Mitigating Adversarial Gray-Box Attacks Against Phishing Detectors

Supplementary Material

I. Configuration Parameters

We provide in Table I the most influential parameter settings for each considered classifier of our evaluation.

II. Standardized Attack Tables

To facilitate understanding the effectiveness of our attacks, we provide the detailed results of the *attack* cases in Tables II to XIII. These tables have the same format as the those of the *no attack* case, i.e., Table IV and Table V of the main paper. In other words, Tables II to XIII show the F1-score, Accuracy, TPR and FPR of all the considered attacks (simple and complex) against the baseline and the POC classifiers. We report the results for for GBA-1-GBA-3, and three of the seven variants of GBA-Δ (namely, when Δ=10,40,70%).

By observing these tables we note that the FPR score is always the same between the *no-attack* and the *attack* scenarios. This is because our Gray Box attacks are evasion attacks, and aim to have a malicious (phishing) sample to be classified as benign. Hence, the false positive rate is not affected at all, because this metric would require a change in the "negative" (that is, benign) samples, which we do not touch in our evaluation. This reason motivates our focus on the Recall (or detection rate) metric for measuring the Impact of our attacks.

Table I: Most important parameters of each considered classifier. The DnW is an *EnsembleVotingClassifier* composed of DnW-1 and DnW-2.

Classifier	**Type** (as in *SciKit-learn*)	**First Parameter**	**Second Parameter**	**Third Parameter**
RF	RandomForestClassifier	*Estimators: 150*	*Criterion:* Gini	*maxFeatures:* auto
SVM	LinearSVC	*Loss:* squaredHinge	*Penalty:* L2	*C:* 0.6476
KNN	KNeighborsClassifier	*Neighbors:* 1	*Metric:* minkowski	*Algorithm:* ballTree
SGD	SGDClassifier	*Loss:* squaredHinge	*Penalty:* ElasticNet	*LearningRate:* adaptive
DT	DecisionTreeClassifier	*Splitter:* best	*Criterion:* Gini	*MaxDepth:* none
LR	LogisticRegression	*Penalty:* L2	*TOL;* 0.0001	*C:* 1.0
NB	GaussianNB	*Priors:* None	*Smoothing:* 9e-07	N/A
MLP	MLPClassifier	*Layers:* 100	*Activation:* ReLU	*Solver:* adam
AB	AdaBoostClassifier	*Estimators:* 50	*Algorithm:* SAMME.R	*LearningRate:* 1.0
ET	ExtraTreesClassifier	*Estimators:* 340	*Criterion:* Entropy	*MaxDepth:* 40
GB	GradientBoostingClassifier	*Estimators:* 200	*Criterion:* MSE	*Loss:* deviance
DnW-1	MLPClassifier	*Layers:* 64, 16, 16, 4, 4	*Activation:* ReLU	*Solver:* adam
DnW-2	MLPClassifier	*Layers:* 256	*Activation:* ReLU	*Solver:* adam
Bag	BaggingClassifier	*Estimators:* 230	*MaxSamples:* 1000	*BootstrapFeatures:* true

