Drift-Adaptive Intrusion Detection for Enterprise Networks

Dheeraj Ramasahayam

***Abstract*—Enterprise intrusion detection must balance strong in-domain accuracy against robustness to distribution shift in live traffic. This paper presents a reproducible benchmark that compares a transparent flow-signature IDS, Random Forest, LSTM, Transformer, a static Drift-Aware Hybrid, and an online Drift-Adaptive Hybrid across official UNSW-NB15 and NSL-KDD splits plus external transfer to CICIDS2017 and CSE-CIC-IDS2018. The static hybrid is strongest on UNSW-NB15 with weighted F1 90.72%, while LSTM leads NSL-KDD with 81.01% and full-corpus CICIDS2017 transfer with 71.53%. Under the same CICIDS2017 stream, the online Drift-Adaptive Hybrid improves the static hybrid from 61.35% to 68.69% weighted F1 without retraining base detectors. A drift-detector comparison shows Isolation Forest is the strongest controller signal, reaching 70.58% post-adaptation weighted F1 with zero source-domain false positives. Results on CSE-CIC-IDS2018 and family-level failure analysis show that cross-dataset generalization remains fragile, making adaptation, reproducibility, and deployment-oriented evaluation essential for enterprise IDS design.**

***Index Terms*—intrusion detection, cybersecurity, concept drift, online adaptation, deep learning, benchmark, network traffic, transfer learning**

I. Introduction

Network intrusion detection systems (IDSs) remain central to enterprise defense because perimeter controls alone do not prevent compromise, lateral movement, or data exfiltration. Classical signature-based systems such as Snort prioritize precision and speed for known patterns, but novel attack behaviors and distribution shift continue to challenge rule maintenance and coverage [1]. In parallel, machine learning and deep learning have become standard tools for anomaly and attack detection over labeled network telemetry [2]. However, a recurring weakness in the literature is that many studies stop at single-dataset evaluation and do not pair accuracy with reproducibility, external transfer, failure analysis, or deployment-oriented latency analysis. A second weakness is that models are often evaluated as static objects even though concept drift is a known operational property of live data streams [3].

This paper addresses that gap with a repo-backed benchmark that compares one traditional rule baseline, three established learning models, a static hybrid stack, and an online drift-adaptive controller across official splits and two external transfer corpora. The study is grounded in public datasets that are widely used in intrusion-detection research: UNSW-NB15 [4], NSL-KDD [5], CICIDS2017 [6], and a cleaned CSE-CIC-IDS2018 follow-on corpus. All data are mapped into a shared 41-feature flow representation so that the comparison remains fair across datasets and model families.

The contributions of this paper are sevenfold:

- A transparent flow-signature IDS baseline that serves as a reproducible traditional IDS comparator on public flow datasets.
- A static Drift-Aware Hybrid detector that fuses signature scores, Random Forest probabilities, LSTM probabilities, Transformer probabilities, and an Isolation Forest drift signal.
- An online Drift-Adaptive controller that derives stable and stressed ensemble regimes offline, then interpolates between them during deployment using a streaming drift estimate.
- Full official-split evaluation on UNSW-NB15 and NSL-KDD, plus external cross-dataset transfer from UNSW-NB15 and NSL-KDD into both CICIDS2017 and CSE-CIC-IDS2018.
- A formal drift-detector comparison that evaluates Isolation Forest, ADWIN, DDM, and Page-Hinkley by detection delay, false positives, and post-adaptation IDS quality.
- A family-level failure analysis over the dominant CICIDS2017 attack types, showing where transfer fails and where adaptation still recovers signal.
- Deployment-oriented analysis through latency-under-load, real-time streaming evaluation, a replayed packet-capture case study, explainability-by-ablation, and a production deployment architecture tied to Kafka-style flow ingestion and SIEM/SOC handoff.

