# No More Shortcuts: Network Traffic Anomaly Detection via Bidirectional Prediction

**Xinglin Lian**[1], **Chengtai Cao**[2], **Fanglin Yu**[1], **Ting Zhong**[1] and **Fan Zhou**[1,3*]
[1]University of Electronic Science and Technology of China
[2]City University of Hong Kong
[3]Key Laboratory of Intelligent Digital Media Technology of Sichuan Province
kenshin.lian24@gmail.com, chengtcao2-c@my.cityu.edu.hk, 202522090630@std.uestc.edu.cn,
{zhongting, fan.zhou}@uestc.edu.cn

## Abstract

Network Traffic Anomaly Detection (NTAD), particularly under zero-positive settings, is a critical task in cybersecurity. Existing zero-positive NTAD approaches primarily rely on reconstruction-based pipelines. Nevertheless, these methods are susceptible to an "identical shortcut" issue, where models indiscriminately reconstruct both normal and anomalous inputs, leading to anomaly overgeneralization. To address this limitation, we propose BiPred, the first prediction-based detection paradigm for NTAD. BiPred symmetrically divides each network traffic sample into two segments and reformulates the NTAD task as a bidirectional prediction across these two segments. Unlike reconstruction-based methods, our BiPred not only eliminates shortcut learning by design but also establishes explicit bidirectional contextual dependencies, making it more sensitive to anomalous traffic. Moreover, we develop a novel Residual Scanning Mamba block that encodes multi-view contextual information to support bidirectional prediction. A residual fusion mechanism is proposed for the multi-view scanning Mamba to suppress the accumulation of inter-view redundancy. This design prevents representation degradation and provides multi-view contextual details for prediction. Extensive experiments demonstrate the superiority of our BiPred paradigm, highlighting a new research direction for NTAD. Code is available at https://github.com/ikun0124/BiPred.

## 1 Introduction

Network Traffic Anomaly Detection (NTAD) is a fundamental task in modern cybersecurity. It monitors real-time network traffic to identify unauthorized or malicious connections [He *et al.*, 2026], ensuring high-quality service availability, protected user privacy, and reliable data security [Zhang *et al.*, 2025a; Fu *et al.*, 2026b]. However, accurately detecting anomalous traffic remains challenging, as attackers continuously modify and obfuscate their strategies.

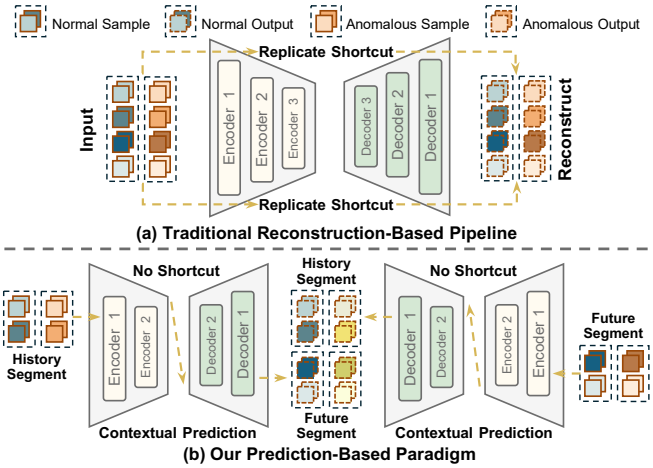


Figure 1: Comparison with existing paradigms. (a) Reconstruction-based methods suffer from shortcut learning and anomaly overgeneralization. (b) Our prediction-based paradigm eliminates shortcuts and highlights anomalous deviations.

These evolving and previously unseen attack patterns often render traditional supervised approaches ineffective, particularly when labeled anomalous samples are scarce or outdated [Dang *et al.*, 2025]. Consequently, research efforts in NTAD have increasingly shifted toward more scalable detection paradigms. Reconstruction-based zero-positive learning has emerged as a particularly promising direction. In this paradigm, an autoencoder is trained exclusively on normal traffic to minimize reconstruction error [Zhao *et al.*, 2024b]. During inference, the method assumes that unseen or evolving anomalies will produce higher reconstruction errors, thus enabling their accurate detection [Yang *et al.*, 2025].