Table II: Simple Attack Case (GBA-1): results of the baseline classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.907	0.941	0.013	0.851	0.061	0.853	0.001	0.031	0.799	0.828	0.033	0.687	0.381	0.565	0.045	0.244
SVM	0.926	0.953	0.004	0.870	0.114	0.838	0.025	0.069	0.684	0.739	0.092	0.569	0.383	0.542	0.113	0.259
KNN	0.962	0.975	0.002	0.931	0.690	0.926	0.007	0.548	0.965	0.965	0.037	0.967	0.876	0.875	0.045	0.809
SGD	0.921	0.950	0.003	0.859	0.283	0.853	0.030	0.192	0.729	0.770	0.083	0.621	0.380	0.541	0.114	0.257
DT	0.919	0.949	0.006	0.861	0.107	0.856	0.002	0.057	0.287	0.563	0.056	0.177	0.386	0.577	0.017	0.243
LR	0.984	0.989	0.003	0.974	0.420	0.707	0.293	0.704	0.625	0.709	0.073	0.488	0.363	0.533	0.115	0.243
NB	0.802	0.886	0.010	0.683	0.089	0.816	0.050	0.059	0.754	0.794	0.046	0.633	0.385	0.579	0.008	0.240
MLP	0.917	0.948	0.001	0.848	0.414	0.884	0.007	0.271	0.805	0.834	0.023	0.689	0.435	0.595	0.026	0.284
AB	0.926	0.953	0.002	0.866	0.009	0.834	0.019	0.005	0.806	0.830	0.054	0.712	0.364	0.536	0.107	0.242
ET	0.895	0.935	0.000	0.809	0.066	0.853	0.001	0.034	0.545	0.687	0.006	0.376	0.459	0.610	0.015	0.302
GB	0.929	0.955	0.000	0.867	0.031	0.852	-0.000	0.016	0.761	0.806	0.009	0.620	0.546	0.652	0.020	0.381
DnW	0.984	0.990	0.002	0.973	0.503	0.897	0.005	0.346	0.827	0.849	0.027	0.725	0.815	0.822	0.051	0.717
Bag	0.921	0.950	0.003	0.858	0.038	0.850	0.003	0.020	0.480	0.656	0.012	0.320	0.425	0.593	0.020	0.275

Table III: Simple Attack Case (GBA-2): results of the baseline classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.805	0.887	0.013	0.690	0.907	0.974	0.001	0.834	0.891	0.899	0.033	0.830	0.910	0.906	0.045	0.866
SVM	0.705	0.845	0.004	0.548	0.401	0.871	0.025	0.286	0.890	0.893	0.092	0.877	0.824	0.821	0.113	0.766
KNN	0.601	0.806	0.002	0.432	0.924	0.978	0.007	0.892	0.055	0.499	0.037	0.030	0.761	0.781	0.045	0.637
SGD	0.483	0.768	0.003	0.320	0.217	0.845	0.030	0.142	0.077	0.482	0.083	0.043	0.846	0.840	0.114	0.802
DT	0.936	0.959	0.006	0.889	0.946	0.984	0.002	0.908	0.882	0.889	0.056	0.833	0.902	0.901	0.017	0.833
LR	0.974	0.983	0.003	0.956	0.348	0.685	0.293	0.558	0.776	0.805	0.073	0.681	0.823	0.820	0.115	0.766
NB	0.735	0.856	0.010	0.593	0.000	0.807	0.050	0.000	0.828	0.847	0.046	0.739	0.387	0.580	0.008	0.242
MLP	0.966	0.978	0.001	0.936	0.853	0.960	0.007	0.772	0.910	0.916	0.023	0.855	0.825	0.833	0.026	0.717
AB	0.947	0.966	0.002	0.904	0.793	0.943	0.019	0.728	0.881	0.888	0.054	0.830	0.884	0.876	0.107	0.862
ET	0.895	0.936	0.000	0.810	0.912	0.976	0.001	0.844	0.949	0.951	0.006	0.908	0.943	0.940	0.015	0.903
GB	0.905	0.942	0.000	0.827	0.963	0.989	-0.000	0.929	0.948	0.951	0.009	0.910	0.695	0.739	0.020	0.541
DnW	0.934	0.958	0.002	0.880	0.924	0.978	0.005	0.882	0.793	0.824	0.027	0.675	0.395	0.569	0.051	0.256
Bag	0.909	0.943	0.003	0.838	0.918	0.977	0.003	0.863	0.413	0.628	0.012	0.264	0.899	0.898	0.020	0.830

