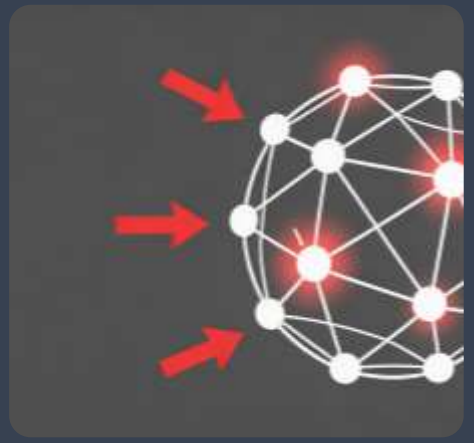
Directional Flow Graph Framework for Detecting Known and Unknown Malicious Tra汪c

DASFAA 2026

Yang Wu, Ruijia Wang, Mingyuan Zang, Jie Wu

China Telecom Cloud Computing Research Institute & Temple University

Background



01. Growing Security Challenges

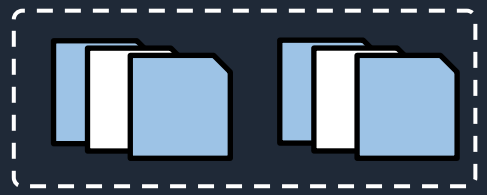
- The increasing complexity and diversity of attack methods
- data breaches, service interruptions…

02. The Dilemma of Traditional Methods

- Rule-based systems have poor adaptability
- Unknown attacks

Detecting Known and Unknown Malicious Traffic

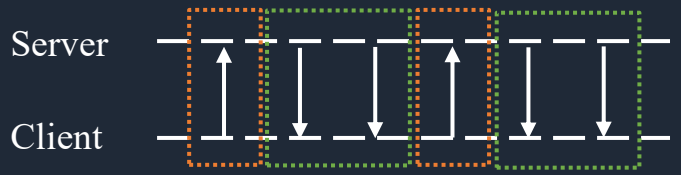
Motivation



Traffic Flow

Packet length [40, 48, 48, 40, 48, 48]

Packet direction [1, -1, -1, 1, -1, -1]



Limitation 1: Feature representation information loss

bidirectional traffic → undirected graphs or sequences

losing key temporal interaction patterns and structural dependencies

Limitation 2: The single paradigm is weak

Supervised learning: Good at known attacks

Unsupervised learning: Good at unknown attacks

Idea and Contribution

Idea淫DGNet (Directional Graph Network)

directed graph modeling + a hybrid learning paradigm

First work: Directional Flow Graph

A breakthrough to capture the deep structure information and time series information of the traffic at the same time.

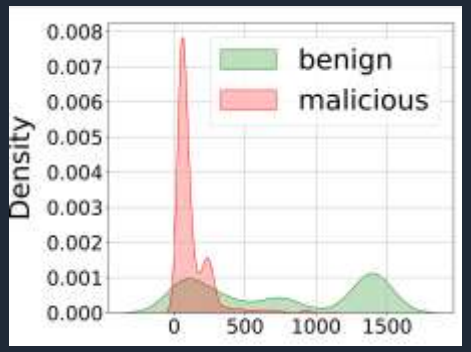
Design: Directional Graph Isomorphism Networks (DIRGIN)

The GNN model is designed for a directional flow graph.

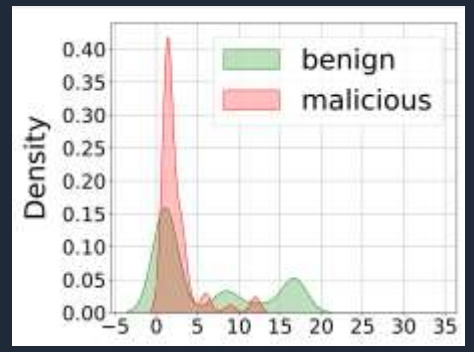
Build: An end-to-end Hybrid Learning Framework

Combining the advantages of supervised learning and unsupervised learning to achieve comprehensive detection.

