Encryption agnostic classifiers of traffic originators and their application to anomaly detection

Traffic statistics dataset and insights on machine learning models

Daniele Canavese[a]
daniele.canavese@polito.it

Leonardo Regano[a]
leonardo.regano@polito.it

Cataldo Basile[a]
cataldo.basile@polito.it

Gabriele Ciravegna[b]
gabriele.ciravegna@unifi.it

Antonio Lioy[a]
lioy@polito.it

[a] Dipartimento di Automatica e Informatica, Politecnico di Torino, 10129 Torino, Italy
[b] Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Firenze, 53100 Siena, Italy

Data set

This section reports several statistics about the data set. Table 1 lists the tools that have been used to generate the traffic that has been considered in the presented data set. Table 2 reports the features that have been used to train and test the machine-learning models.

APPLICATION	CATEGORY	WINDOWS	LINUX
Chrome 48	browser	✔	✔
Chrome 68	browser	✔	✔
Firefox 42	browser	✔	✔
Firefox 62	browser	✔	✔
Firefox 68	browser	○	✔
Edge 42	browser	✔	○
Opera 62	browser	✔	○
GoldenEye 3.49.2	stress tool	✔	✔
HULK 1.0	stress tool	✔	✔
RudyJS 1.0.0	stress tool	✔	✔
SlowHTTPTest 1.6	stress tool	○	✔
SlowLoris 7.70	stress tool	✔	✔
Curl 7.55	web crawler	✔	✔
GrabSite 2.1.16	web crawler	○	✔
Httrack 3.49.2	web crawler	✔	✔
Wget 1.19	web crawler	✔	✔
Wpull 2.0.1	web crawler	✔	✔

Table 1: Tools used to generate the traffic considered in the experiments.

	FEATURE	UNIT
1	# packets (both directions)	*packets*
2	# packets with payload (both directions)	*packets*
3	# retransmitted packets (both directions)	*packets*
4	# out of sequence packets (both directions)	*packets*
5	# packets with ACK set (both directions)	*packets*
6	# packets with ACK set and no payload (both directions)	*packets*
7	# packets with FIN set (both directions)	*packets*
8	# packets with RST set (both directions)[1]	*packets*
9	# packets with SYN set (both directions)	*packets*
10	# payload bytes excluding retransmissions (both directions)	*bytes*
11	# payload bytes including retransmissions (both directions)	*bytes*
12	# retransmitted bytes (both directions)	*bytes*
13	flow duration	*ms*
14	relative time of first payload packet (both directions)	*ms*
15	relative time of last payload packet (both directions)	*ms*
16	relative time of first ACK packet (both directions)	*ms*
17	TCP connection correctly terminated	*boolean*

Table 2: TCP statistics used as classification features.

Finally, several statistics, grouped by labels, are reported: the average number of packets and bytes send by the client or by the server and the average connection duration in milliseconds. Table 3, 4 and 5 respectively reports the averages for the three categories (browsers, crawlers and DoS tools), all the tools and their instances in data set.

[1] Actually this can be only 0 or 1 since a proper TCP implementation will reset a connection after receiving a RST packet.

	SENT BY CLIENT		SENT BY SERVER		
CATEGORY	PACKETS	BYTES	PACKETS	BYTES	DURATION $[ms]$
browser	30.590	2381.743	41.748	46519.913	25495.701
crawler	95.668	1568.130	158.849	344685.003	4589.703
dos	10.974	738.409	16.329	17955.801	2895.279

Table 3: Means of some features for the category in our data set.

	SENT BY CLIENT		SENT BY SERVER		
TOOL	PACKETS	BYTES	PACKETS	BYTES	DURATION $[ms]$
chrome	30.605	2587.957	43.548	46727.950	36023.258
curl	48.260	539.707	71.030	91280.696	608.296
edge	23.135	2024.088	21.693	21782.879	12693.361
firefox	40.620	2744.452	60.108	71260.752	24748.461
goldeneye	13.061	800.137	21.071	24220.021	1409.918
grabsite	368.901	3929.413	583.037	1803017.128	15427.538
httrack	16.424	1009.946	21.192	23018.443	2517.901
hulk	5.711	573.383	4.576	2659.303	5909.654
opera	25.263	1914.553	47.881	53419.715	40428.032
rudy	11.332	713.800	10.997	3342.403	15770.126
slowhttptest	8.640	1406.865	6.826	3494.015	11974.112
slowloris	5.280	164.620	3.890	47.859	13641.368
wget	129.312	2134.542	246.652	328985.405	2756.862
wpull	115.476	1239.060	214.299	296092.743	8558.179

Table 4: Means of some features for the tool in our data set.

	SENT BY CLIENT		SENT BY SERVER		
TOOL INSTANCE	PACKETS	BYTES	PACKETS	BYTES	DURATION $[ms]$
chrome-48.0.2564.109	31.145	2337.780	41.950	41918.843	34469.858
chrome-68.0.3440.84	29.843	2941.171	45.803	53517.726	38216.437
curl-7.55.1	31.203	649.631	53.382	65814.020	431.330
curl-7.61.0	67.340	416.752	90.771	119766.433	806.241
edge-42.17134.1.0	23.135	2024.088	21.693	21782.879	12693.361
firefox-42.0	37.651	3162.645	57.196	66622.033	24610.871
firefox-62.0	49.066	3359.130	72.157	84900.460	33731.822
firefox-68.0	30.374	1155.053	43.732	54509.414	9834.131
goldeneye-2.1	13.061	800.137	21.071	24220.021	1409.918
grabsite-2.1.16	368.901	3929.413	583.037	1803017.128	15427.538
httrack-3.49.2	16.424	1009.946	21.192	23018.443	2517.901
hulk-1.0	5.711	573.383	4.576	2659.303	5909.654
opera-62.0.3331.66	25.263	1914.553	47.881	53419.715	40428.032
rudy-1.0.0	11.332	713.800	10.997	3342.403	15770.126
slowhttptest-1.6	8.640	1406.865	6.826	3494.015	11974.112
slowloris-0.1.4	5.404	164.220	3.973	48.299	13392.927
slowloris-0.1.5	5.159	165.008	3.810	47.434	13881.584
wget-1.11.4	92.605	1024.641	184.089	249312.020	2076.603
wget-1.19.5	176.123	3549.987	326.437	430592.154	3624.391
wpull-2.0.1	115.476	1239.060	214.299	296092.743	8558.179

Table 5: Means of some features for the tool instance in our data set.

Classifiers

This section reports several statistics and plots about the models for classifying the traffic into various classes. Three different models have been considered for each classification task: a random forest (via the `RandomForestClassifier` class in `scikit-learn`), an extra-trees (via the `ExtraTreesClassifier` class in `scikit-learn`) and a neural network (a custom class implemented in `PyTorch` and `skorch`). The optimization process was performed using the `hyperopt` package using a Bayesian optimization procedure.

For each classifier, the following data are reported:

- the plots showing the values of the R_k statistics as our Bayesian hyper-parameters optimization process progressed (Figs. 1, 3, 5, 7, 9, 11, 13, 15 and 17);
- the tables listing the optimal hyper-parameters found by our Bayesian optimization process (Tables 6, 10, 14, 18, 22, 26, 30, 34 and 38) – we normally used the default values for the hyper-parameters not reported[2];
- the tables reporting several classification statistics computed on the training set, development set, known tools test Set and unknown tools test set (Tables 7, 11, 15, 19, 23, 27, 31, 35 and 39);
- the confusion matrices for each classifier (Tables. 8, 12, 16, 20, 24, 28, 32, 36 and 40);
- the plots depicting how the balanced accuracy changes as the number of exchanged packets increases (Figs. 2, 4, 6, 8, 10, 12, 14, 16 and 18);
- the results of the classification of the unknown tools (Tables 9, 13, 17, 21, 25, 29, 33, 37 and 41).

Category classifiers

This section reports several statistics and plots about the models for classifying the traffic into categories (e.g. BROWSER, CRAWLER and DOS a.k.a. network stress tools).