**Challenge.** Although theoretically appealing, this assumption does not consistently hold in practice. Recent studies reveal that reconstruction-based approaches are susceptible to the **"identical shortcut"** phenomenon [Lian *et al.*, 2025a; Lian *et al.*, 2026b]. As illustrated in Figure 1(a), the reconstruction-based training pipeline inherently leaves a learnable shortcut, causing the model to reconstruct inputs via simple replication rather than to encode meaningful contextual representations. When the shortcut extends to anomalous samples, they achieve fidelity comparable to normal samples

*Corresponding Author

through replication, undermining anomaly detection. Some works mitigate this limitation by introducing inter-sample learning, feature extraction, and memory modules into autoencoder models [Lian *et al.*, 2025a; Lian *et al.*, 2026b; Li *et al.*, 2023]. However, the paradigm-level shortcut is unavoidably triggered by their reconstruction-based pipeline, rendering existing efforts inherently limited. This observation raises a key research question: **How can the paradigm-level shortcut curse be fundamentally eliminated?**

**Contribution 1.** To answer this question, we propose *a new paradigm* named **BiPred**: **Bi**directional **Pred**iction-based anomaly detection for NTAD. As illustrated in Figure 1(b), BiPred divides each network traffic sample into two segments, history and future, and reformulates the reconstruction task as a bidirectional prediction process. Specifically, two prediction models are trained exclusively on normal traffic: one predicts future segments from past segments, and the other predicts past segments from future segments. During inference, anomalous traffic disrupts the learned normal contextual relationships across the two segments, resulting in larger prediction deviations. Compared with reconstruction-based approaches, our paradigm offers two intrinsic advantages: (1) **Shortcut Elimination.** The prediction-based pipeline leaves the model with no more shortcut pathways to directly replicate inputs, thereby eliminating shortcut learning by design. (2) **Contextual Learning.** The bidirectional prediction forces the model to encode forward–backward contextual dependencies, enhancing sensitivity to inconsistencies in anomalous traffic patterns.

**Contribution 2.** Contextual representation is critical for BiPred's prediction models, as inadequate learning can cause prediction collapse. Recently, Mamba-based architectures leverage multi-directional scanning mechanisms, achieving strong sequential modeling capacity while maintaining linear computational complexity [Patro and Agneeswaran, 2025]. This renders Mamba particularly suitable for our prediction framework. Nevertheless, its multi-directional scanning mechanism may introduce severe **information redundancy** [Guo *et al.*, 2024; Shi *et al.*, 2024], even degrading Mamba's performance in NTAD tasks [Wang *et al.*, 2024]. This redundancy causes the model to repeatedly emphasize overlapping patterns, diluting fine-grained directional information [Xiao *et al.*, 2025]. To address this limitation, we propose a novel **Multi-View Residual Scanning Mamba (MRSM)** module. Exploiting the hierarchical structure of network traffic, we first design a dual-view scanning strategy to form complementary contextual information at both packet-level and flow-level of network traffic. To explicitly mitigate inter-view redundancy, we propose a residual fusion mechanism: one view is treated as the anchor for primary contextual modeling, while the other serves as an auxiliary view to capture residual representations relative to the anchor. This residual operation isolates auxiliary-specific complementary information while preventing the secondary accumulation issue of inter-view redundancy. The final representation integrates the anchor and auxiliary residual information, forming a fine-grained multi-view context that underpins the bidirectional prediction.