II. Related Work

UNSW-NB15 was introduced to provide a more modern alternative to KDD-era benchmarks and remains a common benchmark for binary and multi-class intrusion detection [4]. NSL-KDD was proposed to correct redundancy and evaluation issues in KDD Cup 99 and is still useful as a historical baseline with official train and test partitions [5]. CICIDS2017 broadened the field with a large flow dataset spanning modern attacks such as DDoS, botnet activity, and web attacks [6]. CSE-CIC-IDS2018 continues that CIC lineage with newer attack scenarios and is therefore useful here as a second external transfer target rather than another in-domain benchmark.

Among learning methods, Random Forest remains a strong baseline for tabular intrusion features and has shown competitive performance on UNSW-NB15 [2]. LSTM is a natural sequential baseline because it can model ordered feature

Dheeraj Ramasahayam is an Independent Researcher based in Durham, North Carolina, USA.

TABLE I
DATASETS AND EVALUATION PROTOCOL.

Dataset	Train	Test	Protocol
UNSW-NB15	82,332	175,341	Full official split
NSL-KDD	125,973	22,544	Full official split
UNSW+NSL → CICIDS2017	208,305	2,830,743	Full external corpus
UNSW+NSL → CSE-CIC-IDS2018	208,305	505,156	Cleaned external corpus

interactions and temporal-style dependencies [7]. Transformer models provide a stronger attention-based alternative for long-range feature interaction modeling [8]. Recent work by Yan et al. reported 89.00% F1 on UNSW-NB15 with a specialized Transformer and dynamic semantic embedding [9], while Xin and Xu reported 55.00% F1 for a recent cross-dataset Transformer-IDS in the NSL→UNSW transfer setting [10]. Graph-based IDS work is also advancing rapidly: Wang et al. report binary F1 scores above 99% for BS-GAT on recent edge/IoT datasets [11]. The present work targets a different niche: a unified 41-feature benchmark that emphasizes external transfer, rule-based comparators, and deployment-time adaptation rather than peak single-dataset accuracy alone. Concept drift adaptation is also well established as a requirement for live data systems [3], yet classical detectors such as ADWIN [12], DDM [13], and Page-Hinkley [14] are still rarely compared directly inside IDS adaptation pipelines. In this work, the model families themselves are not claimed as novel; the paper's novelty comes from pairing them with a reproducible online adaptation layer and a stronger transfer-oriented evaluation protocol.

III. Datasets and Experimental Protocol

Table I summarizes the data used in this study. The benchmark uses the full official train/test files shipped with UNSW-NB15 and NSL-KDD. For external transfer, the UNSW-NB15 and NSL-KDD training data are combined into a joint source-domain training set and evaluated on the full 2,830,743-row CICIDS2017 corpus and on a cleaned 505,156-row CSE-CIC-IDS2018 corpus. The local CSE copy contained 59 repeated header rows, which were removed before evaluation. These additions remove the earlier sampled-holdout caveat and make the transfer results more representative than a single small external slice.

All datasets are mapped to a binary target of `Benign` versus `Attack`. The canonical feature schema contains 41 flow-level indicators including duration, packet counts, byte counts, inter-arrival statistics, packet-length moments, TCP flag counts, and initial window sizes. This common representation avoids giving any model access to dataset-specific metadata that would make cross-dataset evaluation less interpretable.

IV. Methods

A. Traditional IDS Baseline

The public datasets used in this study are distributed primarily as labeled flow records rather than raw packet payload traces. Because direct packet replay through Snort-style signatures is therefore not reproducible from the released benchmark files alone, this paper implements a transparent flow-signature IDS baseline. The baseline calibrates benign percentiles on the training set and fires explicit rules for volumetric floods, SYN-oriented scanning, asymmetric request bursts, flag anomalies, microbursts, and handshake failures. The purpose of this baseline is not to reproduce a specific commercial ruleset, but to preserve the classical rule-based detection paradigm under the same public feature constraints as the learned models.

B. Established Learning Models

Random Forest serves as the classical tabular baseline because it is strong on mixed-scale engineered features and efficient at inference time [15]. LSTM is used as a recurrent deep baseline [7]. Transformer is used as an attention-based deep baseline that can model non-local feature interactions through self-attention [8]. All deep models are trained with weighted sampling and weighted cross-entropy to reduce bias from class imbalance.