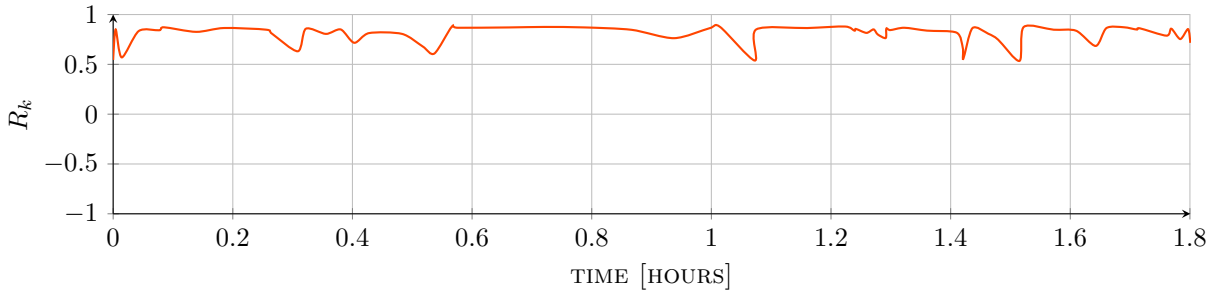

Figure 1: Hyper-parameters optimization plot for the category classifier based on random forest.

HYPER-PARAMETER	VALUE
`criterion`	entropy
`max_depth`	17
`min_samples_leaf`	9
`min_samples_split`	38
`n_estimators`	89

Table 6: Optimal hyper-parameters for the category classifier based on random forest.

[2]The most notable exceptions are given by the neural networks' batch size and number of epochs, that we chose to set to 1024 and 50, respectively.

STATISTIC	TRAINING SET	DEV SET	KTS	UTS
samples	955872	119484	119485	30144
accuracy [%]	97.702	97.542	97.447	40.459
balanced accuracy [%]	95.568	93.997	94.041	34.031
precision [%]	86.660	85.637	85.226	32.723
recall [%]	95.568	93.997	94.041	34.031
Cohen's kappa [%]	86.105	85.021	84.577	1.097
F-score [%]	90.719	89.459	89.230	32.151
Jaccard score [%]	83.556	81.607	81.255	20.631
Hamming loss	0.023	0.025	0.026	0.595
zero-one loss	0.023	0.025	0.026	0.595
R_k	0.865	0.854	0.850	0.011

Table 7: Classification statistics for the category classifier based on random forest.

		INFERRED		
		BROWSER	CRAWLER	DOS
TARGET	BROWSER	6872	194	322
	CRAWLER	120	2112	83
	DOS	1980	352	107450

Table 8: Confusion matrix for the category classifier based on random forest on the KTS.

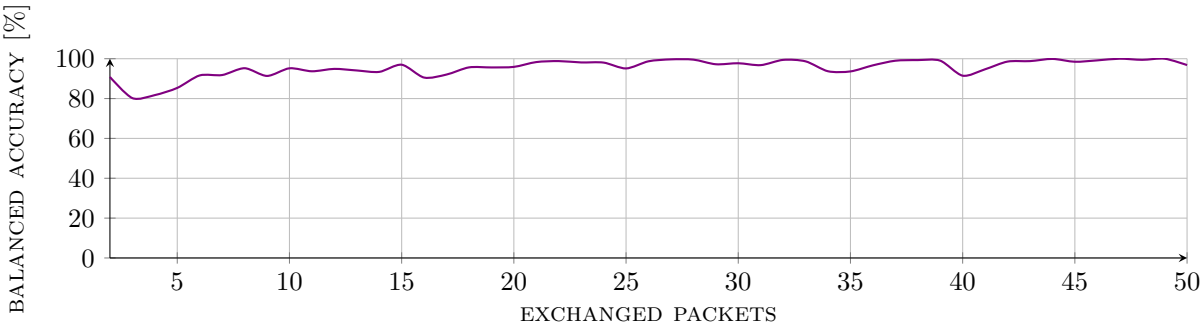

Figure 2: Balanced accuracy vs. exchange packets plot for the category classifier based on random forest on the KTS.

INFERRED CLASS	SAMPLES
browser	3291
crawler	2623
dos	620

(a) Classification of FIREFOX-68.0.

INFERRED CLASS	SAMPLES
browser	1369
crawler	883
dos	1413

(b) Classification of GRABSITE-2.1.16.

INFERRED CLASS	SAMPLES
browser	6149
crawler	605
dos	2196

(c) Classification of OPERA-62.0.3331.66.

INFERRED CLASS	SAMPLES
browser	6657
crawler	2465
dos	1873

(d) Classification of SLOWHTTPTEST-1.6.

Table 9: Classification of unknown tools for the category classifier based on random forest.

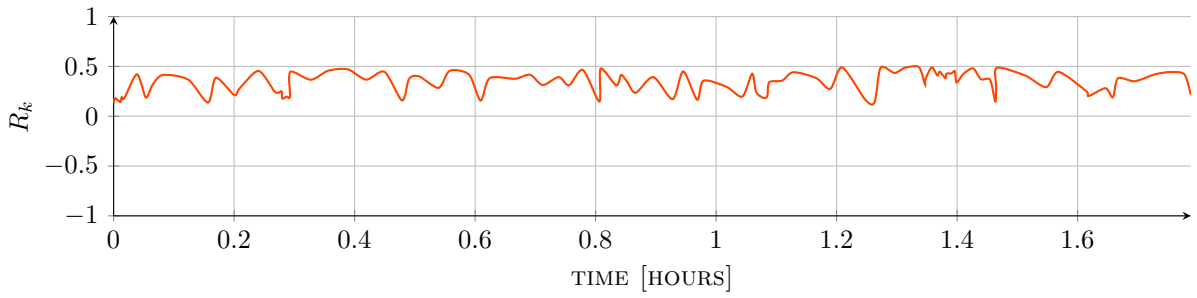

Figure 3: Hyper-parameters optimization plot for the category classifier based on extra-trees.

HYPER-PARAMETER	VALUE
`criterion`	entropy
`max_depth`	20
`min_samples_leaf`	3
`min_samples_split`	37
`n_estimators`	88

Table 10: Optimal hyper-parameters for the category classifier based on extra-trees.

STATISTIC	TRAINING SET	DEV SET	KTS	UTS
samples	955872	119484	119485	30144
accuracy [%]	85.085	85.033	85.015	37.646
balanced accuracy [%]	80.692	80.216	80.373	38.510
precision [%]	62.558	62.265	62.413	33.339
recall [%]	80.692	80.216	80.373	38.510
Cohen's kappa [%]	42.760	42.372	42.547	1.694
F-score [%]	62.138	61.837	61.866	33.949
Jaccard score [%]	51.661	51.322	51.552	21.426
Hamming loss	0.149	0.150	0.150	0.624
zero-one loss	0.149	0.150	0.150	0.624
R_k	0.491	0.487	0.489	0.017

Table 11: Classification statistics for the category classifier based on extra-trees.

		INFERRED		
		BROWSER	CRAWLER	DOS
TARGET	BROWSER	5225	1496	667
	CRAWLER	126	1954	235
	DOS	1439	13942	94401

Table 12: Confusion matrix for the category classifier based on extra-trees on the KTS.

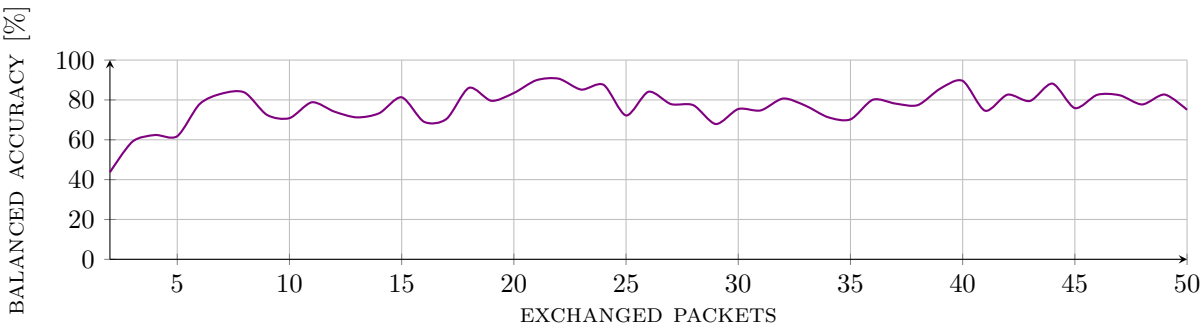

Figure 4: Balanced accuracy vs. exchange packets plot for the category classifier based on extra-trees on the KTS.