**In summary, our contributions are threefold:**

- We propose a novel bidirectional prediction-based anomaly detection paradigm. To the best of our knowledge, it is the first framework to fundamentally eliminate the "identical shortcut" issue in zero-positive NTAD.
- We propose a Multi-View Residual Scanning Mamba module that captures rich multi-view contextual information and explicitly mitigates inter-view redundancy, thereby supporting effective bidirectional prediction.
- Experiments on three real-world datasets demonstrate that BiPred consistently outperforms state-of-the-art zero-positive NTAD methods.

## 2 Related Work

### 2.1 Network Traffic Anomaly Detection

Zero-positive Network Traffic Anomaly Detection (NTAD) aims to identify previously unseen anomalous traffic using only normal training samples [Zhang *et al.*, 2024; Yang *et al.*, 2025]. Existing approaches primarily follow a reconstruction-based paradigm, where autoencoder models are trained to faithfully reconstruct normal traffic patterns. These methods emphasize accurately modeling normal behavior, assuming that anomalies deviate from the learned representations and thus incur larger reconstruction errors. Representative methods include MANomaly [Zhang *et al.*, 2022] and ARCADE [Lunardi *et al.*, 2023], which integrate adversarial training into autoencoder architectures to enhance reconstruction robustness. Trident [Zhao *et al.*, 2024b] employs a U-Net-based autoencoder to preserve fine-grained structural characteristics of network traffic. MFAD [Zheng *et al.*, 2023] and MFR [Lian *et al.*, 2025b] exploit low-pass filtering to highlight texture-level representations for improved normal pattern learning. FreeUp [Lian *et al.*, 2026a] introduces a frequency-decoupled reconstruction framework for modeling encrypted "full-frequency" network traffic. unFlowS [Yang *et al.*, 2025] projects high-dimensional traffic features into a spectral space to facilitate anomaly discrimination.

Despite these reconstruction-oriented enhancements, recent studies have revealed a paradigm-level limitation termed the "identical shortcut" phenomenon [Lian *et al.*, 2025a]. Specifically, autoencoders may converge to shortcut mappings that replicate both normal and anomalous samples with comparable fidelity, undermining anomaly separability. To mitigate this issue, UnDiff [Lian *et al.*, 2025a] introduces inter-sample difference learning to reduce shortcut reliance, ConMD [Lian *et al.*, 2026b] employs a pretrained model to extract discriminative features, while other works adopt memory-augmented mechanisms to reinforce normal pattern memorization [Li *et al.*, 2023]. Nevertheless, these methods remain fundamentally constrained by the reconstruction-based pipeline and cannot fully eliminate the shortcut risk caused by direct input–output replication.

To avoid these paradigm-level limitations, we shift NTAD to a prediction-based formulation. By introducing internal bidirectional prediction between segments of each traffic sample, our approach explicitly removes the direct replication path between inputs and outputs, thereby preventing shortcut learning. Furthermore, this predictive formulation

encourages richer contextual modeling, which is particularly effective for detecting anomalies characterized by subtle contextual inconsistencies [Zhao *et al.*, 2024a], rather than relying solely on deviations from naive reconstruction.

### 2.2 Mamba-Based Representation Learning

The Mamba architecture is built upon selective state-space sequence models, enabling efficient contextual modeling with linear computational complexity. It has demonstrated a favorable efficiency–effectiveness trade-off in domains such as computer vision and time-series analysis [Zhang *et al.*, 2025b; Qin *et al.*, 2025]. By design, Mamba employs multi-directional scanning mechanisms, which are intended to fuse representations from different directions and provide comprehensive global contextual information. However, recent studies have shown that multi-directional scanning often introduces substantial information redundancy across branches [Guo *et al.*, 2024; Shi *et al.*, 2024; Xiao *et al.*, 2025]. This redundancy causes the model to repeatedly emphasize overlapping components, thereby diluting fine-grained directional information and, in extreme cases, degenerating the model into a single-branch scanning behavior.