Table IV: Simple Attack Case (GBA-3): results of the baseline classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.790	0.880	0.013	0.670	0.000	0.848	0.001	0.000	0.481	0.649	0.033	0.327	0.000	0.431	0.045	0.000
SVM	0.615	0.811	0.004	0.448	0.016	0.829	0.025	0.009	0.503	0.639	0.092	0.367	0.000	0.400	0.113	0.000
KNN	0.511	0.777	0.002	0.345	0.427	0.886	0.007	0.283	0.052	0.498	0.037	0.028	0.478	0.610	0.045	0.326
SGD	0.455	0.760	0.003	0.296	0.041	0.827	0.030	0.025	0.003	0.462	0.083	0.002	0.000	0.400	0.114	0.000
DT	0.819	0.895	0.006	0.702	0.107	0.856	0.002	0.057	0.274	0.558	0.056	0.168	0.000	0.444	0.017	0.000
LR	0.978	0.985	0.003	0.962	0.206	0.646	0.293	0.304	0.259	0.545	0.073	0.160	0.000	0.400	0.115	0.000
NB	0.661	0.825	0.010	0.503	0.000	0.807	0.050	0.000	0.771	0.806	0.046	0.657	0.000	0.448	0.008	0.000
MLP	0.897	0.937	0.001	0.815	0.113	0.853	0.007	0.062	0.686	0.757	0.023	0.534	0.004	0.441	0.026	0.002
AB	0.916	0.947	0.002	0.848	0.000	0.833	0.019	0.000	0.638	0.721	0.054	0.494	0.000	0.403	0.107	0.000
ET	0.890	0.933	0.000	0.801	0.000	0.848	0.001	0.000	0.535	0.682	0.006	0.368	0.000	0.445	0.015	0.000
GB	0.898	0.937	0.000	0.814	0.000	0.849	-0.000	0.000	0.649	0.739	0.009	0.485	0.000	0.442	0.020	0.000
DnW	0.932	0.957	0.002	0.876	0.461	0.891	0.005	0.308	0.666	0.744	0.027	0.512	0.002	0.429	0.051	0.001
Bag	0.821	0.897	0.003	0.701	0.000	0.847	0.003	0.000	0.091	0.520	0.012	0.048	0.111	0.475	0.020	0.059

Table V: Simple Attack Case (GBA-1): results of the POC classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.986	0.991	0.001	0.975	0.170	0.863	0.000	0.093	0.954	0.955	0.033	0.943	0.480	0.616	0.028	0.323
SVM	0.962	0.975	0.009	0.942	0.300	0.862	0.020	0.196	0.669	0.657	0.382	0.697	0.434	0.544	0.184	0.320
KNN	0.986	0.991	0.001	0.975	0.820	0.953	0.004	0.710	0.909	0.906	0.124	0.937	0.887	0.886	0.032	0.818
SGD	0.872	0.905	0.121	0.956	0.234	0.792	0.105	0.211	0.818	0.815	0.208	0.838	0.384	0.522	0.174	0.272
DT	0.991	0.994	0.004	0.990	0.283	0.874	0.000	0.165	0.877	0.880	0.104	0.864	0.421	0.589	0.026	0.273
LR	0.954	0.970	0.008	0.927	0.316	0.586	0.422	0.632	0.777	0.795	0.128	0.718	0.360	0.506	0.187	0.254
NB	0.813	0.879	0.071	0.780	0.358	0.799	0.125	0.371	0.732	0.663	0.599	0.927	0.421	0.528	0.209	0.312
MLP	0.983	0.989	0.001	0.969	0.667	0.919	0.014	0.539	0.878	0.872	0.180	0.925	0.340	0.548	0.045	0.213
AB	0.974	0.983	0.005	0.959	0.116	0.847	0.014	0.067	0.869	0.873	0.108	0.853	0.580	0.636	0.148	0.458
ET	0.989	0.993	0.000	0.978	0.255	0.871	0.000	0.146	0.923	0.927	0.029	0.882	0.550	0.651	0.028	0.388
GB	0.989	0.993	0.001	0.981	0.177	0.863	0.001	0.097	0.945	0.946	0.040	0.932	0.648	0.709	0.022	0.488
DnW	0.984	0.989	0.004	0.975	0.468	0.892	0.006	0.315	0.792	0.809	0.114	0.731	0.827	0.828	0.077	0.750
Bag	0.986	0.991	0.003	0.979	0.183	0.864	0.000	0.101	0.934	0.936	0.044	0.916	0.655	0.712	0.029	0.498