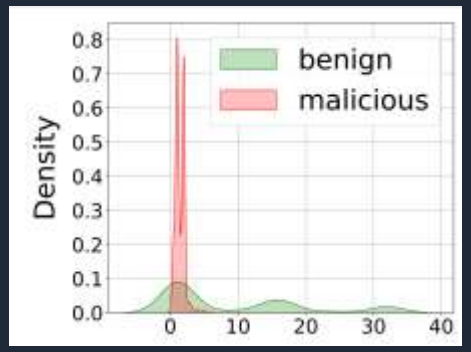
Design Intuition: Traffic Pattern Differences



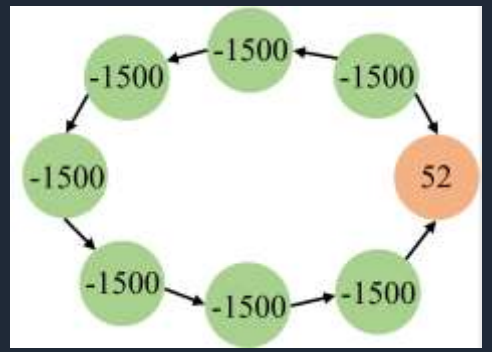
Packet length



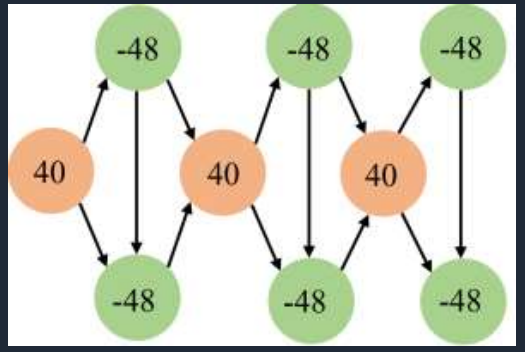
Burst



Down/Up Ratio



Benign



Malicious

Observation淫USTC-TFC2016 dataset

There are significant behavioral pattern differences between benign and malicious traffic in the three dimensions of the length distribution of the underlying packet interaction, burst frequency, and uplink and downlink ratio

Conclusion: Feasibility veri钱ation of the scheme

Examples of directional flow graphs for benign and malicious traffic. The directed graph pattern difference can effectively distinguish malicious traffic from benign traffic.

Method: Directional Flow Graph

How to build a directed flow graph?

Nodes: Data packet

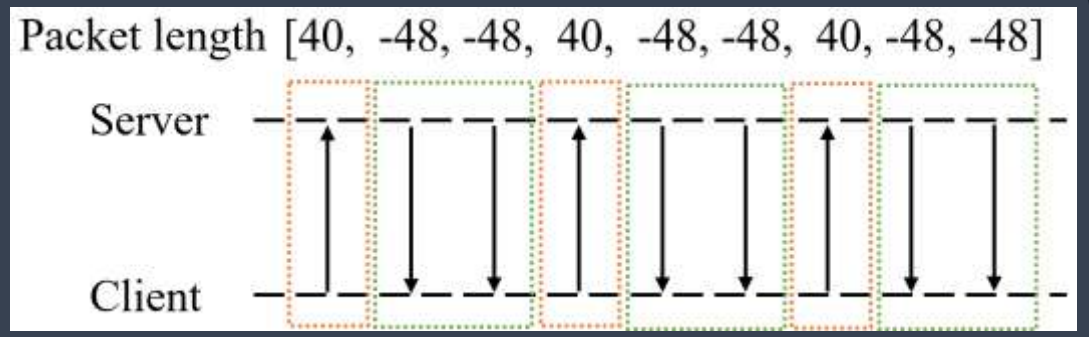
Feature = packet length Ö direction (uplink =1, downlink =-1)

