																				100.00
Ensemble			0.00061	0.99939	0.49309			0.00025	0.99975	0.69234			0.00764	0.99236	0.70672		0.02179	0.97821	1.00000	99.24
Whitebox Predict	25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense	0.87157	0.12843	0.02642	0.97358		0.97537	0.02463	0.01311	0.98689		0.98005	0.01995	0.02364	0.97636			0.07429	0.92571		96.56
Binary LSTM	0.89741	0.10259	0.04701	0.95299		0.98336	0.01664	0.02171	0.97829		0.98330	0.01670	0.03535	0.96465			0.06730	0.93270		95.72
Random Forest																				100.00
SVM Classifier																				100.00
Knearest Neighbors																				100.00
Ensemble			0.02423	0.97577	0.43523			0.00938	0.99062	0.66130			0.00989	0.99011	0.78173		0.01575	0.98425	0.86484	98.52

JOE'S ENVIRONMENT																					
WB Random Attack		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model		TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense		0.93753	0.06247	0.06068	0.93932		0.97777	0.02223	0.02342	0.97658		0.98712	0.01288	0.01632	0.98368				0.02257	0.97743	
Binary LSTM		0.92522	0.07478	0.06832	0.93168		0.97623	0.02377	0.02463	0.97537		0.98499	0.01501	0.01598	0.98402				0.03443	0.96557	
Random Forest		0.88441	0.11559	0.10042	0.89958		0.96453	0.03547	0.03841	0.96159		0.98437	0.01563	0.02390	0.97610				0.02177	0.97823	
SVM Classifier		0.70937	0.29063	0.30437	0.69563		0.84881	0.15119	0.17641	0.82359		0.91133	0.08867	0.10609	0.89391				0.08828	0.91172	
Knearest Neighbors		0.68448	0.31552	0.32277	0.67723		0.79102	0.20898	0.22997	0.77003		0.83887	0.16113	0.20936	0.79064				0.23673	0.76327	
Ensemble				0.04221	0.95779	0.60958			0.01455	0.98531	0.85804			0.00957	0.99043	0.94597			0.02025	0.97975	1.00000
WB Strategic-Timed Attack		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model		TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense		0.98126	0.01874	0.03587	0.96413		0.97918	0.02082	0.03870	0.96130		0.96831	0.03169	0.03505	0.96495				0.03445	0.96555	
Binary LSTM		0.97612	0.02388	0.02871	0.97129		0.97438	0.02562	0.03155	0.96845		0.96433	0.03567	0.02668	0.97332				0.01785	0.98215	
Random Forest		0.98397	0.01603	0.08568	0.91432		0.97781	0.02219	0.08834	0.91166		0.95148	0.04852	0.06492	0.93508				0.03575	0.96425	
SVM Classifier		0.96503	0.03497	0.23447	0.76553		0.94904	0.05096	0.20922	0.79078		0.89941	0.10059	0.15657	0.84343				0.09038	0.90962	
Knearest Neighbors		0.90830	0.09170	0.20169	0.79831							0.84421	0.15579	0.17602	0.82398				0.13229	0.86771	
Ensemble				0.01693	0.98307	0.63904			0.01820	0.98180	0.85570			0.01544	0.95370	0.99097			0.01021	0.98979	1.00000
Bbox prediction		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model		TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense		0.82195	0.17805	0.06191	0.93809		0.90504	0.09496	0.07138	0.92862		0.95411	0.04589	0.06613	0.93387				0.05643	0.94357	
Binary LSTM		0.84537	0.15463	0.11212	0.88788		0.90858	0.09142	0.10712	0.89288		0.95154	0.04846	0.08359	0.91641				0.06099	0.93901	
Random Forest																					
SVM Classifier																					
Knearest Neighbors																					
Ensemble				0.07799	0.92201	0.55288			0.07666	0.92334	0.82688			0.06083	0.93917	0.95071			0.04381	0.95619	1.00000
Bbox random		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model		TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense		0.93287	0.06713	0.03291	0.96709		0.96343	0.03657	0.01947	0.98053		0.97198	0.02802	0.01589	0.98411				0.02592	0.97408	
Binary LSTM		0.94200	0.05800	0.04782	0.95218		0.96902	0.03098	0.02573	0.97427		0.97869	0.02131	0.01972	0.98028				0.02347	0.97653	
Random Forest																					
SVM Classifier																					
Knearest Neighbors																					
Ensemble				0.02861	0.97139	0.61871			0.01569	0.98431	0.84578			0.01182	0.98818	0.94094			0.01348	0.98652	1.00000
Bbox timed		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model		TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense		0.98573	0.01427	0.01755	0.98245		0.98182	0.01818	0.03169	0.96831		0.97475	0.02525	0.06752	0.93248				0.10125	0.89875	
Binary LSTM		0.98847	0.01153	0.01564	0.98436		0.98593	0.01407	0.03078	0.96922		0.97933	0.02067	0.05971	0.94029				0.08025	0.91970	
Random Forest																					
SVM Classifier																					
Knearest Neighbors																					
Ensemble				0.00866	0.99134	0.62186			0.01642	0.98358	0.83620			0.03314	0.96686	0.94423			0.04472	0.95528	1.00000
Whitebox Predict		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model		TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision
Binary Dense		0.82278	0.17722	0.06123	0.93877		0.90563	0.09437	0.07155	0.92845		0.95442	0.04558	0.06678	0.93322				0.05620	0.94380	
Binary LSTM		0.83763	0.16237	0.10879	0.89121		0.90158	0.09842	0.10536	0.89464		0.94686	0.05314	0.08263	0.91737				0.06103	0.93897	
Random Forest																					
SVM Classifier																					
Knearest Neighbors																					
Ensemble				0.07516	0.92484	0.5476312			0.07567	0.92433	0.82296			0.05992	0.94008	0.94956			0.04338	0.95662	1.00000