C. Proposed Drift-Adaptive Hybrid

The proposed method contains two layers. First, a static Drift-Aware Hybrid forms a second-stage meta-classifier over the following signals:

- flow-signature IDS attack probability,
- Random Forest attack probability,
- LSTM attack probability,
- Transformer attack probability,
- a normalized drift score from Isolation Forest.

The static hybrid is trained on validation data and learns when to trust deterministic rules, when to trust fast tabular classification, and when to defer to deep models under shift. On top of that, an online adaptation controller derives two ensemble regimes: a stable regime from the original validation distribution and a stressed regime from synthetically drifted validation batches. During deployment, the controller tracks an exponential moving average of the Isolation Forest drift score and interpolates between those regimes. As drift grows, the controller gradually reduces reliance on the more brittle tabular branch and increases reliance on the temporal branches and the stacked meta-classifier. This design is motivated by the observation that no single family dominates all three evaluation settings in this benchmark.

D. Drift-Adaptive Ensemble Formulation

Let $M_i(x_t)$ denote the ith detector's attack probability for flow x_t, with $K = 5$ component detectors in the hybrid. The deployed ensemble prediction is

$$p(x_t) = \sum_{i=1}^{K} w_i(t) M_i(x_t). \tag{1}$$

The online controller interpolates between two offline regimes:

$$\tilde{w}_i(t) = (1 - \alpha_t) w_i^{\text{stable}} + \alpha_t w_i^{\text{stress}}, \tag{2}$$

$$w_i(t) = \frac{\tilde{w}_i(t)}{\sum_{j=1}^{K} \tilde{w}_j(t)}. \tag{3}$$

Here $\alpha_t \in [0, 1]$ is the drift-controlled interpolation coefficient,

$$\alpha_t = \text{clip}\left(\frac{d_t - \tau_s}{\tau_e - \tau_s}, 0, 1\right), \tag{4}$$

where d_t is the current drift score and τ_s, τ_e are the adaptation start and end thresholds. In stable regimes, the controller favors the tabular and stacked branches. In stressed regimes, it shifts mass toward the temporal models and disagreement-aware boosting.

E. Formal Drift Detector Comparison

The deployed system uses a label-free Isolation Forest drift score, but that design choice should be justified empirically rather than assumed. This paper therefore compares Isolation Forest against ADWIN, DDM, and Page-Hinkley under the same controller and windowed transfer stream. ADWIN, DDM, and Page-Hinkley consume the static hybrid's window error rate and are therefore label-dependent post-hoc detectors; Isolation Forest instead operates directly on the feature stream and remains deployable when labels are delayed or unavailable. The comparison reports detection delay, source-domain false positives, and final post-adaptation IDS quality.

F. Production Deployment Scenario

Figure 2 translates the benchmark into an enterprise deployment path. Mirrored traffic is converted into flow records, ingested through Kafka or file-backed streams, aligned to the shared 41-feature schema, scored by the Drift-Adaptive Hybrid, and forwarded to a security alerting layer before SIEM/SOC triage. The measured full-stream controller latency of 0.2630 ms/flow on CICIDS2017 indicates that the adaptation layer is compatible with practical low-latency deployment.

V. Experimental Results

Table II reports the official split results and the primary full-corpus CICIDS2017 transfer result. On the official UNSW-NB15 split, the static Drift-Aware Hybrid achieves the highest weighted F1 score of 90.72%, narrowly beating Random Forest at 90.65%. On the official NSL-KDD split, LSTM achieves the best weighted F1 score of 81.01%, with Transformer close behind at 80.52%. On full cross-dataset transfer into CICIDS2017, LSTM leads with a weighted F1 score of 71.53%, while the online Drift-Adaptive Hybrid improves substantially over its own static counterpart.

