

Supplementary Tables, and Algorithms

Explaining Cross-Dataset False Positives in Encrypted Traffic Classification for Forensic Interpretation

TABLE I
DATASETS USED IN THIS STUDY

Dataset	Role	Benign	Malicious
CIRA-CIC-DoHBrw-2020	Training and internal testing	917,300	249,969
CESNET-TLS-Year22	External false positive evaluation	947,997	0

TABLE II
SELECTED FEATURES AND DESCRIPTIONS

Feature	Description
Flow Duration	Total duration of a flow
Byte Count	Total bytes transferred
Packet Count	Number of packets in a flow
Packet Size Mean	Average packet size in a flow
Inter-Arrival Time Mean	Average time between packets

TABLE III
CIRA-CIC-DOHBRW-2020 INTERNAL DATASET SPLIT DISTRIBUTION

Split	Flows	Percentage
Training	817,071	70.0%
Validation	175,544	15.0%
Test	174,654	15.0%
Total	1,167,269	100.0%

TABLE IV
EVALUATION DESIGN

Phase	Dataset	Purpose	Evaluation
Training	CIRA train	Learn traffic patterns	Supervised learning
Validation	CIRA validation	Select model	Model selection
Internal testing	CIRA test	Measure detection performance	Standard metrics
External testing	CESNET TLS	Measure false positives	Cross-dataset test

TABLE V
INTERNAL MODEL PERFORMANCE ON THE CIRA TEST DATASET

Model	Acc.	AUC	Prec.	Recall	F1
Random Forest	0.9962	0.9995	0.9929	0.9895	0.9912
XGBoost	0.9935	0.9996	0.9773	0.9930	0.9851
LightGBM	0.9902	0.9993	0.9649	0.9903	0.9774
CatBoost	0.9892	0.9989	0.9621	0.9886	0.9752

TABLE VI
CROSS-DATASET EVALUATION ON THE CESNET TLS DATASET

Metric	Value
Total benign flows	947,997
False positives	41,238
False positive rate	4.35%
95% confidence interval	4.31–4.39%

Algorithm 1 Cross-dataset false positive analysis

Require: CIRA-CIC-DoHBrw-2020 dataset, CESNET-TLS-Year22 dataset

Ensure: False positive rate and feature-level explanation

- 1: Extract the aligned flow-level features from both datasets
- 2: Map CIRA labels into benign and malicious classes
- 3: Treat selected CESNET flows as benign external traffic
- 4: Train the classifier using the CIRA training data
- 5: Apply the trained classifier to the CESNET dataset
- 6: Count each malicious prediction on CESNET as a false positive
- 7: Compute the false positive rate using $FPR = FP/(FP + TN)$
- 8: Apply SHAP to explain feature contributions for predictions
- 9: Compare SHAP findings with permutation importance
- 10: Interpret false positive alerts as decision support for forensic review