96.92

96.42

95.39

83.12

75.03

97.84

96.40

97.38

93.13

82.73

87.25

98.48

93.60

90.90

100.00

100.00

100.00

93.52

97.64

97.08

100.00

100.00

100.00

98.26

94.55

95.34

100.00

100.00

100.00

97.43

93.61

91.05

100.00

100.00

100.00

93.65

Frank's Env		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
WB Random	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	
Binary Dense	0.92493	0.07507	0.06226	0.93774		0.95005	0.04995	0.04377	0.95623		0.94257	0.05743	0.04292	0.95708			0.05456	0.94544		94.91	
Binary LSTM	0.91917	0.08083	0.06729	0.93271		0.94892	0.05108	0.04263	0.95737		0.94769	0.05231	0.04769	0.95231			0.06785	0.93215		94.36	
Random Forest	0.90606	0.09394	0.05950	0.94050		0.93787	0.06213	0.04790	0.95210		0.93579	0.06421	0.05015	0.94985			0.06053	0.93947		94.55	
SVM Classifier	0.81914	0.18086	0.33701	0.66299		0.86302	0.13698	0.26987	0.73013		0.86639	0.13361	0.20334	0.79666			0.17617	0.82434		75.34	
Knearest Neighbors	0.84657	0.15343	0.18292	0.81708		0.89349	0.10651	0.26987	0.85493		0.88990	0.11010	0.14862	0.85138			0.17324	0.82676		80.63	
Ensemble			0.04038	0.95962	0.51228			0.02558	0.97442	0.77540			0.02902	0.97098	0.90898		0.04178	0.95822	1.00000	96.58	
WB Timed		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	
Binary Dense	0.93966	0.06034	0.03004	0.96996		0.93804	0.06196	0.02903	0.97097		0.85792	0.14208	0.05488	0.94512			0.07571	0.92429		95.26	
Binary LSTM	0.96536	0.03464	0.04611	0.95389		0.95145	0.04855	0.06701	0.93299		0.90132	0.09868	0.08341	0.91659			0.10793	0.89207		92.39	
Random Forest																				100.00	
SVM Classifier																				100.00	
Knearest Neighbors																				100.00	
Ensemble			0.02680	0.97320	0.54243			0.03861	0.96139	0.75738			0.04431	0.95569	0.88349		0.05419	0.94581	1.00000	95.90	
BBox Predict		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	
Binary Dense	0.84348	0.15652	0.05408	0.94592		0.92105	0.07895	0.11327	0.88673		0.96162	0.03838	0.20119	0.79881			0.26626	0.73374		84.13	
Binary LSTM	0.79785	0.20215	0.04617	0.95383		0.89081	0.10919	0.09435	0.90565		0.94524	0.05476	0.18714	0.81286			0.25845	0.74155		85.35	
Random Forest																				100.00	
SVM Classifier																				100.00	
Knearest Neighbors																				100.00	
Ensemble		0.20294	0.03101	0.96899	0.40212		0.21762	0.06225	0.93775	0.72108		0.22916	0.12005	0.87995	0.91593		0.15950	0.84050	1.00000	90.68	
BBox Random		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	
Binary Dense	0.94514	0.05486	0.04245	0.95755		0.96385	0.03615	0.02255	0.97745		0.95824	0.04176	0.02682	0.97318			0.04170	0.95830		96.66	