INFERRED CLASS	SAMPLES
browser	2570
crawler	1242
dos	2722

(a) Classification of FIREFOX-68.0.

INFERRED CLASS	SAMPLES
browser	837
crawler	1809
dos	1019

(b) Classification of GRABSITE-2.1.16.

INFERRED CLASS	SAMPLES
browser	5237
crawler	1046
dos	2667

(c) Classification of OPERA-62.0.3331.66.

INFERRED CLASS	SAMPLES
browser	5077
crawler	4186
dos	1732

(d) Classification of SLOWHTTPTEST-1.6.

Table 13: Classification of unknown tools for the category classifier based on extra-trees.

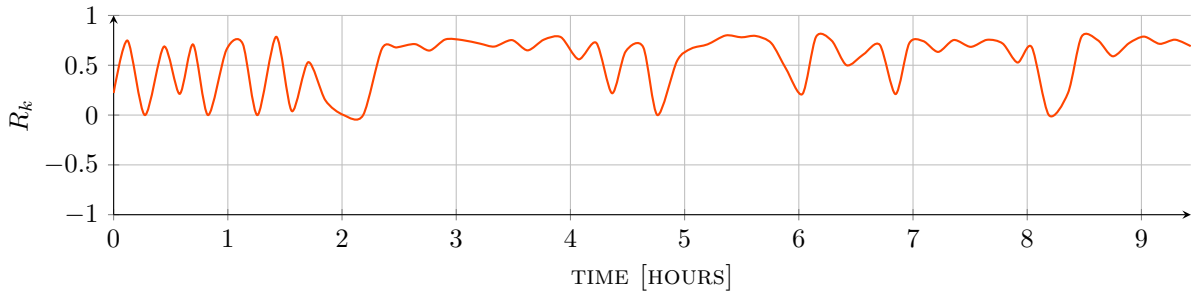

Figure 5: Hyper-parameters optimization plot for the category classifier based on neural network.

HYPER-PARAMETER	VALUE
`lr`	0.0027014308955057255
`module__layers`	4
`module__neurons_per_layer`	434
`module__p`	0.1292524544974843

Table 14: Optimal hyper-parameters for the category classifier based on neural network.

STATISTIC	TRAINING SET	DEV SET	KTS	UTS
samples	955872	119484	119485	30144
accuracy [%]	96.132	96.005	96.017	41.912
balanced accuracy [%]	91.123	90.207	90.308	35.058
precision [%]	75.600	74.957	74.653	33.739
recall [%]	91.123	90.207	90.308	35.058
Cohen's kappa [%]	77.644	76.827	76.988	2.587
F-score [%]	80.923	80.166	79.902	33.950
Jaccard score [%]	70.873	69.961	69.772	21.943
Hamming loss	0.039	0.040	0.040	0.581
zero-one loss	0.039	0.040	0.040	0.581
R_k	0.784	0.776	0.778	0.026

Table 15: Classification statistics for the category classifier based on neural network.

		INFERRED		
		BROWSER	CRAWLER	DOS
TARGET	BROWSER	6423	489	476
	CRAWLER	148	2018	149
	DOS	1270	2227	106285

Table 16: Confusion matrix for the category classifier based on neural network on the KTS.

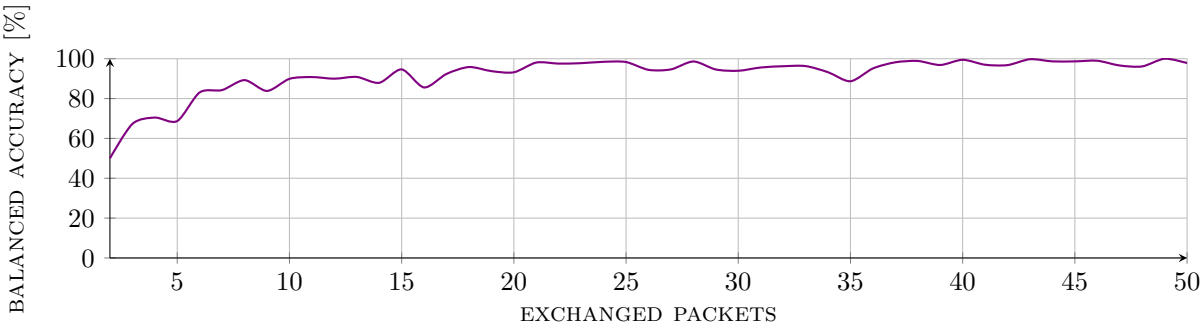

Figure 6: Balanced accuracy vs. exchange packets plot for the category classifier based on neural network on the KTS.

INFERRED CLASS	SAMPLES
browser	3366
crawler	358
dos	2810

(a) Classification of FIREFOX-68.0.

INFERRED CLASS	SAMPLES
browser	2012
crawler	835
dos	818

(b) Classification of GRABSITE-2.1.16.

INFERRED CLASS	SAMPLES
browser	6086
crawler	738
dos	2126

(c) Classification of OPERA-62.0.3331.66.

INFERRED CLASS	SAMPLES
browser	5201
crawler	3447
dos	2347

(d) Classification of SLOWHTTPTEST-1.6.

Table 17: Classification of unknown tools for the category classifier based on neural network.

Tool classifiers

This section reports several statistics and plots about the models for classifying the traffic into tools (e.g. GOLDENEYE, HULK, FIREFOX, WGET, EDGE, HTTRACK, CHROME, RUDY, SLOWLORIS, CURL and WPULL).

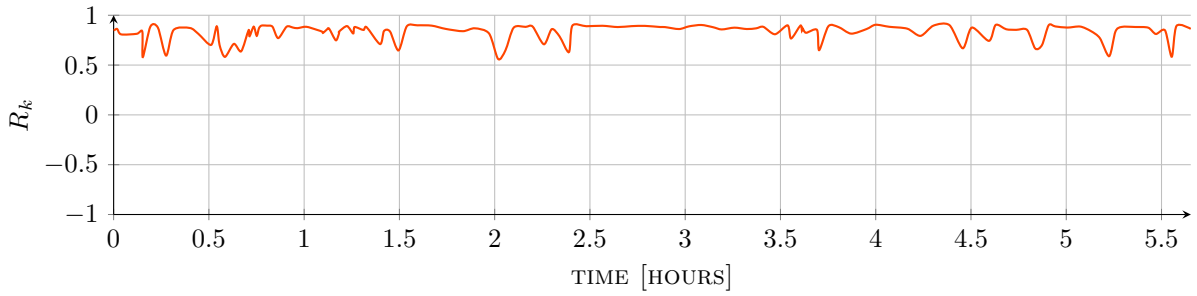

Figure 7: Hyper-parameters optimization plot for the tool classifier based on random forest.

HYPER-PARAMETER	VALUE
`criterion`	entropy
`max_depth`	20
`min_samples_leaf`	5
`min_samples_split`	22
`n_estimators`	417

Table 18: Optimal hyper-parameters for the tool classifier based on random forest.

STATISTIC	TRAINING SET	DEV SET	KTS	UTS
samples	955872	119484	119485	30144
accuracy [%]	95.716	94.977	94.953	7.680
balanced accuracy [%]	95.590	89.399	90.366	8.858
precision [%]	80.233	76.446	76.447	2.624
recall [%]	95.590	89.399	90.366	2.531
Cohen's kappa [%]	91.752	90.300	90.285	3.298
F-score [%]	85.925	81.341	81.477	2.577
Jaccard score [%]	77.432	71.180	71.442	1.572
Hamming loss	0.043	0.050	0.050	0.923
zero-one loss	0.043	0.050	0.050	0.923
R_k	0.919	0.904	0.904	0.041

Table 19: Classification statistics for the tool classifier based on random forest.