Table IV adds a second modern external dataset and materially changes the interpretation of robustness. On the cleaned CSE-CIC-IDS2018 corpus, the transparent Signature IDS baseline becomes the strongest transfer model at 66.37% weighted F1, while the learned models collapse under the combined distribution shift and attack-heavy class prior. This result is intentionally sobering: the current source-domain training mixture and light-weight online adaptation are not sufficient to guarantee generalization to every modern external corpus.

Table V answers a likely reviewer question about the detector choice. Isolation Forest is not just convenient; it is the strongest detector in this benchmark, achieving 70.58% weighted F1 after adaptation with zero source-domain false positives. DDM and Page-Hinkley react immediately but finish slightly lower at 69.77%, while ADWIN never triggers in this setting. This also matters operationally because Isolation Forest is label-free, whereas the error-stream detectors require post-label feedback.

Table VI provides the requested failure analysis. The source-trained LSTM collapses entirely on the six most frequent CICIDS2017 attack families under a family-versus-benign protocol, while the adaptive hybrid only recovers meaningful signal on FTP-Patator and a smaller amount on SSH-Patator. The result suggests that the adaptation layer helps most when attack families still preserve explicit flag, count, or asymmetry cues that the hybrid can exploit, but it does not fully solve severe family-level covariate shift.

Table VII places the current benchmark alongside representative recent papers. The main message is not that one number dominates every setting; it is that this repository now situates its contribution relative to recent Transformer and graph-based IDS work while adding a fuller deployment story through formal drift comparison, two external corpora, and real-time streaming evaluation.

The latency results still reveal a clear speed hierarchy. On UNSW-NB15 at batch size 1,024, the Signature IDS baseline exceeds 1.2 million flows per second, Random Forest reaches roughly 71 thousand flows per second, LSTM reaches about 13.4 thousand flows per second, Transformer reaches about 11.8 thousand flows per second, and the hybrid model reaches about 5.2 thousand flows per second. The full-corpus online adaptation study shows that deployment-time adaptation can recover 7.34 weighted F1 points over the static hybrid on the external stream without touching the already-trained base detectors, while increasing average hybrid latency from 0.1240 to 0.2630 ms/flow. In a separate file-backed real-time emulation over 29 sequential windows, the same online controller preserved its aggregate weighted F1 of 68.69%, achieved a mean window F1 of 73.80%, and operated with a mean adaptation alpha of 0.9985, indicating that the external stream almost immediately pushes the model into its stressed regime. The tradeoff is therefore favorable for enterprise settings that can tolerate sub-millisecond model inference in exchange for better robustness under shift.

Table VIII adds the requested real-world-style validation. The case study replays a local packet capture collected from a controlled internal lab environment simulating enterprise traffic, containing benign warm-up traffic, a sustained internal service attack from 192.168.76.9 to 192.168.12.56:3000/UDP, and benign cool-down traffic. The packet stream is aggregated into one-second bidirectional flow windows and passed through the transfer-trained adaptive controller. The hybrid detects the dominant attack flow immediately at onset with attack probability 0.5382 and maintains that flow between 0.5382 and 0.5703 throughout the attack period. On these replay windows, the hybrid reaches 98.05% weighted F1, while the deterministic signature baseline overfires on benign