Binary LSTM	0.94742	0.05258	0.05410	0.94590		0.96745	0.03255	0.03283	0.96717		0.96674	0.03326	0.03798	0.96202			0.05720	0.94280		95.45	
Random Forest																				100.00	
SVM Classifier																				100.00	
Knearest Neighbors																				100.00	
Ensemble		0.24179	0.03166	0.96834	0.52107		0.23821	0.01874	0.98126	0.78381		0.24601	0.02150	0.97850	0.91354		0.03297	0.96703	1.00000	97.38	
BBox Timed		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	
Binary Dense	0.97990	0.02010	0.02065	0.97935		0.97309	0.02691	0.04889	0.95111		0.96310	0.03690	0.07899	0.92101			0.13270	0.86730		92.97	
Binary LSTM	0.98278	0.01722	0.02107	0.97893		0.97868	0.02132	0.04779	0.95221		0.97153	0.02847	0.08025	0.91975			0.11805	0.88195		93.32	
Random Forest																				100.00	
SVM Classifier																				100.00	
Knearest Neighbors																				100.00	
Ensemble		0.26664	0.01267	0.98733	0.55185		0.27209	0.02744	0.97256	0.77567		0.27840	0.04236	0.95764	0.89833		0.05601	0.94399	1.00000	96.54	
Whitebox Predict		25% Adversarial					50% Adversarial NOT TRAINING SET					75% Adversarial					100% Adversarial				
Model	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	TN	FP	FN	TP/Recall	Precision	
Binary Dense	0.84865	0.15135	0.14802	0.85198		0.88556	0.11444	0.13160	0.86840		0.90658	0.09342	0.12246	0.87754			0.14354	0.85646		86.36	
Binary LSTM	0.80462	0.19538	0.13292	0.86708		0.84357	0.15643	0.11080	0.88920		0.86636	0.13364	0.09928	0.90072			0.11710	0.88290		88.50	
Random Forest																				100.00	
SVM Classifier																				100.00	
Knearest Neighbors																				100.00	
Ensemble		0.19198	0.09275	0.90725	0.45307		0.18820	0.07757	0.92243	0.74929		0.17277	0.06883	0.93117	0.90776		0.07749	0.92251	1.00000	92.08	

94.91
94.36
94.55
75.34
80.63
96.58

95.26
92.39
100.00
100.00
100.00
95.90

84.13
85.35
100.00
100.00
100.00
90.68

96.66
95.45
100.00
100.00
100.00
97.38

92.97
93.32
100.00
100.00
100.00
96.54

86.36
88.50
100.00
100.00
100.00
92.08

ENSEMBLE	<u>0.25</u>	<u>0.5</u>	<u>0.75</u>	<u>1</u>	ROUND1 FN		
STARCRAFT							
Whitebox Predict	0.02423	0.00938	0.00989	0.01575	2.42,0.94,0.99,1.57,		
Whitebox Random	0.02240	0.03693	0.07384	0.09940	2.24,3.69,7.38,9.94,	97.76,96.31,92.62,90.06	94.19
Whitebox Timed	0.01003	0.00003	0.00217	0.00016	1,0,0.22,0.02,	99,100,99.78,99.98	99.69
Blackbox Predict	0.01743	0.00608	0.00845	0.01609	1.74,0.61,0.85,1.61,	98.26,99.39,99.15,98.39	98.8
Blackbox Random	0.00946	0.00492	0.00312	0.00128	0.95,0.49,0.31,0.13,	99.05,99.51,99.69,99.87	99.53
Blackbox Timed	0.00061	0.00025	0.00764	0.02179	0.06,0.03,0.76,2.18,	99.94,99.97,99.24,97.82	99.24