Table VI: Simple Attack Case (GBA-2): results of the POC classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.881	0.928	0.001	0.790	0.689	0.928	0.000	0.525	0.911	0.916	0.033	0.864	0.957	0.954	0.028	0.939
SVM	0.952	0.969	0.009	0.925	0.278	0.859	0.020	0.180	0.806	0.775	0.382	0.935	0.880	0.865	0.184	0.905
KNN	0.647	0.823	0.001	0.480	0.080	0.852	0.004	0.042	0.898	0.896	0.124	0.917	0.928	0.925	0.032	0.889
SGD	0.887	0.915	0.121	0.986	0.089	0.771	0.105	0.074	0.667	0.699	0.208	0.605	0.879	0.865	0.174	0.897
DT	0.938	0.960	0.004	0.889	0.822	0.954	0.000	0.698	0.835	0.844	0.104	0.792	0.904	0.902	0.026	0.842
LR	0.980	0.986	0.008	0.975	0.318	0.587	0.422	0.638	0.785	0.801	0.128	0.730	0.880	0.864	0.187	0.906
NB	0.835	0.891	0.071	0.816	0.074	0.753	0.125	0.065	0.699	0.631	0.599	0.863	0.894	0.877	0.209	0.948
MLP	0.907	0.942	0.001	0.831	0.887	0.967	0.014	0.860	0.849	0.846	0.180	0.872	0.870	0.869	0.045	0.798
AB	0.938	0.960	0.005	0.892	0.560	0.901	0.014	0.419	0.839	0.847	0.108	0.801	0.918	0.906	0.148	0.951
ET	0.887	0.932	0.000	0.798	0.773	0.944	0.000	0.630	0.886	0.895	0.029	0.819	0.957	0.954	0.028	0.938
GB	0.903	0.940	0.001	0.824	0.687	0.928	0.001	0.526	0.907	0.912	0.040	0.863	0.858	0.861	0.022	0.765
DnW	0.917	0.948	0.004	0.853	0.775	0.943	0.006	0.653	0.894	0.894	0.114	0.901	0.532	0.628	0.077	0.385
Bag	0.875	0.924	0.003	0.782	0.653	0.922	0.000	0.485	0.880	0.889	0.044	0.821	0.930	0.926	0.029	0.890