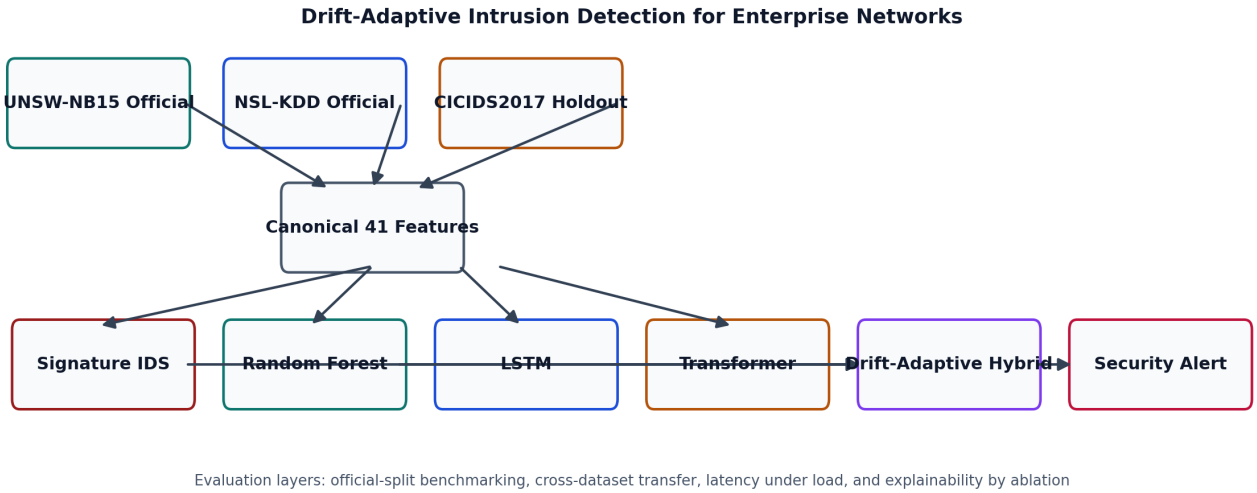

Fig. 1. System overview of the multi-dataset benchmark and the proposed Drift-Adaptive Hybrid pipeline.

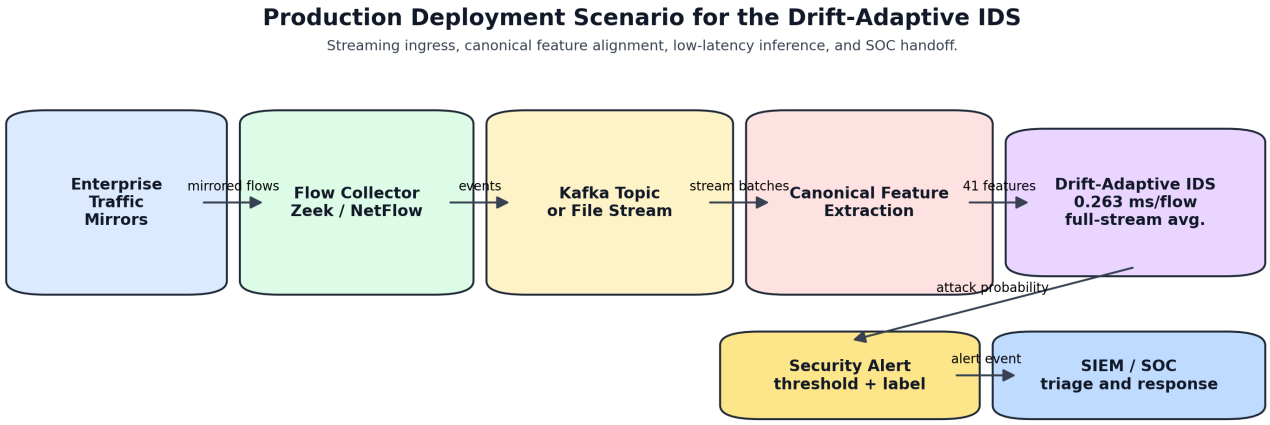

Fig. 2. Production deployment scenario with mirrored traffic, flow collection, Kafka-style ingestion, canonical feature alignment, Drift-Adaptive IDS inference, and SIEM/SOC handoff.

windows despite assigning high confidence to the attack flow itself.

VI. Discussion

Four findings stand out. First, the official split results still support the core premise of the paper: learned detectors substantially outperform the transparent rule baseline on UNSW-NB15 and NSL-KDD, and the static hybrid reaches 90.72% weighted F1 on UNSW-NB15 under a constrained unified feature schema. That slightly exceeds the 89.00% UNSW-NB15 F1 reported by Yan et al. for a specialized recent Transformer [9], while this repository also adds rule-based comparators, external transfer, and deployment analysis.

Second, the formal detector comparison strengthens the adaptation claim. Isolation Forest is not only deployment-feasible without labels; it is also the strongest detector in the measured study. The immediate detectors DDM and Page-Hinkley are close, but they still underperform slightly, and ADWIN fails to fire. This removes the criticism that the adaptation layer uses an arbitrary detector.