ENSEMBLE MODEL	<u>0.25000</u>	<u>0.50000</u>	<u>0.75000</u>	<u>1.00000</u>			
Cooperative Navigation Attacks							
Whitebox Predict	0.07516	0.07567	0.05992	0.04338	7.52,7.57,5.99,4.34,	92.48,92.43,94.01,95.66	93.65
Whitebox Random	0.04221	0.01455	0.00957	0.02025	4.22,1.45,0.96,2.02,	95.78,98.55,99.04,97.98	97.84
Whitebox Timed	0.01693	0.01820	0.01544	0.01021	1.69,1.82,1.54,1.02,	98.31,98.18,98.46,98.98	98.48
Blackbox Predict	0.07799	0.07666	0.06083	0.04381	7.8,7.67,6.08,4.38,	92.2,92.33,93.92,95.62	93.52
Blackbox Random	0.02861	0.01569	0.01182	0.01348	2.86,1.57,1.18,1.35,	97.14,98.43,98.82,98.65	98.26
Blackbox Timed	0.00866	0.01642	0.03314	0.04472	0.87,1.64,3.31,4.47,	99.13,98.36,96.69,95.53	97.43

	<u>0.25000</u>	<u>0.50000</u>	<u>0.75000</u>	<u>1.00000</u>			
Physical Deception Attacks							
Whitebox Predict	0.09275	0.07757	0.06883	0.07749	9.27,7.76,6.88,7.75,	90.73,92.24,93.12,92.25	92.08
Whitebox Random	0.04038	0.02558	0.02902	0.04178	4.04,2.56,2.9,4.18,	95.96,97.44,97.1,95.82	96.58
Whitebox Timed	0.02680	0.03861	0.04431	0.05419	2.68,3.86,4.43,5.42,	97.32,96.14,95.57,94.58	95.9
Blackbox Predict	0.03101	0.06225	0.12005	0.15950	3.1,6.23,12,15.95,	96.9,93.77,88,84.05	90.68
Blackbox Random	0.03166	0.01874	0.02150	0.03297	3.17,1.87,2.15,3.3,	96.83,98.13,97.85,96.7	97.38
Blackbox Timed	0.01267	0.02744	0.04236	0.05601	1.27,2.74,4.24,5.6,	98.73,97.26,95.76,94.4	96.54

DENSE	<u>0.25000</u>	<u>0.50000</u>	<u>0.75000</u>	<u>1.00000</u>			
STARCRAFT							
Whitebox Predict	0.02642	0.01311	0.02364	0.07429	2.64,1.31,2.36,7.43,	97.36,98.69,97.64,92.57	96.56
Whitebox Random	0.03569	0.06809	0.15941	0.25171	3.57,6.81,15.94,25.17,	96.43,93.19,84.06,74.83	87.13
Whitebox Timed	0.00846	0.00051	0.03989	0.00153	0.85,0.05,3.99,0.15,	99.15,99.95,96.01,99.85	98.74
Blackbox Predict	0.05493	0.02443	0.02944	0.06882	5.49,2.44,2.94,6.88,	94.51,97.56,97.06,93.12	95.56
Blackbox Random	0.03583	0.02270	0.02343	0.02515	3.58,2.27,2.34,2.51,	96.42,97.73,97.66,97.49	97.32
Blackbox Timed	0.00200	0.00232	0.01632	0.07356	0.2,0.23,1.63,7.36,	99.8,99.77,98.37,92.64	97.65

BINARY DENSE	<u>0.25000</u>	<u>0.50000</u>	<u>0.75000</u>	<u>1.00000</u>
---------------------	-----------------------	-----------------------	-----------------------	-----------------------

Cooperative Navigation Attacks

Whitebox Predict	0.06123	0.07155	0.06678	0.05620
Whitebox Random	0.06068	0.02342	0.01632	0.02257
Whitebox Timed	0.03587	0.03870	0.03505	0.03445
Blackbox Predict	0.06191	0.07138	0.06613	0.05643
Blackbox Random	0.03291	0.01947	0.01589	0.02592
Blackbox Timed	0.01755	0.03169	0.06752	0.10125