The redundancy problem is particularly pronounced in the context of network traffic analysis. In traffic classification tasks, multi-directional Mamba scanning can degrade performance, whereas simple unidirectional scanning often yields better results [Wang *et al.*, 2024]. Consequently, some recent works adopt unidirectional Mamba architectures to avoid redundancy-induced degradation [Deng *et al.*, 2025; Xia *et al.*, 2025]. However, network traffic inherently exhibits rich multi-view characteristics, where anomalies often manifest through intricate flow-view and packet-view interactions [Zhao *et al.*, 2024a]. Purely unidirectional scanning, therefore, risks overlooking critical multi-view contextual cues and inducing biased representations.

To address this challenge, we propose a novel Multi-View Residual Scanning Mamba module with explicit redundancy elimination. Instead of simply fusing scan outputs, our approach designates one view as the primary anchor and extracts complementary information from auxiliary views via residual learning. This mechanism mitigates information degradation caused by redundant component accumulation while preserving fine-grained multi-perspective contextual details, thereby supporting more effective bidirectional prediction and contextual anomaly awareness.

## 3 Methodology

### 3.1 Problem Definition

Let $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N\}$ denote a training set of $N$ normal samples. During training, the model exclusively learns to characterize normal network traffic patterns [Lian *et al.*, 2024; Cai *et al.*, 2026]. During inference, for a given test sample $\boldsymbol{x}$, the model outputs an anomaly score $s$, where higher values indicate a greater likelihood of anomalous activity. Each sample $\boldsymbol{x}$ corresponds to a network flow, i.e., a sequence of packets exchanged between endpoints within a connection session. To capture the hierarchical characteristics of network traffic, we represent each flow as a packet-level matrix $\boldsymbol{x} \in \mathbb{R}^{F \times P}$, where $F$ is the number of packets in the flow and $P$ denotes the packet size in bytes. Each packet follows the prior traffic preprocessing protocol [Lian *et al.*, 2025b; Fu *et al.*, 2026a; Zheng *et al.*, 2024].

### 3.2 Mamba-Based Bidirectional-Prediction

This section presents the proposed BiPred, as illustrated in Figure 2(a). Each traffic sample is first symmetrically divided into two segments. Then, zero-positive NTAD is reformulated as a bidirectional prediction task, in which one segment is used to predict its counterpart. As shown in Figure 2(b), each prediction model is built upon a Multi-View Residual Scanning Mamba encoder to capture rich contextual dependencies. The encoder incorporates an explicit residual mechanism to suppress inter-view redundancy while highlighting view-specific information. The resulting multi-view features provide strong contextual cues for the prediction decoder. During inference, the anomaly score $s$ is quantified as the deviation between the predicted segment and the corresponding input segment, where lower deviations indicate conformity to normal traffic patterns.

**Bidirectional Prediction Paradigm for NTAD**

Existing NTAD methods suffer from a paradigm-level limitation. Reconstruction-based pipelines are prone to triggering the "identical shortcut", which leads to anomaly overgeneralization. To overcome this, we reformulate NTAD as a bidirectional prediction problem that removes the shortcut learning by design. As illustrated in Figure 2(a), each traffic sample is symmetrically divided along the flow dimension into two segments: the first half serves as the history segment $\boldsymbol{x}_{\text{his}}$, while the second half serves as the future segment $\boldsymbol{x}_{\text{fut}}$.

$$\boldsymbol{x}_{\text{his}} = \boldsymbol{x}[:F/2, :], \quad \boldsymbol{x}_{\text{fut}} = \boldsymbol{x}[F/2:, :]. \tag{1}$$

The two segments are then processed by independent embedding layers to tokenize byte-level representations:

$$\boldsymbol{h}^0_{\text{his}} = \text{Embed}_{\text{his}}(\boldsymbol{x}_{\text{his}}), \boldsymbol{h}^0_{\text{fut}} = \text{Embed}_{\text{fut}}(\boldsymbol{x}_{\text{fut}}), \tag{2}$$

where $\text{Embed}(\cdot)$ comprises a Conv1D-based data embedding layer and a sine–cosine positional encoding layer [Li *et al.*, 2025; Chen *et al.*, 2026b]. The resulting embeddings encode sequential packet information at the byte level, yielding feature representations in $\mathbb{R}^{(\frac{F}{2} \cdot P) \times D}$. To explicitly preserve the inherent flow–packet hierarchy of network traffic, we further reshape them into $\boldsymbol{h}^0_{\text{his}}, \boldsymbol{h}^0_{\text{fut}} \in \mathbb{R}^{\frac{F}{2} \times P \times D}$, enabling subsequent hierarchical contextual modeling.