Table VII: Simple Attack Case (GBA-3): results of the POC classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.810	0.892	0.001	0.682	0.191	0.865	0.000	0.106	0.775	0.811	0.033	0.654	0.381	0.571	0.028	0.241
SVM	0.982	0.988	0.009	0.982	0.037	0.835	0.020	0.021	0.586	0.597	0.382	0.575	0.802	0.792	0.184	0.771
KNN	0.555	0.791	0.001	0.385	0.888	0.969	0.004	0.817	0.898	0.896	0.124	0.917	0.839	0.844	0.032	0.742
SGD	0.651	0.783	0.121	0.597	0.237	0.792	0.105	0.214	0.053	0.414	0.208	0.033	0.784	0.777	0.174	0.737
DT	0.917	0.948	0.004	0.853	0.432	0.891	0.000	0.275	0.739	0.773	0.104	0.648	0.639	0.703	0.026	0.480
LR	0.661	0.826	0.008	0.502	0.382	0.611	0.422	0.798	0.838	0.844	0.128	0.815	0.791	0.781	0.187	0.755
NB	0.628	0.791	0.071	0.521	0.065	0.752	0.125	0.057	0.762	0.693	0.599	0.987	0.869	0.851	0.209	0.901
MLP	0.860	0.917	0.001	0.756	0.294	0.865	0.014	0.186	0.671	0.709	0.180	0.597	0.654	0.708	0.045	0.504
AB	0.931	0.956	0.005	0.879	0.074	0.844	0.014	0.041	0.742	0.774	0.108	0.654	0.631	0.668	0.148	0.517
ET	0.836	0.905	0.000	0.719	0.306	0.876	0.000	0.180	0.756	0.799	0.029	0.626	0.608	0.684	0.028	0.447
GB	0.909	0.943	0.001	0.834	0.136	0.859	0.001	0.073	0.752	0.794	0.040	0.626	0.376	0.571	0.022	0.236
DnW	0.925	0.953	0.004	0.868	0.422	0.886	0.006	0.277	0.719	0.757	0.114	0.626	0.497	0.610	0.077	0.352
Bag	0.798	0.886	0.003	0.667	0.204	0.866	0.000	0.114	0.763	0.801	0.044	0.644	0.443	0.598	0.029	0.291

Table VIII: Complex Attack Case (GBA-Δ, $\Delta = 10\%$): results of the baseline classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.973	0.982	0.013	0.971	0.939	0.983	0.001	0.891	0.961	0.962	0.033	0.956	0.958	0.954	0.045	0.954
SVM	0.946	0.965	0.004	0.904	0.420	0.874	0.025	0.303	0.823	0.837	0.092	0.765	0.862	0.855	0.113	0.829
KNN	0.996	0.997	0.002	0.996	0.965	0.989	0.007	0.969	0.903	0.909	0.037	0.854	0.892	0.889	0.045	0.834
SGD	0.838	0.905	0.003	0.725	0.465	0.877	0.030	0.354	0.910	0.911	0.083	0.904	0.820	0.817	0.114	0.760
DT	0.988	0.992	0.006	0.987	0.987	0.996	0.002	0.986	0.896	0.901	0.056	0.858	0.812	0.824	0.017	0.694
LR	0.943	0.963	0.003	0.897	0.400	0.700	0.293	0.662	0.894	0.898	0.073	0.868	0.824	0.820	0.115	0.767
NB	0.920	0.949	0.010	0.869	0.668	0.904	0.050	0.642	0.835	0.853	0.046	0.750	0.449	0.607	0.008	0.291
MLP	0.937	0.960	0.001	0.883	0.882	0.967	0.007	0.820	0.913	0.919	0.023	0.860	0.837	0.843	0.026	0.735
AB	0.984	0.989	0.002	0.972	0.870	0.962	0.019	0.852	0.924	0.926	0.054	0.905	0.829	0.825	0.107	0.770
ET	1.000	1.000	0.000	1.000	0.982	0.995	0.001	0.970	0.882	0.895	0.006	0.794	0.932	0.929	0.015	0.883
GB	0.976	0.984	0.000	0.954	0.986	0.996	0.000	0.972	0.973	0.973	0.009	0.955	0.923	0.920	0.020	0.871
DnW	0.982	0.988	0.002	0.968	0.920	0.977	0.005	0.875	0.950	0.951	0.027	0.929	0.856	0.856	0.051	0.780
Bag	0.985	0.990	0.003	0.977	0.936	0.982	0.003	0.895	0.917	0.923	0.012	0.857	0.959	0.956	0.020	0.936