Edges: structured traffic sequences

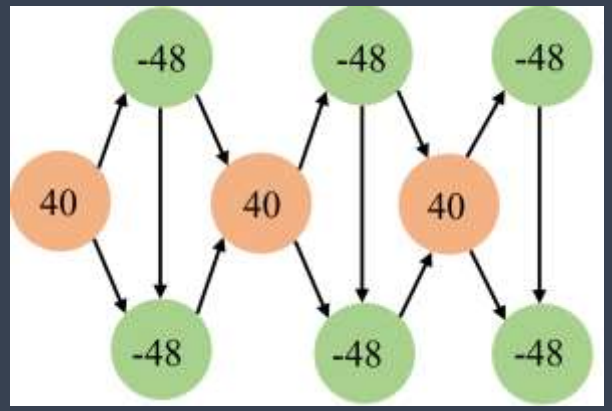
Intro-Burst: sequential join;
Inter-Burst: head-to-head, tail-to-tail;

Advantage: Differentiation of behavioral patterns

It preserves the order and direction information of data packets, and can effectively distinguish the completely different bidirectional interaction behaviors.



Bi-flow traffic



Corresponding directional graph

Method: DGNet

Supervised Learning

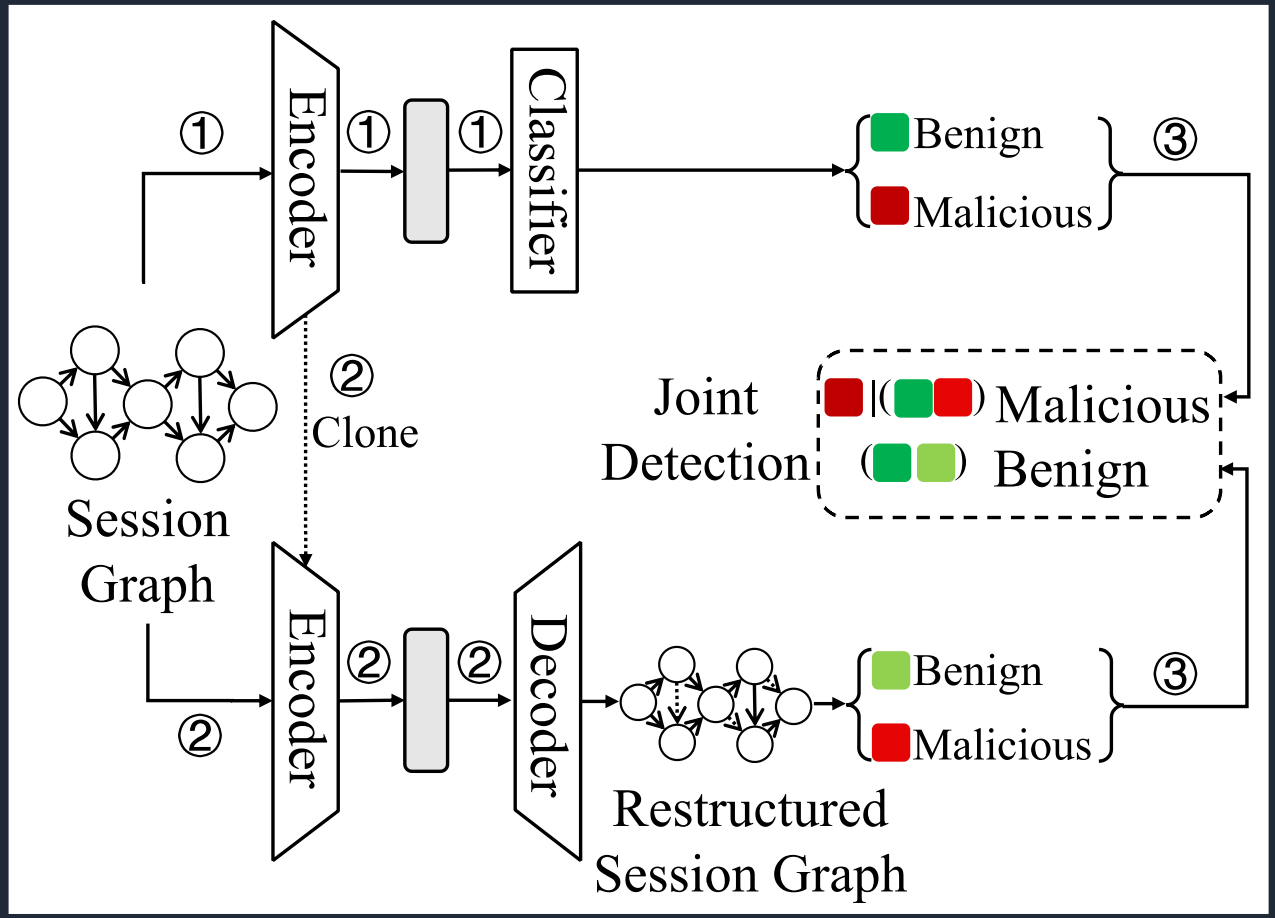
DIRGIN is used to train a classifier to accurately identify known attack behaviors in the system through labeled data.