Normal and anomalous traffic exhibit distinct communication patterns [Yang *et al.*, 2025; Lian *et al.*, 2025b], which are reflected in their contextual dependencies across history and future segments. For normal traffic, these segments maintain stable and predictable cross-segment relationships. In contrast, anomalous traffic disrupts such dependencies, preventing the formation of coherent cross-segment context under zero-positive training. Motivated by this observation, we employ two Mamba-based encoders to capture contextual information from the input segments. One prediction branch estimates the future representation $\boldsymbol{h}^L_{\text{fut}}$ conditioned on the history segment $\boldsymbol{h}^0_{\text{his}}$, while the other predicts the history representation $\boldsymbol{h}^L_{\text{his}}$ conditioned on the future segment $\boldsymbol{h}^0_{\text{fut}}$.

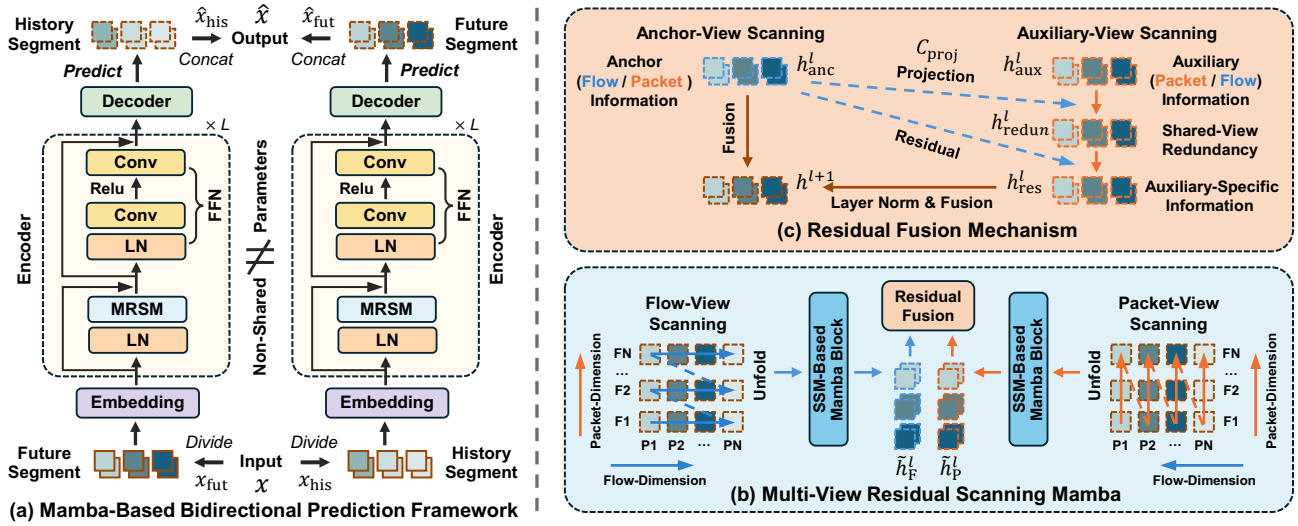


Figure 2: **Overview of our BiPred**. (a) **Bidirectional Prediction Framework**: BiPred reformulates NTAD as a bidirectional prediction pipeline. (b) **Multi-View Residual Scanning Mamba**: a dual-view scanning strategy is employed to encode multi-view contextual information. (c) **Residual Fusion Mechanism**: residual learning is applied to suppress inter-view redundancy and preserve multi-view context.