Third, the new CSE-CIC-IDS2018 evaluation reveals a sharper limitation than the CICIDS2017 result alone would suggest. The fact that the Signature IDS baseline is strongest on that corpus indicates that some external distributions remain easier to capture through explicit rules than through source-trained learned models. In other words, lightweight drift adaptation helps, but it is not a substitute for broader source coverage or continual learning when the target domain changes substantially.

Fourth, the failure analysis shows where the system still breaks. The adaptive hybrid can partially recover brute-force traffic such as FTP-Patator, but it remains near zero on large volumetric DoS families under the family-versus-benign protocol. This suggests that the current controller helps when cross-dataset transfer preserves compact count and flag signatures, but not when the dominant attack family's geometry changes too far from the source-domain feature manifold.

TABLE II
MODEL COMPARISON ACROSS OFFICIAL SPLITS AND EXTERNAL TRANSFER.

Split	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	ROC AUC	Latency (ms/flow)
UNSW-NB15	Drift-Aware Hybrid (Static)	90.53	91.60	90.53	90.72	0.9816	0.1804
	Random Forest	90.46	91.55	90.46	90.65	0.9792	0.0068
	LSTM	72.84	83.67	72.84	73.60	0.9065	0.0780
	Transformer	71.15	83.42	71.15	71.87	0.8800	0.1040
	Signature IDS	68.06	46.32	68.06	55.13	0.7418	0.0004
NSL-KDD	LSTM	81.08	85.23	81.08	81.01	0.9288	0.0743
	Transformer	80.61	85.04	80.61	80.52	0.9346	0.0862
	Drift-Aware Hybrid (Static)	79.23	84.52	79.23	79.06	0.9508	0.1845
	Random Forest	77.71	83.75	77.71	77.44	0.9366	0.0074
	Signature IDS	54.66	51.32	54.66	49.47	0.3541	0.0004
UNSW+NSL → CICIDS2017	LSTM	80.30	64.48	80.30	71.53	0.2904	0.0406
	Transformer	80.23	64.58	80.23	71.49	0.4886	0.0762
	Drift-Adaptive Hybrid	74.26	64.13	74.26	68.69	0.4041	0.2630
	Random Forest	58.51	65.55	58.51	61.53	0.4621	0.0060
	Signature IDS	53.34	68.33	53.34	58.09	0.4716	0.0004

TABLE III
ONLINE DRIFT ADAPTATION ON THE EXTERNAL CICIDS2017 TRANSFER STREAM.

Variant	Acc.	Prec.	Rec.	F1
Static Hybrid	57.65	67.10	57.65	61.35
Online Drift-Adaptive Hybrid	74.26	64.13	74.26	68.69

TABLE IV
SECOND EXTERNAL TRANSFER EVALUATION ON CLEANED CSE-CIC-IDS2018.

Model	Acc.	F1	ROC AUC
Signature IDS	57.95	66.37	0.5573
Static Hybrid	22.55	27.91	0.5025
Random Forest	22.45	27.58	0.5353
Drift-Adaptive Hybrid	11.89	7.39	0.3870
Transformer	9.88	2.06	0.2637
LSTM	9.90	1.78	0.2334

TABLE V
FORMAL DRIFT DETECTOR COMPARISON ON THE FULL CICIDS2017 TRANSFER STREAM.

Detector	Delay	FP	F1
Isolation Forest	2	0	70.58
DDM	0	0	69.77
Page-Hinkley	0	0	69.77
ADWIN	–	0	62.93

TABLE VI
FAMILY-LEVEL CICIDS2017 FAILURE ANALYSIS AGAINST BENIGN TRAFFIC.