		INFERRED										
		CHROME	CURL	EDGE	FIREFOX	GOLDENEYE	HTTRACK	HULK	RUDY	SLOWLORIS	WGET	WPULL
TARGET	CHROME	1999	25	166	113	110	9	12	1	3	14	22
	CURL	1	273	13	3	4	11	0	1	1	12	4
	EDGE	73	16	2717	45	26	9	10	0	1	10	9
	FIREFOX	140	17	74	1588	71	41	17	3	2	16	29
	GOLDENEYE	1053	22	159	366	74967	150	1511	107	10	31	318
	HTTRACK	17	6	4	16	20	1229	2	0	0	0	4
	HULK	409	4	21	49	357	15	28464	20	0	9	93
	RUDY	3	0	0	1	7	2	7	321	2	0	5
	SLOWLORIS	0	2	1	9	0	20	0	1	1263	0	3
	WGET	2	4	3	3	12	0	1	1	0	449	7
	WPULL	2	3	4	2	10	4	0	0	0	3	184

Table 20: Confusion matrix for the tool classifier based on random forest on the KTS.

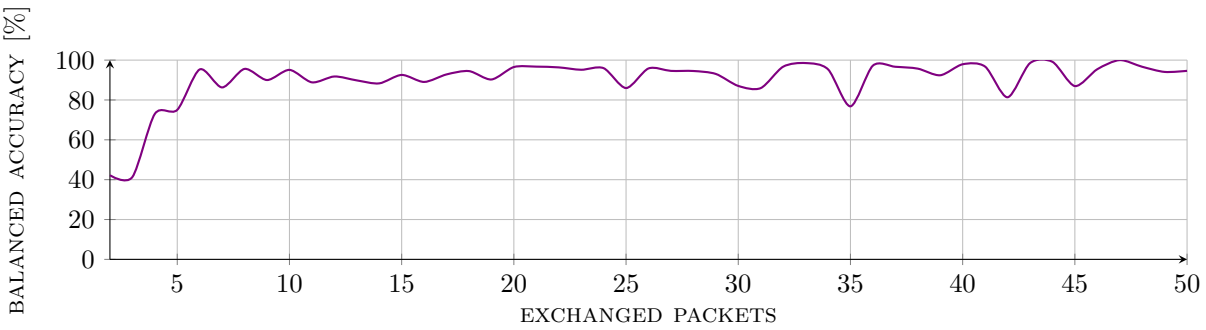

Figure 8: Balanced accuracy vs. exchange packets plot for the tool classifier based on random forest on the KTS.

INFERRED CLASS	SAMPLES
chrome	688
curl	6
edge	329
firefox	2315
goldeneye	405
httrack	113
hulk	55
rudy	37
slowloris	2
wget	53
wpull	2531

(a) Classification of FIREFOX-68.0.

INFERRED CLASS	SAMPLES
chrome	348
curl	41
edge	262
firefox	639
goldeneye	1440
httrack	186
hulk	185
rudy	26
slowloris	6
wget	207
wpull	325

(b) Classification of GRABSITE-2.1.16.

INFERRED CLASS	SAMPLES
chrome	5006
curl	78
edge	288
firefox	1013
goldeneye	2139
httrack	182
hulk	34
rudy	15
slowloris	3
wget	20
wpull	172

(c) Classification of OPERA-62.0.3331.66.

INFERRED CLASS	SAMPLES
chrome	175
curl	57
edge	1831
firefox	2334
goldeneye	1032
httrack	372
hulk	17
rudy	4022
slowloris	39
wget	39
wpull	1077

(d) Classification of SLOWHTTPTEST-1.6.

Table 21: Classification of unknown tools for the tool classifier based on random forest.

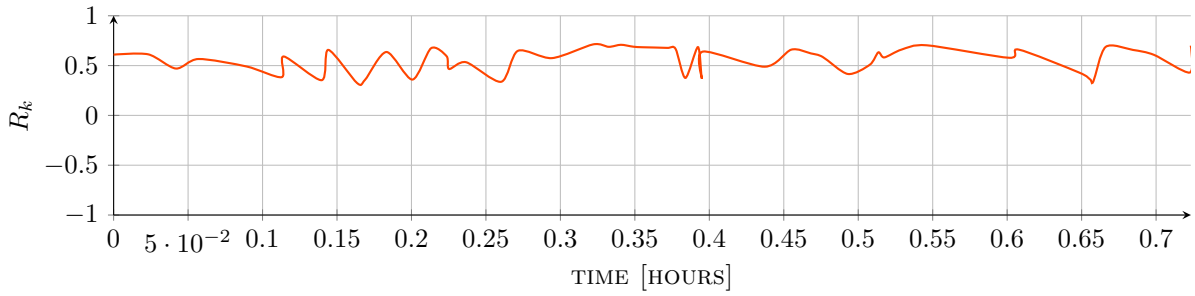

Figure 9: Hyper-parameters optimization plot for the tool classifier based on extra-trees.

HYPER-PARAMETER	VALUE
`criterion`	entropy
`max_depth`	20
`min_samples_leaf`	21
`min_samples_split`	49
`n_estimators`	96

Table 22: Optimal hyper-parameters for the tool classifier based on extra-trees.

STATISTIC	TRAINING SET	DEV SET	KTS	UTS
samples	955872	119484	119485	30144
accuracy [%]	83.376	83.256	83.226	4.668
balanced accuracy [%]	69.140	66.884	66.701	5.383
precision [%]	50.983	49.885	49.685	3.105
recall [%]	69.140	66.884	66.701	1.538
Cohen's kappa [%]	70.282	70.017	70.042	2.396
F-score [%]	52.837	51.522	51.257	2.057
Jaccard score [%]	39.406	38.294	38.123	1.202
Hamming loss	0.166	0.167	0.168	0.953
zero-one loss	0.166	0.167	0.168	0.953
R_k	0.713	0.710	0.711	0.030

Table 23: Classification statistics for the tool classifier based on extra-trees.

		INFERRED										
		CHROME	CURL	EDGE	FIREFOX	GOLDENEYE	HTTRACK	HULK	RUDY	SLOWLORIS	WGET	WPULL
TARGET	CHROME	1256	73	208	112	358	126	43	13	153	64	68
	CURL	0	179	22	0	41	38	2	1	26	7	7
	EDGE	124	35	2025	30	103	140	51	10	284	11	103
	FIREFOX	204	52	171	837	286	169	27	17	68	63	104
	GOLDENEYE	92	237	1084	57	66908	5511	1061	173	2759	458	354
	HTTRACK	17	15	15	14	122	978	1	1	94	13	28
	HULK	26	6	431	38	1496	1225	25393	15	641	20	150
	RUDY	0	2	4	0	32	19	5	249	33	2	2
	SLOWLORIS	1	0	2	0	0	35	0	0	1260	0	1
	WGET	4	18	19	0	40	51	4	3	68	255	20
	WPULL	1	5	8	1	36	38	0	6	12	3	102

Table 24: Confusion matrix for the tool classifier based on extra-trees on the KTS.

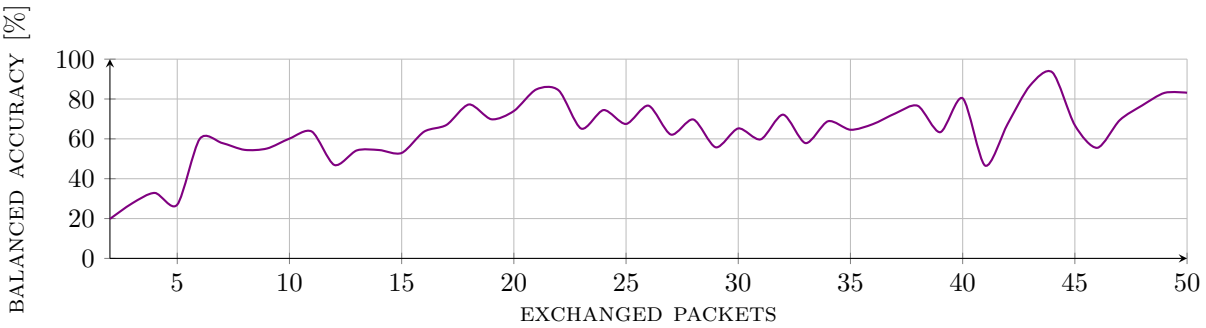

Figure 10: Balanced accuracy vs. exchange packets plot for the tool classifier based on extra-trees on the KTS.