Lightweight decoders with linear layers are then used to map the predicted representations $\boldsymbol{h}^L_{\text{his}}$ and $\boldsymbol{h}^L_{\text{fut}}$ to the corresponding outputs $\hat{\boldsymbol{x}}_{\text{his}}$ and $\hat{\boldsymbol{x}}_{\text{fut}}$. Finally, the bidirectional prediction outputs are concatenated to form the overall prediction result $\hat{\boldsymbol{x}}$ for zero-positive learning. The bidirectional prediction process is formalized as follows:

$$\boldsymbol{h}^L_{\text{fut}} = \text{Encoder}_{\text{his}}(\boldsymbol{h}^0_{\text{his}}), \quad \boldsymbol{h}^L_{\text{his}} = \text{Encoder}_{\text{fut}}(\boldsymbol{h}^0_{\text{fut}}), \tag{3}$$

$$\hat{\boldsymbol{x}}_{\text{fut}} = \text{Decoder}_{\text{his}}(\boldsymbol{h}^L_{\text{fut}}), \quad \hat{\boldsymbol{x}}_{\text{his}} = \text{Decoder}_{\text{fut}}(\boldsymbol{h}^L_{\text{his}}), \tag{4}$$

$$\hat{\boldsymbol{x}} = \text{Concat}[\hat{\boldsymbol{x}}_{\text{his}}, \hat{\boldsymbol{x}}_{\text{fut}}], \tag{5}$$

where $L$ denotes the number of Mamba layers in the encoder.

Overall, our design philosophy goes beyond traditional reconstruction-based frameworks at both the paradigm and learning-mechanism levels. (1) The proposed bidirectional-prediction paradigm removes trivial input–output replication through its construction. By constraining data flow into two independent predictive paths, it prevents unintended information leakage between inputs and outputs, thereby fundamentally eliminating the "identical shortcut". (2) The framework establishes explicit contextual prediction dependencies, where each model is forced to infer the missing segment from partial observations. This design is particularly effective at exposing concealed anomalies manifested across segments.

### Multi-View Residual Scanning Mamba

In BiPred, strong contextual representations of $\text{Encoder}_{\text{his}}$ and $\text{Encoder}_{\text{fut}}$ are essential. Insufficient representation learning may cause the prediction paradigm to collapse and unreliable anomaly detection. Motivated by its strong sequential modeling capacity [Qin *et al.*, 2025], we adopt a Mamba-based backbone and design the Multi-View Residual Scanning Mamba (MRSM) module, as shown in Figure 2(b).

Existing traffic analysis works primarily employ unidirectional Mamba scanning [Deng *et al.*, 2025; Xia *et al.*, 2025]. However, unidirectional scanning would yield biased representations for $\boldsymbol{h}^0_{\text{his}}$ and $\boldsymbol{h}^0_{\text{fut}}$ [Qu *et al.*, 2024] and fail to capture the flow–packet structure of network traffic [Zhao *et al.*, 2024a]. Therefore, we design a multi-view scanning strategy. As illustrated in Figure 2(b), for both $\boldsymbol{h}^0_{\text{his}}$ and $\boldsymbol{h}^0_{\text{fut}}$ segments, we independently rearrange each representation into two complementary views: a flow view $\boldsymbol{h}^0_{\text{F}} \in \mathbb{R}^{(\frac{F}{2}\cdot P)\times D}$ and a packet view $\boldsymbol{h}^0_{\text{P}} \in \mathbb{R}^{(P\cdot\frac{F}{2})\times D}$. These views are processed by shared Mamba blocks [Gu and Dao, 2024] to capture multi-view contextual representations $\tilde{\boldsymbol{h}}^0_{\text{F}}$ and $\tilde{\boldsymbol{h}}^0_{\text{P}}$. The $l$-th MRSM module is represented as follows:

$$\tilde{\boldsymbol{h}}^l_{\text{F}} = \text{MambaBlock}(\text{LN}(\boldsymbol{h}^l_{\text{F}})), \tag{6}$$

$$\tilde{\boldsymbol{h}}^l_{\text{P}} = \text{MambaBlock}(\text{LN}(\boldsymbol{h}^l_{\text{P}})), \tag{7}$$

where LN denotes the layer normalization.