Attack	Support	LSTM F1	Adaptive F1
FTP-Patator	7,938	0.0000	0.1793
SSH-Patator	5,897	0.0000	0.0194
DoS GoldenEye	10,293	0.0000	0.0015
DDoS	128,027	0.0000	0.0010
DoS Hulk	231,073	0.0000	0.0008
PortScan	158,930	0.0000	0.0000

The packet-capture replay adds a more operationally grounded perspective. Unlike the benchmark corpora, it starts from packet-level evidence and reconstructs one-second flow windows from a real recorded trace. The immediate zero-delay detection in that setting is encouraging because it shows the controller can still behave sensibly outside the public benchmark files. At the same time, the strong case-study performance should not be over-generalized: the replayed attack is a concentrated single-service event, not a full enterprise week of traffic.

Explainability-by-ablation still supports the feature engineering design. On the official UNSW-NB15 study, replacing the top Random Forest importance features with benign-reference values causes a weighted F1 drop of 0.1284 on a sampled evaluation set, while replacing random features causes only a 0.0165 drop. This gap suggests that the most important canonical features are not arbitrary proxies.

VII. FUTURE WORK

Three directions follow naturally from the current results. First, continual learning with delayed labels would test whether the adaptive controller can be upgraded from reweighting to actual parameter refresh on long-running streams. Second, reinforcement-learning or contextual-bandit control over the ensemble weights could optimize the precision, recall, and latency tradeoff directly instead of using a fixed interpolation policy. Third, multi-domain training over enterprise, cloud, and IoT corpora, potentially with graph-based side information, would address the sharp failures seen on CSE-CIC-IDS2018 and the dominant CICIDS2017 DoS families. These directions would move the repository from a strong benchmark-and-adaptation paper toward a longer-term research program on robust multi-environment intrusion detection.

VIII. CONCLUSION

This paper presented a reproducible IEEE-style benchmark for enterprise network intrusion detection across two official benchmark splits and two external transfer settings. A transparent flow-signature IDS baseline, Random Forest, LSTM, Transformer, a static Drift-Aware Hybrid, and an online Drift-

TABLE VII
POSITIONING AGAINST REPRESENTATIVE RECENT IDS PAPERS. RESULTS ARE NOT STRICTLY LIKE-FOR-LIKE BECAUSE DATASETS AND PROTOCOLS DIFFER.

Study	Architecture	Setting	F1
Yan et al. 2025	Transformer	UNSW-NB15 within-dataset	89.00
Xin & Xu 2025	Transformer-IDS	NSL→UNSW transfer	55.00
Wang et al. 2025	BS-GAT	Edge/IoT binary IDS	>99
This work	Drift-Adaptive Hybrid	UNSW+NSL→full CICIDS2017	68.69

TABLE VIII
PACKET-CAPTURE REPLAY CASE STUDY FROM THE LOCAL REAL-TIME TRACE.

Quantity	Value
Capture packets	221,253
Flow windows	201
Benign warm-up	10 s
Attack duration	35 s
Hybrid delay	0 s
Dominant attack flow	UDP/3000
Hybrid F1	98.05%
Signature IDS F1	5.16%

Adaptive Hybrid controller were compared under a shared 41-feature representation. The static hybrid was strongest on UNSW-NB15, LSTM was strongest on NSL-KDD and on the full CICIDS2017 transfer corpus, and the Signature IDS baseline was unexpectedly strongest on cleaned CSE-CIC-IDS2018. The key novelty result remains that online drift adaptation improved the hybrid from 61.35 to 68.69 weighted F1 on the full CICIDS2017 stream without retraining the base detectors, and the formal detector study showed that Isolation Forest was the strongest choice in this controller. The broader result is more nuanced than a simple accuracy win: learned detectors can strongly outperform rules on in-domain splits, but cross-dataset generalization remains fragile, modern external corpora can invert the ranking, and deployment-time adaptation should be treated as a first-class IDS design concern rather than a late-stage add-on.

