Scaling Efficiency and Robustness Trade-offs in GNN-Based NIDS via Gradient Bypass

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does the computational efficiency of bypassing obfuscated gradients in GNN-based NIDS models scale with increasing network size, and what is the trade-off between robustness and inference time on. Machine Learning has been steadily gaining traction for its use in Anomaly-based Network Intrusion Detection Systems (A-NIDS). Research into this domain is frequently performed using the KDD~CUP~99 dataset as a benchmark. 17 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.1/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives. Research question: How does the computational efficiency of bypassing obfuscated gradients in GNN-based NIDS models scale with increasing network size, and what is the trade-off between robustness and inference time on the KDD Cup 99 dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.1/10.

3 Results

12 papers retrieved. 17 claims extracted; 1 independently verified. Quality review score: 4.1/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Classifiers trained on UNSW-NB15 match or better the Weighted F1-Score of those trained on NSL-KDD and KDD-99 in the bin	✓	0.37
M-NIDSs detect intrusions by an exact matching of network traffic to known attack signatures.	×	0.08
Statistical-based A-NIDS techniques rely on the quasi-stationary process assumption.	×	0.05
The UNSW-NB15 dataset was simulated using the IXIA PerfectStorm tool at the ACCS (Australian Center of Cyber Security).	×	0.04
The UNSW-NB15 simulation was conducted over two days in sessions of 16 hours and 15 hours.	×	0.02
45 unique IP addresses were used over 3 networks for the UNSW-NB15 dataset.	×	0.03
11 IP addresses on 2 networks were used for the KDD dataset.	×	0.03
Attacks for UNSW-NB15 were chosen from a constantly-updated CVE site.	×	0.03
Normal behavior was not simulated for the UNSW-NB15 dataset.	×	0.04
Packet-level traffic for UNSW-NB15 was captured via TCPdump.	×	0.03
A total of 2,540,044 records were generated for the UNSW-NB15 dataset.	×	0.03
The UNSW-NB15 dataset incorporates 10 target classes: one Normal and 9 anomalous (Fuzzers, Analysis, Backdoors, DoS, Exp	×	0.05
The KDD dataset has 5 target classes.	×	0.12
The Null Error Rate for UNSW-NB15 is 55.06%.	×	0.03
The Null Error Rate for KDD is 26.1%.	×	0.04
The UNSW-NB15 train set contains 175,341 data points.	×	0.04
The UNSW-NB15 test set contains 82,332 data points.	×	0.05

References

- `http://arxiv.org/abs/2212.00966v1`

- http://arxiv.org/abs/1811.05372v1
- http://arxiv.org/abs/2207.06819v5