Unsupervised Learning

Copy the parameters of the DIRGIN encoder, and construct an autoencoder network to learn the deep feature patterns of benign traffic, so as to effectively find unknown attacks.

Inference and Detection

The supervised classification results and unsupervised reconstruction loss were combined to detect malicious traffic.



Method: DIRGIN Encoder

$$\mathbf{m}_v^{\text{in}} = \sum_{u \in \mathcal{N}_{\text{in}}(v)} \mathbf{h}_u^{k-1} \mathbf{W}_{\text{in}}$$

$$\mathbf{m}_v^{\text{out}} = \sum_{u \in \mathcal{N}_{\text{out}}(v)} \mathbf{h}_u^{k-1} \mathbf{W}_{\text{out}}$$

$$\mathbf{h}_v^{\text{self}} = \mathbf{h}_v^{k-1} \mathbf{W}_h$$

$$\mathbf{m}_v^{\text{com}} = \sigma\left(\left[\mathbf{m}_v^{\text{in}} \parallel \mathbf{m}_v^{\text{out}}\right]\right) \mathbf{W}_{\text{com}}$$

$$\mathbf{h}_v^k = \sigma\left(\mathbf{m}_v^{\text{com}} + (1+\epsilon^k) \cdot \mathbf{h}_v^{\text{self}}\right) \mathbf{W}_{\text{final}}$$

Why DIRGIN is needed豬

Standard GNNS (such as GIN) ignore the directions of edges when dealing with directed graphs, and cannot distinguish incoming edges from outgoing edges, resulting in the loss of key interactive information such as requests and responses.

Core mechanism: Separation and aggregation of incoming and outgoing edges

Incoming edge

$m_v = \Sigma(h_u \cdot W_{in})$

Outgoing edge

$m_v = \Sigma(h_u \cdot W_{out})$

Information Fusion

The fully connected layer after concatenation

Core advantage: Accurate capture of causality

The “influencers” (incoming edges) and “influenced” (outgoing edges) of nodes are clearly distinguished.