6.12,7.15,6.68,5.62,	93.88,92.85,93.32,94.38	93.61
6.07,2.34,1.63,2.26,	93.93,97.66,98.37,97.74	96.92
3.59,3.87,3.5,3.44,	96.41,96.13,96.5,96.56	96.4
6.19,7.14,6.61,5.64,	93.81,92.86,93.39,94.36	93.6
3.29,1.95,1.59,2.59,	96.71,98.05,98.41,97.41	97.64
1.75,3.17,6.75,10.13,	98.25,96.83,93.25,89.88	94.55

<u>0.25000</u>	<u>0.50000</u>	<u>0.75000</u>	<u>1.00000</u>
-----------------------	-----------------------	-----------------------	-----------------------

Physical Deception Attacks

Whitebox Predict	0.14802	0.13160	0.12246	0.14354
Whitebox Random	0.06226	0.04377	0.04292	0.05456
Whitebox Timed	0.03004	0.02903	0.05488	0.07571
Blackbox Predict	0.05408	0.11327	0.20119	0.26626
Blackbox Random	0.04245	0.02255	0.02682	0.04170
Blackbox Timed	0.02065	0.04889	0.07899	0.13270

14.8,13.16,12.25,14.35,	85.2,86.84,87.75,85.65	86.36
6.23,4.38,4.29,5.46,	93.77,95.62,95.71,94.54	94.91
3,2.9,5.49,7.57,	97,97.1,94.51,92.43	95.26
5.41,11.33,20.12,26.63,	94.59,88.67,79.88,73.37	84.13
4.24,2.25,2.68,4.17,	95.76,97.75,97.32,95.83	96.66
2.07,4.89,7.9,13.27,	97.93,95.11,92.1,86.73	92.97

BINARY LSTM	<u>0.25000</u>	<u>0.50000</u>	<u>0.75000</u>	<u>1.00000</u>
--------------------	-----------------------	-----------------------	-----------------------	-----------------------

STARCRAFT

Whitebox Predict	0.04701	0.02171	0.03535	0.06730
Whitebox Random	0.04158	0.05840	0.18799	0.28558
Whitebox Timed	0.02397	0.00033	0.02241	0.00177
Blackbox Predict	0.03706	0.01657	0.02908	0.07303
Blackbox Random	0.02777	0.01698	0.01865	0.03322
Blackbox Timed	0.00492	0.00180	0.01550	0.07302

4.7,2.17,3.53,6.73,	95.3,97.83,96.47,93.27	95.72
4.16,5.84,18.8,28.56,	95.84,94.16,81.2,71.44	85.66
2.4,0.03,2.24,0.18,	97.6,99.97,97.76,99.82	98.79
3.71,1.66,2.91,7.3,	96.29,98.34,97.09,92.7	96.11
2.78,1.7,1.86,3.32,	97.22,98.3,98.14,96.68	97.58
0.49,0.18,1.55,7.3,	99.51,99.82,98.45,92.7	97.62

BINARY LSTM	<u>0.25000</u>	<u>0.50000</u>	<u>0.75000</u>	<u>1.00000</u>
--------------------	-----------------------	-----------------------	-----------------------	-----------------------

Cooperative Navigation Attacks

Whitebox Predict	0.10879	0.10536	0.08263	0.06103
Whitebox Random	0.06832	0.02463	0.01598	0.03443
Whitebox Timed	0.02871	0.03155	0.02668	0.01785
Blackbox Predict	0.11212	0.10712	0.08359	0.06099
Blackbox Random	0.04782	0.02573	0.01972	0.02347
Blackbox Timed	0.01564	0.03078	0.05971	0.08025

10.88,10.54,8.26,6.1,	89.12,89.46,91.74,93.9	91.05
6.83,2.46,1.6,3.44,	93.17,97.54,98.4,96.56	96.42
2.87,3.16,2.67,1.79,	97.13,96.84,97.33,98.21	97.38
11.21,10.71,8.36,6.1,	88.79,89.29,91.64,93.9	90.9
4.78,2.57,1.97,2.35,	95.22,97.43,98.03,97.65	97.08
1.56,3.08,5.97,8.03,	98.44,96.92,94.03,91.98	95.34