INFERRED CLASS	SAMPLES
chrome	260
curl	350
edge	663
firefox	1407
goldeneye	495
httrack	585
hulk	2249
rudy	8
slowloris	41
wget	215
wpull	261

(a) Classification of FIREFOX-68.0.

INFERRED CLASS	SAMPLES
chrome	188
curl	142
edge	230
firefox	187
goldeneye	1097
httrack	468
hulk	32
rudy	33
slowloris	132
wget	293
wpull	863

(b) Classification of GRABSITE-2.1.16.

INFERRED CLASS	SAMPLES
chrome	3498
curl	412
edge	494
firefox	523
goldeneye	3009
httrack	354
hulk	39
rudy	147
slowloris	156
wget	66
wpull	252

(c) Classification of OPERA-62.0.3331.66.

INFERRED CLASS	SAMPLES
chrome	373
edge	1458
firefox	1120
goldeneye	2033
httrack	1598
hulk	61
rudy	3516
slowloris	463
wget	53
wpull	320

(d) Classification of SLOWHTTPTEST-1.6.

Table 25: Classification of unknown tools for the tool classifier based on extra-trees.

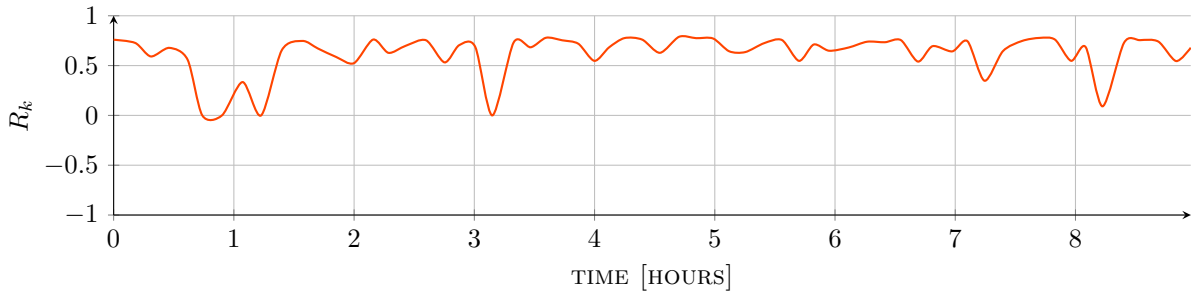

Figure 11: Hyper-parameters optimization plot for the tool classifier based on neural network.

HYPER-PARAMETER	VALUE
`lr`	0.0015639059764891423
`module__layers`	4
`module__neurons_per_layer`	196
`module__p`	0.34048448616373395

Table 26: Optimal hyper-parameters for the tool classifier based on neural network.

STATISTIC	TRAINING SET	DEV SET	KTS	UTS
samples	955872	119484	119485	30144
accuracy [%]	87.735	87.516	87.516	7.467
balanced accuracy [%]	81.721	79.971	80.332	8.613
precision [%]	53.469	52.862	52.688	2.854
recall [%]	81.721	79.971	80.332	2.461
Cohen's kappa [%]	77.740	77.340	77.379	3.560
F-score [%]	60.582	59.641	59.479	2.643
Jaccard score [%]	48.020	47.214	47.164	1.621
Hamming loss	0.123	0.125	0.125	0.925
zero-one loss	0.123	0.125	0.125	0.925
R_k	0.785	0.782	0.782	0.044

Table 27: Classification statistics for the tool classifier based on neural network.

		INFERRED										
		CHROME	CURL	EDGE	FIREFOX	GOLDENEYE	HTTRACK	HULK	RUDY	SLOWLORIS	WGET	WPULL
TARGET	CHROME	1509	50	204	300	214	42	35	4	2	75	39
	CURL	1	207	14	5	37	14	2	2	0	28	13
	EDGE	61	16	2552	66	36	28	38	10	1	51	57
	FIREFOX	165	19	119	1269	184	80	50	10	1	34	67
	GOLDENEYE	1189	313	821	562	67851	710	3873	1273	106	1176	820
	HTTRACK	13	1	37	38	98	1087	10	1	0	7	6
	HULK	367	4	53	48	488	115	27981	147	0	133	105
	RUDY	1	0	2	1	5	2	11	306	2	9	9
	SLOWLORIS	2	0	0	19	0	15	0	3	1258	1	1
	WGET	4	15	16	1	17	5	2	6	0	383	33
	WPULL	1	2	5	5	12	4	3	4	0	10	166

Table 28: Confusion matrix for the tool classifier based on neural network on the KTS.

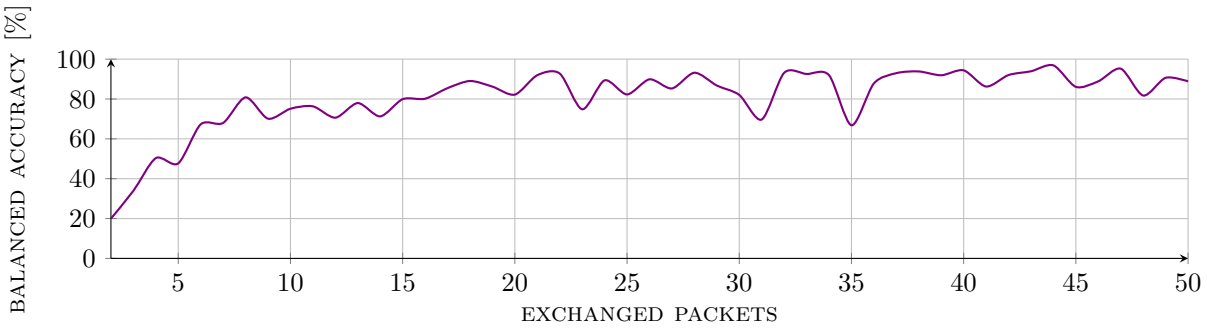

Figure 12: Balanced accuracy vs. exchange packets plot for the tool classifier based on neural network on the KTS.

INFERRED CLASS	SAMPLES
chrome	339
curl	0
edge	717
firefox	2251
goldeneye	2612
httrack	244
hulk	154
rudy	1
slowloris	0
wget	76
wpull	140

(a) Classification of FIREFOX-68.0.

INFERRED CLASS	SAMPLES
chrome	364
curl	13
edge	458
firefox	558
goldeneye	670
httrack	521
hulk	290
rudy	133
slowloris	0
wget	243
wpull	415

(b) Classification of GRABSITE-2.1.16.

INFERRED CLASS	SAMPLES
chrome	4097
curl	138
edge	134
firefox	1502
goldeneye	2342
httrack	367
hulk	80
rudy	29
slowloris	4
wget	59
wpull	198

(c) Classification of OPERA-62.0.3331.66.

INFERRED CLASS	SAMPLES
chrome	1831
curl	109
edge	704
firefox	1323
goldeneye	270
httrack	378
hulk	109
rudy	4144
slowloris	338
wget	640
wpull	1149

(d) Classification of SLOWHTTPTEST-1.6.

Table 29: Classification of unknown tools for the tool classifier based on neural network.

Tool instance classifiers

This section reports several statistics and plots about the models for classifying the traffic into tool instances (e.g. GOLDENEYE-2.1, FIREFOX-62.0, HULK-1.0, WGET-1.11.4, EDGE-42.17134.1.0, HTTRACK-3.49.2, CHROME-48.0.2564.109, RUDY-1.0.0, CHROME-68.0.3440.84, FIREFOX-42.0, SLOWLORIS-0.1.5, CURL-7.55.1, CURL-7.61.0, SLOWLORIS-0.1.4, WPULL-2.0.1 and WGET-1.19.5).