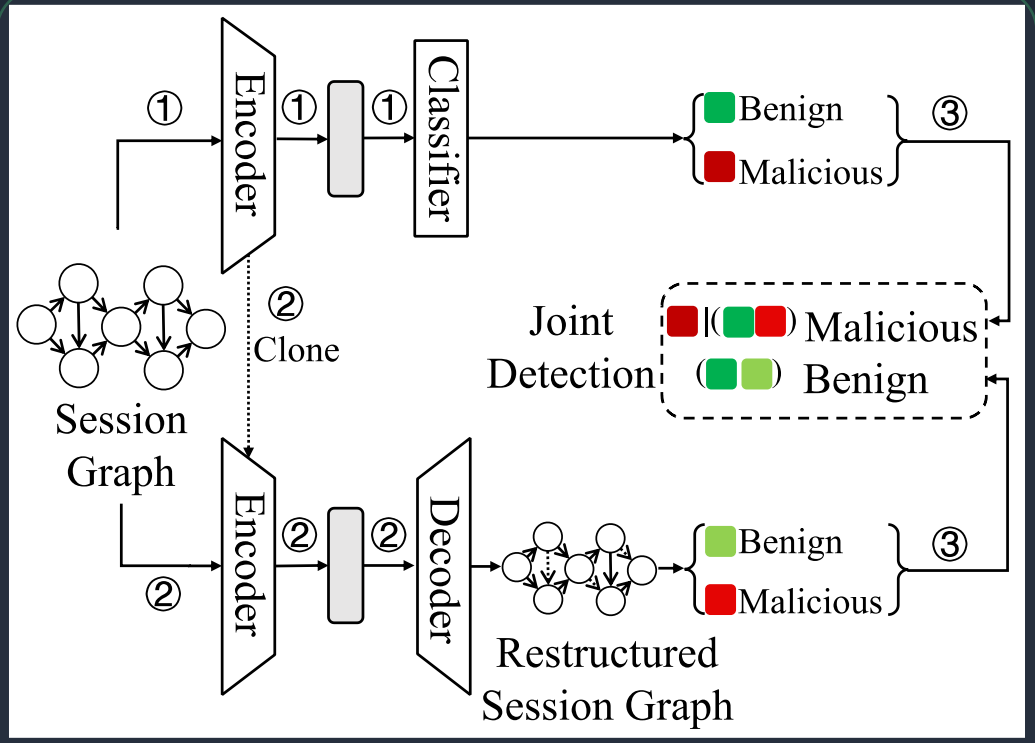
Method: Hybrid Learning and Inference

Final decision logic

- If the supervised model predicts malicious, the result is directly determined to be malicious.
- If the supervised model predicts benign, the unsupervised model reconstruction loss is examined 淫
  -- loss> δ: unknown| loss < δ: benign

Threshold δ

A Data-Driven Approach Based on Kernel Density Estimation (KDE)
The normal and malicious traffic reconstruction loss distribution bimodal valley or 95% quantile is chosen as the optimal threshold δ on the validation dataset.



$$\hat{\mathbf{y}}' = \begin{cases} 1 & \text{if } \hat{\mathbf{y}}^s = 1 \\ 1 & \text{if } \hat{\mathbf{y}}^s = 0 \text{ and } \ell^u > \delta \\ 0 & \text{if } \hat{\mathbf{y}}^s = 0 \text{ and } \ell^u < \delta \end{cases}$$

Experiment Settings

Datasets

USTC-TFC2016

A classic malicious traffic benchmark dataset in the field of network security.

CICIoT2024

The large-scale dataset for IoT scenarios contains more complex attack behaviors and real network traffic characteristics.

Scenarios

I: Known Attacks

The training and testing set contains the same type of attack samples.

II: Unknown Attacks

The training and testing set contains completely different attack types.

Baselines

Supervised

FS-Net, GraphDAPP, Graph2Vec, FG-SAT

Unsupervised

Kitsune

Hybrid

IoTArgos

Other SOTA

CVAE-EVT, ECNet

Experimental results: Known Malicious Traffic Detection

USTC-TFC2016-I

F1-Score

The benchmark attack detection scenario has a relatively canonical structure

0.9973

CICIoT2024-I

F1-Score

Complex real attack scenario data distribution is more challenging

0.9741

The DGNet has the best results

DGNet > graph-based / sequence-based

Directed > Undirected

The direction information is effectively modeled, enabling DGNet to outperform similar graph-based baselines.

Table 3: Performance comparison on known attacks on the two datasets.