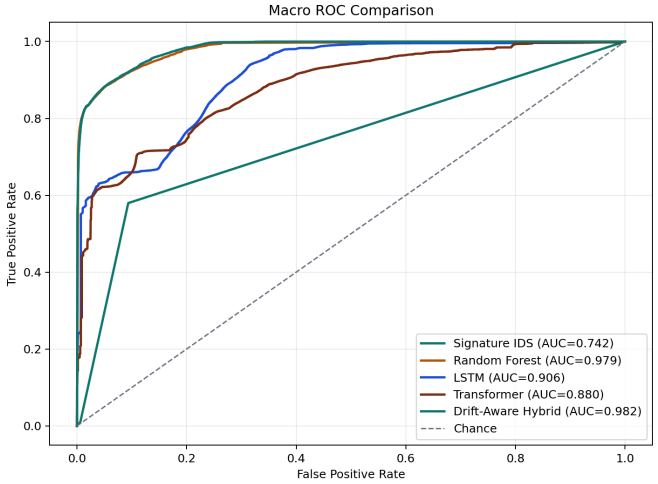

Fig. 3. ROC comparison on the official UNSW-NB15 split.

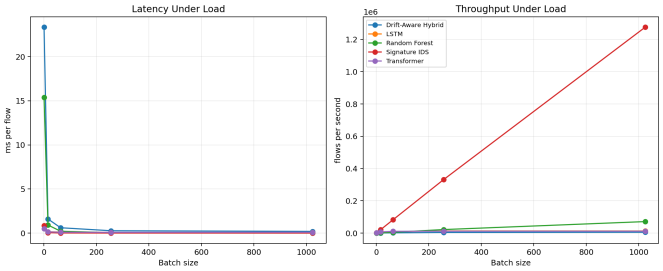

Fig. 4. Latency and throughput under load for the official UNSW-NB15 evaluation.

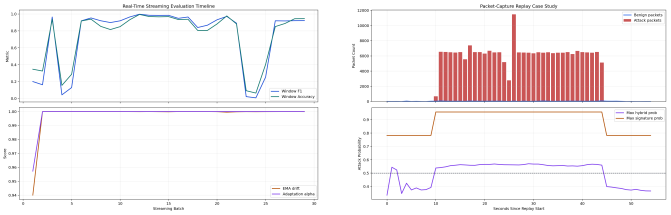

Fig. 5. Deployment-time behavior under external replay. Left: 29-window CICIDS2017 streaming timeline. Right: packet-capture replay with benign warm-up, sustained UDP/3000 attack traffic, and benign cool-down.

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this study.

REFERENCES

[1] M. Roesch, "Snort: Lightweight intrusion detection for networks," in *Proceedings of the 13th USENIX Conference on System Administration*, 1999, pp. 229–238.

[2] S. M. Kasongo and Y. Sun, "Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset," *Journal of Big Data*, vol. 7, no. 1, p. 105, 2020.

[3] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys*, vol. 46, no. 4, pp. 44:1–44:37, 2014.

[4] N. Moustafa and J. Slay, "Unsw-nb15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *2015 Military Communications and Information Systems Conference (MilCIS)*, 2015, pp. 1–6.

[5] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 2009, pp. 1–6.

[6] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, 2018, pp. 108–116.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[9] H. Yan, X. Pang, S. Zhou, and H. Fan, "Transformer-based intrusion detection for post-5g and 6g telecommunication networks using dynamic semantic embedding," *Future Internet*, vol. 17, no. 12, p. 544, 2025.

[10] C. Xin and K. Xu, "Cross-dataset transformer-ids with calibration and auc optimization (evaluated on nsl-kdd, unsw-nb15, cic-ids2017)," *Journal of Cyber Security*, vol. 7, no. 1, pp. 483–503, 2025.

[11] Y. Wang, Z. Han, Y. Du, J. Li, and X. He, "Bs-gat: a network intrusion

detection system based on graph neural network for edge computing," *Cybersecurity*, vol. 8, p. 27, 2025.

[12] A. Bifet and R. Gavaldà, "Learning from time-changing data with adaptive windowing," in *Proceedings of the 2007 SIAM International Conference on Data Mining*, 2007, pp. 443–448.

[13] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with drift detection," in *Advances in Artificial Intelligence – SBIA 2004*, 2004, pp. 286–295.

[14] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[15] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.