Table IX: Complex Attack Case (GBA-Δ, $\Delta = 40\%$): results of the baseline classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.895	0.934	0.013	0.831	0.720	0.934	0.001	0.566	0.670	0.745	0.033	0.520	0.656	0.709	0.045	0.506
SVM	0.655	0.825	0.004	0.491	0.421	0.874	0.025	0.304	0.754	0.785	0.092	0.661	0.541	0.623	0.113	0.405
KNN	0.983	0.989	0.002	0.971	0.645	0.918	0.007	0.494	0.519	0.665	0.037	0.363	0.597	0.673	0.045	0.442
SGD	0.767	0.871	0.003	0.625	0.404	0.868	0.030	0.296	0.830	0.844	0.083	0.769	0.578	0.644	0.114	0.445
DT	0.935	0.958	0.006	0.887	0.881	0.968	0.002	0.797	0.652	0.729	0.056	0.511	0.748	0.776	0.017	0.606
LR	0.792	0.883	0.003	0.659	0.233	0.653	0.293	0.350	0.694	0.750	0.073	0.570	0.514	0.607	0.115	0.379
NB	0.845	0.908	0.010	0.747	0.430	0.860	0.050	0.351	0.902	0.907	0.046	0.859	0.612	0.691	0.008	0.444
MLP	0.760	0.869	0.001	0.614	0.688	0.925	0.007	0.545	0.852	0.869	0.023	0.760	0.516	0.635	0.026	0.356
AB	0.986	0.991	0.002	0.977	0.738	0.931	0.019	0.647	0.721	0.771	0.054	0.595	0.701	0.726	0.107	0.588
ET	0.975	0.984	0.000	0.952	0.842	0.959	0.001	0.731	0.502	0.667	0.006	0.337	0.556	0.659	0.015	0.390
GB	0.792	0.883	0.000	0.655	0.871	0.966	0.000	0.772	0.798	0.831	0.009	0.670	0.550	0.654	0.020	0.385
DnW	0.959	0.973	0.002	0.925	0.920	0.977	0.005	0.876	0.813	0.839	0.027	0.703	0.648	0.702	0.051	0.499
Bag	0.831	0.901	0.003	0.714	0.792	0.947	0.003	0.667	0.580	0.702	0.012	0.413	0.640	0.705	0.020	0.479

Table X: Complex Attack Case (GBA-Δ, $\Delta = 70\%$): results of the baseline classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.900	0.937	0.013	0.839	0.570	0.909	0.001	0.401	0.558	0.685	0.033	0.400	0.950	0.946	0.045	0.938
SVM	0.466	0.763	0.004	0.306	0.564	0.896	0.025	0.448	0.471	0.624	0.092	0.337	0.734	0.748	0.113	0.634
KNN	0.979	0.986	0.002	0.963	0.585	0.908	0.007	0.430	0.443	0.631	0.037	0.295	0.303	0.533	0.045	0.185
SGD	0.547	0.788	0.003	0.379	0.404	0.868	0.030	0.296	0.732	0.772	0.083	0.626	0.686	0.713	0.114	0.571
DT	0.873	0.923	0.006	0.784	0.832	0.956	0.002	0.720	0.600	0.700	0.056	0.453	0.000	0.444	0.017	0.000
LR	0.704	0.844	0.003	0.546	0.242	0.655	0.293	0.365	0.644	0.719	0.073	0.510	0.811	0.809	0.115	0.746
NB	0.496	0.769	0.010	0.336	0.276	0.838	0.050	0.205	0.850	0.864	0.046	0.773	0.911	0.909	0.008	0.841
MLP	0.511	0.777	0.001	0.344	0.338	0.875	0.007	0.211	0.747	0.795	0.023	0.610	0.832	0.839	0.026	0.728
AB	0.987	0.992	0.002	0.979	0.429	0.879	0.019	0.302	0.661	0.734	0.054	0.521	0.741	0.755	0.107	0.641
ET	0.954	0.970	0.000	0.912	0.467	0.895	0.001	0.307	0.351	0.606	0.006	0.214	0.945	0.942	0.015	0.907
GB	0.639	0.821	0.000	0.470	0.733	0.937	0.000	0.579	0.549	0.688	0.009	0.382	0.687	0.734	0.020	0.532
DnW	0.836	0.904	0.002	0.721	0.776	0.943	0.005	0.651	0.515	0.666	0.027	0.356	0.520	0.629	0.051	0.366
Bag	0.785	0.880	0.003	0.650	0.733	0.935	0.003	0.588	0.200	0.552	0.012	0.112	0.458	0.608	0.020	0.302

Table XI: Complex Attack Case (GBA-Δ, $\Delta = 10\%$): results of the POC classifiers for each dataset.