Method	USTC-TFC2016-I				CICIoT2024-I			
	Acc	F1	Precision	Recall	Acc	F1	Precision	Recall
FSNet	0.9966	0.9964	0.9960	0.9938	0.9689	0.9676	0.9695	0.9660
GraphDAPP	0.9697	0.9669	0.9729	0.9618	0.7167	0.6741	0.7445	0.6747
Graph2Vec	0.8851	0.8531	0.9237	0.8255	0.8233	0.6942	0.9033	0.6646
Kitsune	0.8331	0.8061	0.8458	0.7901	0.7746	0.7723	0.7721	0.7798
IoTArgos	0.9732	0.9657	0.9788	0.9714	0.9423	0.9383	0.9454	0.9411
CVAE-EVT	0.9962	0.9958	0.9956	0.9960	0.9626	0.9602	0.9504	0.9677
ECNet	0.9937	0.9932	0.9925	0.9938	0.9416	0.9396	0.9418	0.9376
FG-SAT	0.9969	0.9967	**0.9972**	0.9962	0.9607	0.9592	0.9644	0.9552
DGNet(Mean Pool)	**0.9975**	**0.9973**	0.9967	**0.9979**	0.9714	0.9704	0.9716	0.9694
DGNet(Add Pool)	0.9959	0.9956	0.9957	0.9955	0.9366	0.9387	0.9307	0.9341
DGNet(Max Pool)	0.9926	0.9920	0.9917	0.9924	**0.9739**	**0.9741**	**0.9722**	**0.9731**

Experimental results: Unknown Malicious Traffic Detection

Excellent generalization ability

The performance degradation of DGNet is significantly lower than that of other comparison methods, demonstrating the strong robustness of the model under non-stationary data distribution.

Advantages of Hybrid Learning

It makes up for the performance deficiency of pure supervised learning on unknown samples.

Table 4: Performance comparison on unknown attacks on the two datasets.

Method	USTC-TFC2016-II				CICIoT2024-II			
	Acc	F1	Precision	Recall	Acc	F1	Precision	Recall
FSNet	0.9938	0.9934	0.9924	0.9944	0.9330	0.9311	0.9434	0.9258
GraphDAPP	0.9209	0.9110	0.9406	0.8948	0.6157	0.5574	0.6678	0.5948
Graph2Vec	0.9372	0.9239	0.9575	0.9027	0.9023	0.8674	0.9394	0.8328
Kitsune	0.7906	0.7593	0.7931	0.7473	0.8051	0.8051	0.8096	0.8090
IoTArgos	0.9745	0.9677	0.9797	0.9730	0.9357	0.9352	0.9368	0.9356
CVAE-EVT	0.9935	0.9930	0.9925	0.9936	0.9457	0.9420	0.9375	0.9474
ECNet	0.9964	0.9961	0.9969	0.9954	0.8613	0.8570	0.8825	0.8537
FG-SAT	0.9945	0.9941	0.9955	0.9927	0.9425	0.9419	0.9474	0.9397
DGNet(Mean Pool)	**0.9976**	**0.9974**	**0.9971**	**0.9977**	0.9581	0.9579	0.9594	0.9569
DGNet(Add Pool)	0.9975	0.9973	0.9971	0.9976	0.9228	0.9218	0.9280	0.9195
DGNet(Max Pool)	0.9961	0.9958	0.9948	0.9969	**0.9646**	**0.9644**	**0.9661**	**0.9633**

Experimental results: ablation study

w/ Super

Only supervised

It performs well in known attack scenarios, but decreases significantly when facing unknown attacks.

w/ Unsuper

Only unsupervised

It is limited by the fuzziness of its feature representation; the overall classification accuracy is always lower than the supervised learning baseline.

DGNet (Ours)

Supervised + unsupervised

It verifies the effectiveness and robustness of the hybrid architecture.

Scenario	Method	USTC-TFC2016		CICIoT2024	
		Acc	F1	Acc	F1
Known	DGNet	0.9975	0.9973	0.9739	0.9741
	w/ Super	0.9741	0.9716	0.9668	0.9658
	w/ Unsuper	0.9544	0.9493	0.8703	0.8575
Unknown	DGNet	0.9976	0.9974	0.9646	0.9644
	w/ Super	0.9355	0.9281	0.9379	0.9374
	w/ Unsuper	0.9289	0.9202	0.9258	0.9243

Validation of complementary

The two learning paradigms exhibit extremely strong complementarity in DGNet. 1+1 > 2