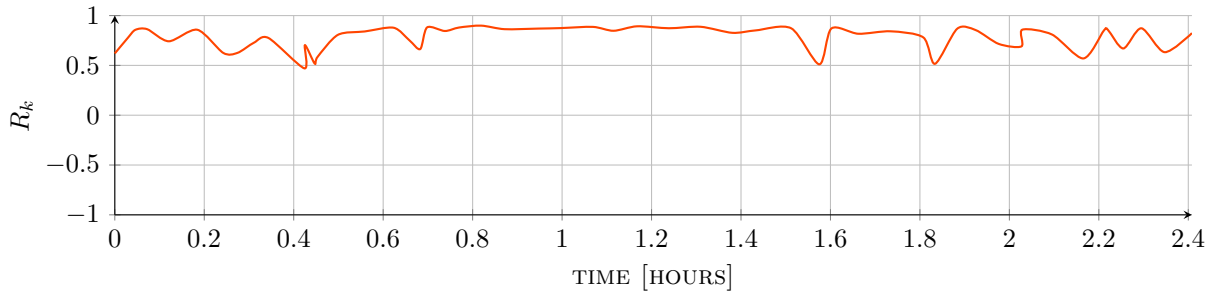

Figure 13: Hyper-parameters optimization plot for the tool instance classifier based on random forest.

HYPER-PARAMETER	VALUE
`criterion`	entropy
`max_depth`	20
`min_samples_leaf`	6
`min_samples_split`	18
`n_estimators`	314

Table 30: Optimal hyper-parameters for the tool instance classifier based on random forest.

STATISTIC	TRAINING SET	DEV SET	KTS	UTS
samples	955872	119484	119485	30144
accuracy [%]	95.565	94.642	94.601	0.000
balanced accuracy [%]	95.196	87.183	88.021	0.000
precision [%]	79.996	74.641	75.061	0.000
recall [%]	95.196	87.183	88.021	0.000
Cohen's kappa [%]	91.495	89.690	89.648	0.000
F-score [%]	85.628	79.254	79.707	0.000
Jaccard score [%]	76.831	68.057	68.871	0.000
Hamming loss	0.044	0.054	0.054	1.000
zero-one loss	0.044	0.054	0.054	1.000
R_k	0.916	0.898	0.898	0.000

Table 31: Classification statistics for the tool instance classifier based on random forest.

		INFERRED															
		CH-48.0	CH-68.0	CU-7.55.1	CU-7.61.0	ED-42	FI-42.0	FI-62.0	GO-2.1	HT-3.49.2	HU-1.0	RU-1.0.0	SL-0.1.4	SL-0.1.5	WG-1.11.4	WG-1.19.5	WP-2.0.1
TARGET	CH-48.0	1128	80	14	4	91	18	20	61	4	4	0	0	3	3	10	10
	CH-68.0	49	810	7	1	36	18	33	53	6	4	0	0	0	0	1	6
	CU-7.55.1	0	1	150	6	0	0	3	4	5	0	0	0	1	2	0	0
	CU-7.61.0	0	0	2	122	9	2	0	0	4	0	1	0	0	2	6	3
	ED-42	75	37	2	21	2656	32	19	35	9	10	0	0	2	4	6	8
	FI-42.0	26	22	9	2	41	646	51	34	12	5	2	0	2	4	3	10
	FI-62.0	27	45	9	1	28	46	894	37	17	6	0	0	3	0	5	11
	GO-2.1	114	1191	50	2	157	74	320	74802	130	1459	93	4	5	18	10	265
	HT-3.49.2	4	15	7	1	7	5	6	23	1224	2	0	0	1	0	0	3
	HU-1.0	35	485	11	0	21	16	53	293	12	28435	18	0	3	6	0	53
	RU-1.0.0	3	1	0	1	0	1	0	6	3	7	321	1	0	0	0	4
	SL-0.1.4	0	0	0	0	0	0	0	0	0	0	1	607	6	0	0	0
	SL-0.1.5	0	0	4	0	1	6	1	0	13	0	1	48	608	0	0	3
	WG-1.11.4	2	1	1	2	4	3	0	8	0	1	1	0	0	259	0	1
	WG-1.19.5	1	0	0	1	0	0	0	0	0	0	0	0	0	0	196	1
	WP-2.0.1	1	5	1	5	4	4	2	6	3	1	0	0	1	3	0	176

Table 32: Confusion matrix for the tool instance classifier based on random forest on the KTS (where GO = GOLDENEYE, FI = FIREFOX, HU = HULK, WG = WGET, ED = EDGE, HT = HTTRACK, CH = CHROME, RU = RUDY, SL = SLOWLORIS, CU = CURL and WP = WPULL.

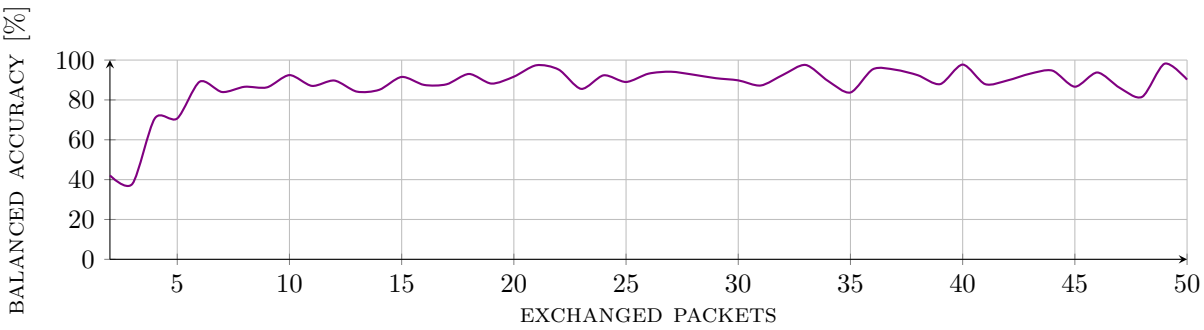

Figure 14: Balanced accuracy vs. exchange packets plot for the tool instance classifier based on random forest on the KTS.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	367
chrome-68.0.3440.84	226
curl-7.55.1	9
edge-42.17134.1.0	356
firefox-42.0	607
firefox-62.0	1805
goldeneye-2.1	377
httrack-3.49.2	47
hulk-1.0	49
rudy-1.0.0	33
slowloris-0.1.5	52
wget-1.19.5	60
wpull-2.0.1	2546

(a) Classification of FIREFOX-68.0.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	112
chrome-68.0.3440.84	168
curl-7.55.1	53
curl-7.61.0	2
edge-42.17134.1.0	260
firefox-42.0	224
firefox-62.0	296
goldeneye-2.1	1579
httrack-3.49.2	164
hulk-1.0	223
rudy-1.0.0	25
slowloris-0.1.5	11
wget-1.11.4	84
wget-1.19.5	23
wpull-2.0.1	441

(b) Classification of GRABSITE-2.1.16.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	2365
chrome-68.0.3440.84	2178
curl-7.55.1	103
curl-7.61.0	5
edge-42.17134.1.0	382
firefox-42.0	275
firefox-62.0	714
goldeneye-2.1	2431
httrack-3.49.2	231
hulk-1.0	35
rudy-1.0.0	18
slowloris-0.1.5	6
wget-1.11.4	12
wget-1.19.5	5
wpull-2.0.1	190

(c) Classification of OPERA-62.0.3331.66.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	165
chrome-68.0.3440.84	27
curl-7.55.1	78
curl-7.61.0	22
edge-42.17134.1.0	1795
firefox-42.0	394
firefox-62.0	1220
goldeneye-2.1	1212
httrack-3.49.2	341
hulk-1.0	33
rudy-1.0.0	4061
slowloris-0.1.4	1
slowloris-0.1.5	56
wget-1.11.4	25
wget-1.19.5	8
wpull-2.0.1	1557

(d) Classification of SLOWHTTPTEST-1.6.

Table 33: Classification of unknown tools for the tool instance classifier based on random forest.