	<u>0.25000</u>	<u>0.50000</u>	<u>0.75000</u>	<u>1.00000</u>			
Physical Deception Attacks							
Whitebox Predict	0.13292	0.11080	0.09928	0.11710	13.29,11.08,9.93,11.71,	86.71,88.92,90.07,88.29	88.5
Whitebox Random	0.06729	0.04263	0.04769	0.06785	6.73,4.26,4.77,6.78,	93.27,95.74,95.23,93.22	94.36
Whitebox Timed	0.04611	0.06701	0.08341	0.10793	4.61,6.7,8.34,10.79,	95.39,93.3,91.66,89.21	92.39
Blackbox Predict	0.04617	0.09435	0.18714	0.25845	4.62,9.44,18.71,25.84,	95.38,90.56,81.29,74.16	85.35
Blackbox Random	0.05410	0.03283	0.03798	0.05720	5.41,3.28,3.8,5.72,	94.59,96.72,96.2,94.28	95.45
Blackbox Timed	0.02107	0.04779	0.08025	0.11805	2.11,4.78,8.02,11.81,	97.89,95.22,91.98,88.19	93.32

Cooperative Navigation			
Epsilon	Success Replica	Ensemble Undetected Positive Rate	
0.05	0.14014	0.16562	0.166,0.313,0.429,0.499,0.543,0.572,0.586,0.583,0.569,0.551,0.529,0.511,0.484,0.465,0.446,0.421,0.401,0.385,0.374,0.358
0.1	0.16608	0.31339	0.166,0.313,0.429,0.499,0.543,0.572,0.586,0.583,0.569,0.551,0.529,0.511,0.484,0.465,0.446,0.421,0.401,0.385,0.374,0.358
0.15	0.19421	0.42870	
0.2	0.22815	0.49904	Take all successful perturbations from replica model and apply them to true ensemble.
0.25	0.25330	0.54321	
0.3	0.27484	0.57230	
0.35	0.29093	0.58568	
0.4	0.30675	0.58269	
0.45	0.32012	0.56853	
0.5	0.33322	0.55066	
0.55	0.34763	0.52883	
0.6	0.35950	0.51113	
0.65	0.37515	0.48418	
0.7	0.39063	0.46455	
0.75	0.40426	0.44606	
0.8	0.41920	0.42093	
0.85	0.42966	0.40147	
0.9	0.44101	0.38537	
0.95	0.45323	0.37401	
1	0.46132	0.35754	

Physical Deception			
Epsilon	Success Replica	Ensemble Undetected Positive Rate	
0.05	0.13553	0.05595	0.056,0.113,0.164,0.216,0.274,0.315,0.337,0.361,0.374,0.387,0.392,0.391,0.394,0.393,0.393,0.393,0.393,0.396,0.392,0.389
0.1	0.14207	0.11348	0.056,0.113,0.164,0.216,0.274,0.315,0.337,0.361,0.374,0.387,0.392,0.391,0.394,0.393,0.393,0.393,0.393,0.396,0.392,0.389
0.15	0.15229	0.16352	
0.2	0.16657	0.21610	
0.25	0.18230	0.27408	
0.3	0.19994	0.31457	Take all successful perturbations from replica model and apply them to true ensemble.
0.35	0.21869	0.33723	
0.4	0.24176	0.36118	
0.45	0.26578	0.37417	
0.5	0.29204	0.38699	
0.55	0.31527	0.39165	
0.6	0.33826	0.39051	
0.65	0.35677	0.39396	
0.7	0.37848	0.39266	
0.75	0.39939	0.39309	
0.8	0.41719	0.39258	