	LNU-Phish				DeltaPhish				Mendeley Phishing				UCI Phishing			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.999	0.999	0.001	0.999	0.954	0.987	0.000	0.911	0.930	0.933	0.033	0.898	0.957	0.954	0.028	0.938
SVM	0.946	0.964	0.009	0.913	0.362	0.869	0.020	0.246	0.721	0.699	0.382	0.781	0.773	0.767	0.184	0.726
KNN	0.999	0.999	0.001	1.000	0.992	0.998	0.004	1.007	0.905	0.903	0.124	0.930	0.936	0.932	0.032	0.903
SGD	0.807	0.865	0.121	0.836	0.446	0.829	0.105	0.457	0.820	0.816	0.208	0.840	0.676	0.693	0.174	0.584
DT	0.990	0.993	0.004	0.988	0.947	0.985	0.000	0.899	0.797	0.814	0.104	0.732	0.849	0.853	0.026	0.753
LR	0.928	0.954	0.008	0.879	0.291	0.578	0.422	0.575	0.788	0.803	0.128	0.734	0.839	0.824	0.187	0.834
NB	0.883	0.919	0.071	0.900	0.622	0.859	0.125	0.769	0.732	0.662	0.599	0.926	0.720	0.719	0.209	0.659
MLP	0.937	0.960	0.001	0.883	0.812	0.949	0.014	0.738	0.817	0.818	0.180	0.815	0.838	0.841	0.045	0.748
AB	0.989	0.992	0.005	0.987	0.910	0.973	0.014	0.900	0.857	0.862	0.108	0.831	0.798	0.794	0.148	0.745
ET	1.000	1.000	0.000	1.001	0.956	0.987	0.000	0.916	0.943	0.945	0.029	0.919	0.942	0.938	0.028	0.911
GB	0.982	0.988	0.001	0.966	0.968	0.991	0.001	0.944	0.920	0.924	0.040	0.887	0.914	0.911	0.022	0.857
DnW	0.974	0.983	0.004	0.956	0.861	0.962	0.006	0.782	0.915	0.914	0.114	0.941	0.857	0.854	0.077	0.798
Bag	0.989	0.992	0.003	0.983	0.971	0.992	0.000	0.944	0.868	0.878	0.044	0.800	0.947	0.943	0.029	0.921

Table XII: Complex Attack Case (GBA-Δ, $\Delta = 40\%$): results of the POC classifiers for each dataset.