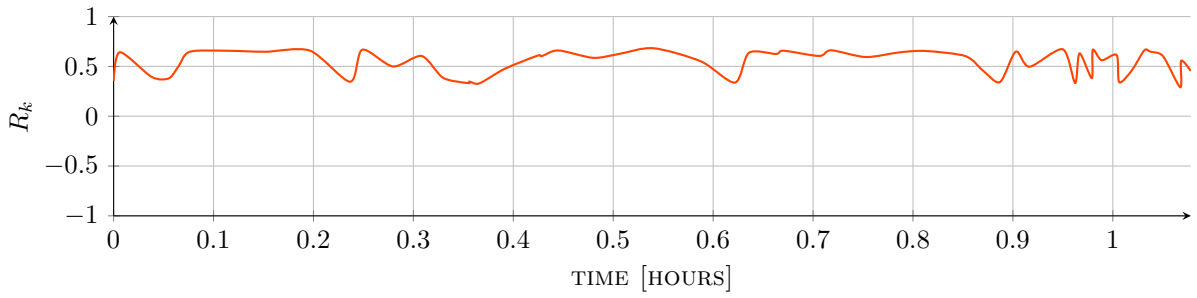

Figure 15: Hyper-parameters optimization plot for the tool instance classifier based on extra-trees.

HYPER-PARAMETER	VALUE
`criterion`	gini
`max_depth`	20
`min_samples_leaf`	8
`min_samples_split`	18
`n_estimators`	417

Table 34: Optimal hyper-parameters for the tool instance classifier based on extra-trees.

STATISTIC	TRAINING SET	DEV SET	KTS	UTS
samples	955872	119484	119485	30144
accuracy [%]	80.889	80.672	80.644	0.000
balanced accuracy [%]	66.162	63.546	62.533	0.000
precision [%]	44.696	42.762	42.506	0.000
recall [%]	66.162	63.546	62.533	0.000
Cohen's kappa [%]	66.786	66.362	66.387	0.000
F-score [%]	45.703	43.713	43.182	0.000
Jaccard score [%]	33.113	31.442	31.131	0.000
Hamming loss	0.191	0.193	0.194	1.000
zero-one loss	0.191	0.193	0.194	1.000
R_k	0.681	0.677	0.677	0.000

Table 35: Classification statistics for the tool instance classifier based on extra-trees.

TARGET	INFERRED															
	CH-48.0	CH-68.0	CU-7.55.1	CU-7.61.0	ED-42	FI-42.0	FI-62.0	GO-2.1	HT-3.49.2	HU-1.0	RU-1.0.0	SL-0.1.4	SL-0.1.5	WG-1.11.4	WG-1.19.5	WP-2.0.1
CH-48.0	721	89	62	5	44	20	23	180	66	0	9	98	48	26	39	20
CH-68.0	92	484	22	4	85	26	20	151	45	32	0	2	6	18	20	17
CU-7.55.1	0	0	98	5	2	0	0	34	8	2	1	2	7	13	0	0
CU-7.61.0	0	0	2	86	12	0	0	0	18	0	0	13	10	3	3	4
ED-42	52	93	13	58	1950	17	11	95	133	47	2	209	126	31	23	56
FI-42.0	45	32	31	12	80	268	95	80	72	3	7	23	42	26	20	33
FI-62.0	70	34	32	6	46	70	498	184	65	3	5	2	20	26	39	29
GO-2.1	59	72	346	366	901	11	179	64648	5382	767	24	2364	1052	1754	660	109
HT-3.49.2	8	5	14	1	11	2	4	118	990	1	0	6	73	32	4	29
HU-1.0	2	42	140	52	457	3	171	1333	1129	24964	0	612	52	43	296	145
RU-1.0.0	0	0	1	2	6	0	1	29	17	3	236	26	13	10	4	0
SL-0.1.4	0	0	0	0	0	0	0	0	0	0	0	525	89	0	0	0
SL-0.1.5	0	0	1	0	0	0	0	0	25	0	0	205	453	0	0	1
WG-1.11.4	1	1	1	6	5	0	0	22	11	0	0	5	3	212	6	10
WG-1.19.5	0	0	1	5	3	0	0	0	21	3	0	30	0	2	132	2
WP-2.0.1	2	0	3	9	8	0	1	35	33	0	4	4	7	12	1	93

Table 36: Confusion matrix for the tool instance classifier based on extra-trees on the KTS (where GO = GOLDENEYE, FI = FIREFOX, HU = HULK, WG = WGET, ED = EDGE, HT = HTTRACK, CH = CHROME, RU = RUDY, SL = SLOWLORIS, CU = CURL and WP = WPULL.

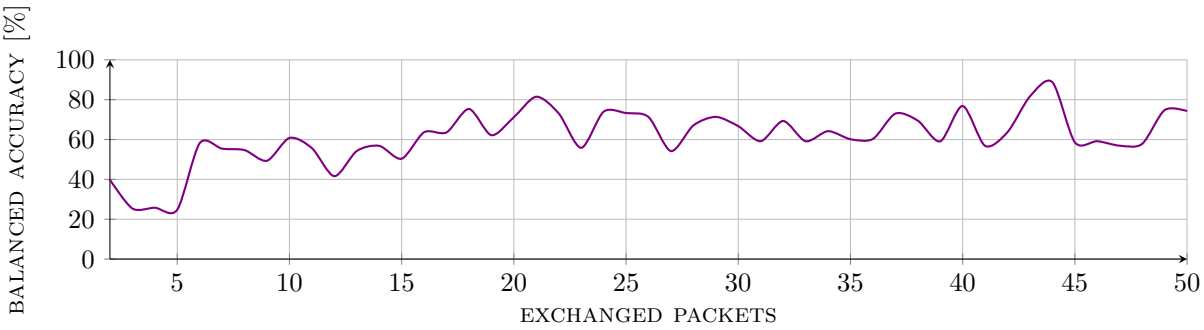

Figure 16: Balanced accuracy vs. exchange packets plot for the tool instance classifier based on extra-trees on the KTS.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	450
chrome-68.0.3440.84	14
curl-7.55.1	183
curl-7.61.0	58
edge-42.17134.1.0	573
firefox-42.0	342
firefox-62.0	1026
goldeneye-2.1	472
httrack-3.49.2	596
hulk-1.0	2241
rudy-1.0.0	22
slowloris-0.1.4	2
slowloris-0.1.5	23
wget-1.11.4	49
wget-1.19.5	260
wpull-2.0.1	223

(a) Classification of FIREFOX-68.0.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	148
chrome-68.0.3440.84	24
curl-7.55.1	80
curl-7.61.0	42
edge-42.17134.1.0	217
firefox-42.0	46
firefox-62.0	156
goldeneye-2.1	1039
httrack-3.49.2	404
hulk-1.0	22
rudy-1.0.0	20
slowloris-0.1.4	97
slowloris-0.1.5	123
wget-1.11.4	198
wget-1.19.5	170
wpull-2.0.1	879

(b) Classification of GRABSITE-2.1.16.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	1913
chrome-68.0.3440.84	1672
curl-7.55.1	503
curl-7.61.0	22
edge-42.17134.1.0	432
firefox-42.0	192
firefox-62.0	220
goldeneye-2.1	2978
httrack-3.49.2	341
hulk-1.0	41
rudy-1.0.0	166
slowloris-0.1.4	19
slowloris-0.1.5	80
wget-1.11.4	92
wget-1.19.5	81
wpull-2.0.1	198

(c) Classification of OPERA-62.0.3331.66.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	271
chrome-68.0.3440.84	214
curl-7.55.1	6
curl-7.61.0	19
edge-42.17134.1.0	1473
firefox-42.0	211
firefox-62.0	581
goldeneye-2.1	1965
httrack-3.49.2	920
hulk-1.0	32
rudy-1.0.0	2674
slowloris-0.1.4	160
slowloris-0.1.5	1592
wget-1.11.4	226
wget-1.19.5	13
wpull-2.0.1	638

(d) Classification of SLOWHTTPTEST-1.6.

Table 37: Classification of unknown tools for the tool instance classifier based on extra-trees.