Experimental results: ablation study of GNN backbone

Scenario	Method	USTC-TFC2016		CICIoT2024	
		Acc	F1	Acc	F1
Known	DGNet_{DIRGIN}	0.9975	0.9973	0.9739	0.9741
	DGNet_{GIN}	0.9862	0.9851	0.8957	0.8933
	DGNet_{GAT}	0.9617	0.9592	0.8972	0.8948
	DGNet_{GCN}	0.9454	0.9397	0.8544	0.8505
Unknown	DGNet_{DIRGIN}	0.9976	0.9974	0.9646	0.9644
	DGNet_{GIN}	0.9936	0.9931	0.8548	0.8518
	DGNet_{GAT}	0.9645	0.9625	0.7636	0.7620
	DGNet_{GCN}	0.9216	0.9150	0.8807	0.8798

DIRGIN (Ours)

Dir + GIN

It performs better than other undirected GNNs.

Experimental results: Analysis of missed detection recovery

Substantial reduction in missed detections

•**First step (only-super): a lot of missed attacks** (FN). The model is not sensitive to some malicious samples.

•**Final step (hybrid): after adding unsupervised learning, a large** number of missed detections are successfully corrected.

Reconstruction Loss distribution: Characteristics of anomalous samples

The reconstruction loss for malicious samples (FN) missed by the supervised stage is significantly higher than that for benign samples, making them easier to detect.

True Label \ Predicted Label	Super	1	0
	1	48426	3366
	0	323	90527

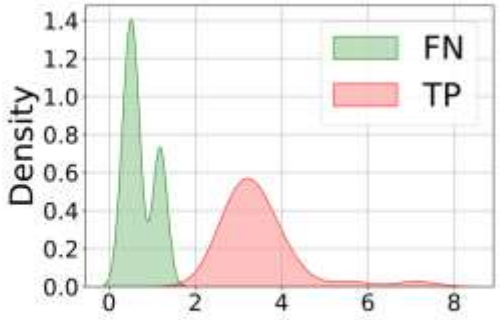
True Label \ Predicted Label	Final	1	0
	1	51760↑	32↓
	0	323	90527

(a) First-stage and final classification confusion matrices for known attack scenarios.

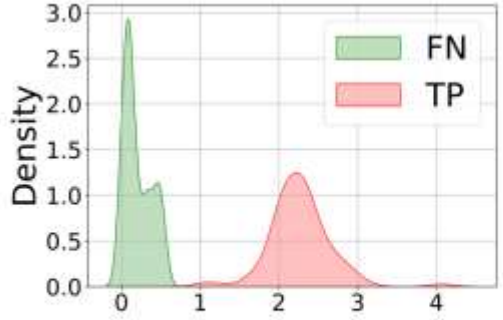
True Label \ Predicted Label	Super	1	0
	1	44330	9037
	0	256	90594

True Label \ Predicted Label	Final	1	0
	1	53273↑	94↓
	0	256	90594

(b) First-stage and final classification confusion matrices for unknown attack scenarios.



(c) Reconstruction loss distribution of FN and TP samples corresponding to (a).



(d) Reconstruction loss distribution of FN and TP samples corresponding to (b).

Conclusion and Future Work

Conclusion

Innovative traffic representation

The directional flow graph structure captures the spatial topology and time series characteristics.

Targeted graph neural network learners

The DIRGIN module is designed to optimize the extraction and deep learning of edge direction information in directed graphs.

Detection performance at the SOTA level

Based on the hybrid learning framework, optimal results are achieved in both known-attack classification and unknown-attack detection tasks.

Future work

Model lightweight

Explore a more streamlined GNN architecture to adapt it to edge computing devices.

Real-time traffic detection

Optimize our stream-processing pipeline and model inference to achieve millisecond response times over high-speed network links.

Combat sophisticated and covert attacks

Extend the framework to deep-detection tasks for adversarial samples, encrypted traffic, and multi-stage APT attacks.

Thank You!

QUESTIONS & ANSWERS

CONTACT US

wuy92@chinatelecom.cn