0.85	0.43036	0.39299
0.9	0.44529	0.39595
0.95	0.45758	0.39194
1	0.46931	0.38861

StarCraft

Epsilon Success Replica Ensemble Undetected Positive Rate

0.05	0.15736	0.16972	0.17,0.226,0.205,0.212,0.227,0.24,0.261,0.281,0.302,0.318,0.331,0.336,0.34,0.341,0.34,0.337,0.339,0.329,0.325,0.316
0.1	0.34374	0.22582	0.17,0.226,0.205,0.212,0.227,0.24,0.261,0.281,0.302,0.318,0.331,0.336,0.34,0.341,0.34,0.337,0.339,0.329,0.325,0.316
0.15	0.58152	0.20486	
0.2	0.72332	0.21171	
0.25	0.81122	0.22750	
0.3	0.86858	0.24013	
0.35	0.89077	0.26079	
0.4	0.89932	0.28053	
0.45	0.90191	0.30177	
0.5	0.90249	0.31797	
0.55	0.89634	0.33087	
0.6	0.89413	0.33641	
0.65	0.89154	0.33987	
0.7	0.88846	0.34148	
0.75	0.88452	0.34007	
0.8	0.87780	0.33742	
0.85	0.87405	0.33887	
0.9	0.86560	0.32863	
0.95	0.85974	0.32495	
1	0.85378	0.31597	

Navigation	25%		50%		75%		100%		FN Rate	Average FN across Attack Rate by Attack
	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision		
White Predict	0.795	0.510	0.818	0.804	0.864	0.945	0.912	1.000	20.51,18.19,13.64,8.78	0.847
White Random	0.544	0.470	0.823	0.835	0.952	0.944	0.966	1.000	45.63,17.75,4.83,3.45	0.821
White Timed	0.914	0.622	0.917	0.847	0.945	0.952	0.976	1.000	8.58,8.3,5.51,2.45	0.938
Black Predict	0.703	0.441	0.816	0.808	0.862	0.947	0.911	1.000	29.73,18.45,13.79,8.9	0.823
Black Random	0.526	0.468	0.772	0.811	0.906	0.936	0.911	1.000	47.42,20.32,9.38,8.9	0.785
Black Timed	0.873	0.591	0.880	0.820	0.884	0.939	0.897	1.000	12.72,12.05,11.59,10.33	0.883
Average Recall Across	0.726	0.517	0.837	0.821	0.902	0.944	0.929	1.000	27.43,16.26,9.79,7.13	

Phys Deception	0.250		0.500		0.750		1.000		FN Rate	Average FN across Attack Rate by Attack
	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision		
White Predict	0.768	0.412	0.800	0.722	0.828	0.897	0.846	1.000	23.18,20.02,17.22,15.41	0.810
White Random	0.717	0.440	0.789	0.737	0.852	0.898	0.871	1.000	28.28,21.09,14.83,12.91	0.807
White Timed	0.799	0.493	0.790	0.720	0.818	0.867	0.852	1.000	20.06,20.98,8.3,14.76	0.840
Black Predict	0.846	0.370	0.828	0.695	0.816	0.910	0.802	1.000	15.42,17.18,18.43,19.84	0.823
Black Random	0.703	0.441	0.797	0.746	0.856	0.902	0.802	1.000	29.73,20.32,14.4,19.84	0.789
Black Timed	0.821	0.506	0.812	0.743	0.824	0.884	0.802	1.000	17.91,18.82,17.58,19.84	0.815
Average	0.776	0.444	0.803	0.727	0.832	0.893	0.829	1.000	22.43,19.74,16.78,17.1	

Starcraft	0.250		0.500		0.750		1.000		FN Rate	Average FN across Attack Rate by Attack
	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision		
White Predict	0.844	0.416	0.890	0.655	0.905	0.783	0.940	1.000	15.57,10.99,9.51,6	0.895
White Random	0.669	0.369	0.616	0.556	0.564	0.745	0.466	1.000	33.11,38.42,43.61,53.39	0.579
White Timed	0.723	0.449	0.719	0.676	0.687	0.894	0.338	1.000	27.67,28.14,31.29,66.19	0.617
Black Predict	0.860	0.416	0.888	0.652	0.871	0.775	0.892	1.000	13.98,11.22,12.92,10.75	0.878
Black Random	0.883	0.436	0.848	0.542	0.815	0.662	0.796	1.000	11.66,15.2,18.48,20.45	0.836
Black Timed	0.881	0.483	0.840	0.673	0.864	0.690	0.803	1.000	11.91,15.98,13.58,19.67	0.847
Average	0.810	0.428	0.800	0.626	0.784	0.758	0.706	1.000	18.98,19.99,21.57,29.41	