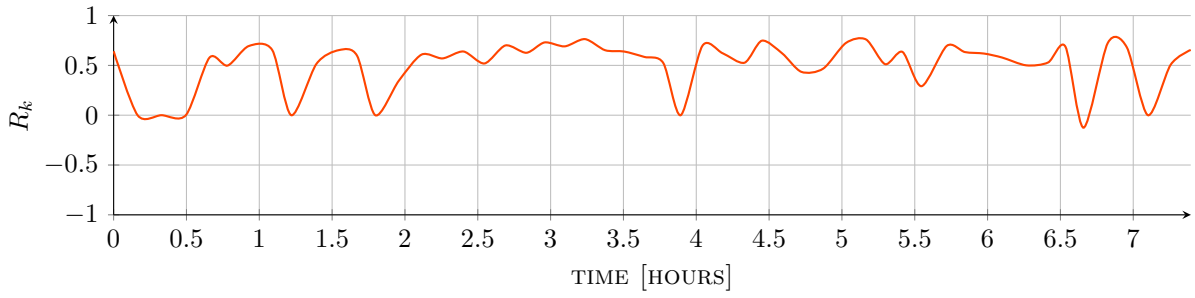

Figure 17: Hyper-parameters optimization plot for the tool instance classifier based on neural network.

HYPER-PARAMETER	VALUE
`lr`	0.001044660236833224
`module__layers`	4
`module__neurons_per_layer`	478
`module__p`	0.33926635120188525

Table 38: Optimal hyper-parameters for the tool instance classifier based on neural network.

STATISTIC	TRAINING SET	DEV SET	KTS	UTS
samples	955872	119484	119485	30144
accuracy [%]	83.050	82.687	82.649	0.000
balanced accuracy [%]	79.137	76.554	76.767	0.000
precision [%]	45.881	44.808	44.725	0.000
recall [%]	79.137	76.554	76.767	0.000
Cohen's kappa [%]	70.892	70.285	70.290	0.000
F-score [%]	50.890	49.582	49.411	0.000
Jaccard score [%]	38.186	37.084	36.897	0.000
Hamming loss	0.170	0.173	0.174	1.000
zero-one loss	0.170	0.173	0.174	1.000
R_k	0.724	0.719	0.719	0.000

Table 39: Classification statistics for the tool instance classifier based on neural network.

		INFERRED															
		CH-48.0	CH-68.0	CU-7.55.1	CU-7.61.0	ED-42	FI-42.0	FI-62.0	GO-2.1	HT-3.49.2	HU-1.0	RU-1.0.0	SL-0.1.4	SL-0.1.5	WG-1.11.4	WG-1.19.5	WP-2.0.1
TARGET	CH-48.0	832	178	26	14	91	57	78	43	16	6	22	0	2	8	41	36
	CH-68.0	60	716	25	1	40	23	80	45	7	8	1	0	2	3	1	12
	CU-7.55.1	0	2	133	5	0	1	1	6	11	0	0	0	1	6	0	6
	CU-7.61.0	0	0	0	121	3	1	0	1	6	0	2	0	0	3	12	2
	ED-42	38	77	2	117	2395	30	40	19	39	33	10	0	1	14	62	39
	FI-42.0	25	55	5	16	47	437	155	19	56	5	6	0	4	4	7	28
	FI-62.0	18	98	36	2	28	68	752	39	35	6	2	0	6	5	11	23
	GO-2.1	132	1922	2542	50	598	123	1033	63582	1334	2713	1230	5	45	993	719	1673
	HT-3.49.2	5	17	22	1	13	7	24	37	1150	1	1	0	4	5	3	8
	HU-1.0	2	573	76	15	60	20	474	264	123	26812	287	0	2	23	431	279
	RU-1.0.0	2	0	1	0	1	1	0	1	5	9	312	2	0	7	0	7
	SL-0.1.4	0	0	0	0	0	0	0	0	0	0	2	538	74	0	0	0
	SL-0.1.5	0	0	0	1	0	5	0	0	27	0	1	248	402	0	1	0
	WG-1.11.4	0	2	2	7	2	1	1	13	8	0	5	0	0	228	2	12
	WG-1.19.5	0	0	0	6	4	0	0	1	0	0	0	0	0	0	185	3
	WP-2.0.1	1	8	2	6	5	3	0	3	9	0	5	0	1	8	3	158

Table 40: Confusion matrix for the tool instance classifier based on neural network on the KTS (where GO = GOLDENEYE, FI = FIREFOX, HU = HULK, WG = WGET, ED = EDGE, HT = HTTRACK, CH = CHROME, RU = RUDY, SL = SLOWLORIS, CU = CURL and WP = WPULL.

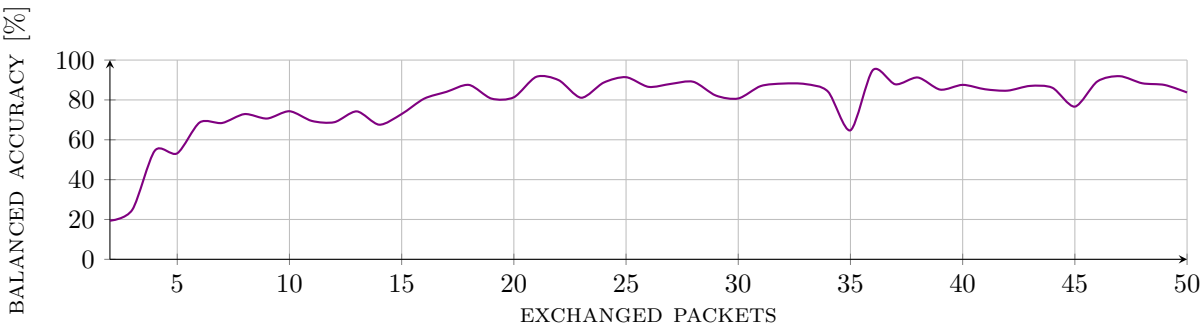

Figure 18: Balanced accuracy vs. exchange packets plot for the tool instance classifier based on neural network on the KTS.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	142
chrome-68.0.3440.84	336
curl-7.55.1	62
curl-7.61.0	3
edge-42.17134.1.0	454
firefox-42.0	660
firefox-62.0	1651
goldeneye-2.1	2386
httrack-3.49.2	351
hulk-1.0	62
rudy-1.0.0	60
slowloris-0.1.4	0
slowloris-0.1.5	75
wget-1.11.4	5
wget-1.19.5	96
wpull-2.0.1	191

(a) Classification of FIREFOX-68.0.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	210
chrome-68.0.3440.84	321
curl-7.55.1	47
curl-7.61.0	19
edge-42.17134.1.0	673
firefox-42.0	139
firefox-62.0	439
goldeneye-2.1	480
httrack-3.49.2	312
hulk-1.0	121
rudy-1.0.0	240
slowloris-0.1.4	0
slowloris-0.1.5	9
wget-1.11.4	99
wget-1.19.5	201
wpull-2.0.1	355

(b) Classification of GRABSITE-2.1.16.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	1636
chrome-68.0.3440.84	3446
curl-7.55.1	509
curl-7.61.0	8
edge-42.17134.1.0	75
firefox-42.0	281
firefox-62.0	1091
goldeneye-2.1	1292
httrack-3.49.2	322
hulk-1.0	26
rudy-1.0.0	20
slowloris-0.1.4	0
slowloris-0.1.5	7
wget-1.11.4	45
wget-1.19.5	28
wpull-2.0.1	164

(c) Classification of OPERA-62.0.3331.66.

INFERRED CLASS	SAMPLES
chrome-48.0.2564.109	920
chrome-68.0.3440.84	239
curl-7.55.1	33
curl-7.61.0	71
edge-42.17134.1.0	1022
firefox-42.0	473
firefox-62.0	1262
goldeneye-2.1	100
httrack-3.49.2	735
hulk-1.0	61
rudy-1.0.0	4359
slowloris-0.1.4	9
slowloris-0.1.5	3
wget-1.11.4	405
wget-1.19.5	129
wpull-2.0.1	1174

(d) Classification of SLOWHTTPTEST-1.6.

Table 41: Classification of unknown tools for the tool instance classifier based on neural network.