	LNU-Phish				**DeltaPhish**				**Mendeley Phishing**				**UCI Phishing**			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.851	0.912	0.001	0.742	0.839	0.958	0.000	0.723	0.736	0.785	0.033	0.602	0.474	0.613	0.028	0.318
SVM	0.792	0.882	0.009	0.668	0.413	0.876	0.020	0.290	0.751	0.725	0.382	0.833	0.488	0.572	0.184	0.372
KNN	0.986	0.991	0.001	0.974	0.916	0.976	0.004	0.865	0.732	0.763	0.124	0.650	0.617	0.688	0.032	0.458
SGD	0.739	0.827	0.121	0.724	0.347	0.810	0.105	0.334	0.767	0.772	0.208	0.752	0.711	0.719	0.174	0.630
DT	0.921	0.950	0.004	0.860	0.739	0.938	0.000	0.586	0.583	0.676	0.104	0.454	0.597	0.678	0.026	0.435
LR	0.828	0.899	0.008	0.717	0.185	0.543	0.422	0.345	0.698	0.739	0.128	0.605	0.489	0.572	0.187	0.374
NB	0.841	0.894	0.071	0.826	0.484	0.825	0.125	0.544	0.752	0.683	0.599	0.968	0.673	0.683	0.209	0.594
MLP	0.824	0.899	0.001	0.702	0.580	0.904	0.014	0.440	0.641	0.689	0.180	0.557	0.282	0.524	0.045	0.170
AB	0.914	0.946	0.005	0.850	0.638	0.913	0.014	0.505	0.695	0.742	0.108	0.590	0.680	0.702	0.148	0.578
ET	0.977	0.985	0.000	0.955	0.777	0.945	0.000	0.636	0.704	0.766	0.029	0.559	0.958	0.955	0.028	0.940
GB	0.864	0.919	0.001	0.761	0.813	0.952	0.001	0.688	0.795	0.824	0.040	0.687	0.560	0.659	0.022	0.396
DnW	0.974	0.983	0.004	0.957	0.910	0.974	0.006	0.863	0.748	0.777	0.114	0.666	0.681	0.718	0.077	0.549
Bag	0.898	0.937	0.003	0.819	0.865	0.964	0.000	0.761	0.612	0.710	0.044	0.461	0.642	0.704	0.029	0.484

Table XIII: Complex Attack Case (GBA-Δ, $\Delta = 70\%$): results of the POC classifiers for each dataset.

	LNU-Phish				**DeltaPhish**				**Mendeley Phishing**				**UCI Phishing**			
Classifier	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*	*F1*	*Acc*	*FPR*	*TPR*
RF	0.623	0.815	0.001	0.454	0.646	0.921	0.000	0.477	0.494	0.655	0.033	0.339	0.930	0.927	0.028	0.890
SVM	0.536	0.782	0.009	0.372	0.450	0.881	0.020	0.323	0.446	0.509	0.382	0.398	0.756	0.753	0.184	0.700
KNN	0.981	0.988	0.001	0.965	0.843	0.958	0.004	0.746	0.374	0.569	0.124	0.259	0.647	0.706	0.032	0.491
SGD	0.502	0.722	0.121	0.414	0.442	0.828	0.105	0.451	0.779	0.782	0.208	0.772	0.840	0.827	0.174	0.828
DT	0.787	0.880	0.004	0.654	0.668	0.925	0.000	0.502	0.678	0.732	0.104	0.567	0.871	0.872	0.026	0.788
LR	0.717	0.848	0.008	0.568	0.239	0.560	0.422	0.457	0.434	0.594	0.128	0.313	0.858	0.843	0.187	0.867
NB	0.502	0.744	0.071	0.382	0.227	0.776	0.125	0.218	0.763	0.694	0.599	0.990	0.915	0.899	0.209	0.989
MLP	0.942	0.963	0.001	0.893	0.842	0.956	0.014	0.784	0.621	0.677	0.180	0.532	0.760	0.779	0.045	0.635
AB	0.764	0.869	0.005	0.624	0.438	0.883	0.014	0.302	0.706	0.749	0.108	0.605	0.817	0.810	0.148	0.775
ET	0.839	0.906	0.000	0.723	0.505	0.900	0.000	0.337	0.495	0.656	0.029	0.339	0.861	0.863	0.028	0.773
GB	0.793	0.884	0.001	0.658	0.703	0.931	0.001	0.545	0.599	0.703	0.040	0.444	0.957	0.953	0.022	0.933
DnW	0.852	0.912	0.004	0.748	0.772	0.942	0.006	0.649	0.510	0.635	0.114	0.382	0.534	0.629	0.077	0.387
Bag	0.972	0.982	0.003	0.952	0.807	0.951	0.000	0.676	0.492	0.650	0.044	0.340	0.735	0.765	0.029	0.595