While multi-view scanning provides effective contextual representations, its application to network traffic presents significant challenges. Naive fusion of the scanning branches can repeatedly aggregate inter-view redundancy, causing the multi-view representation to degenerate toward a single branch [Wang *et al.*, 2024]. To quantify this effect, we measure the linear CKA similarity between the view representations [Tran *et al.*, 2025]. As shown in Figure 3, the similarity values of the Anchor–Auxiliary scheme, i.e., between the flow-view and the packet-view, remain high across three datasets, indicating substantial redundancy.

To address the inter-view redundancy, we propose a novel residual fusion mechanism, as shown in Figure 2(c). At each layer, the flow view $\tilde{\boldsymbol{h}}^l_{\text{F}}$ and the packet view $\tilde{\boldsymbol{h}}^l_{\text{P}}$ are alternately assigned as the anchor view $\boldsymbol{h}^l_{\text{anc}}$ and the auxiliary view $\boldsymbol{h}^l_{\text{aux}}$. For example, the flow view $\tilde{\boldsymbol{h}}^l_{\text{P}}$ serves as the anchor at the $l$-th layer, while the packet view $\tilde{\boldsymbol{h}}^{l+1}_{\text{P}}$ takes the anchor role at the $(l+1)$-th layer. The anchor view encodes the primary scanning information, comprising shared-view redundancy and anchor-specific components. For the

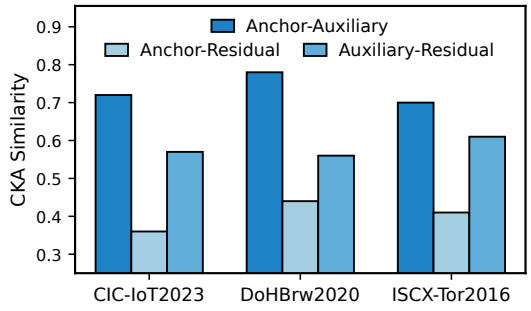


Figure 3: Statistics of redundancy in multi-view scanning.

auxiliary view, only the auxiliary-specific information is retained for merging with the anchor view. This residual fusion mechanism explicitly mitigates inter-view redundancy by removing the shared components from the auxiliary view. To achieve this, we first compute the projection coefficient vector $\boldsymbol{C}_{\text{proj}}$ of the auxiliary view with respect to the anchor view via element-wise vector projection. The shared-view redundancy $\boldsymbol{h}^l_{\text{redun}}$ is then obtained by projecting the auxiliary representation onto the anchor representation using $\boldsymbol{C}_{\text{proj}}$. Finally, the auxiliary-specific information is extracted by subtracting the shared redundancy, yielding the residual representation $\boldsymbol{h}^l_{\text{res}}$. This process is summarized as follows:

$$\boldsymbol{C}_{\text{proj}} = \frac{\boldsymbol{h}^l_{\text{aux}} \odot \boldsymbol{h}^l_{\text{anc}}}{\|\boldsymbol{h}^l_{\text{anc}}\|^2 + \varepsilon}, \tag{8}$$

$$\boldsymbol{h}^l_{\text{redun}} = \boldsymbol{C}_{\text{proj}} \boldsymbol{h}^l_{\text{anc}}, \tag{9}$$

$$\boldsymbol{h}^l_{\text{res}} = \boldsymbol{h}^l_{\text{aux}} - \boldsymbol{h}^l_{\text{redun}}, \tag{10}$$

where $\varepsilon$ is a small constant added for numerical stability.

Our multi-view representation is formed by combining the anchor view with the residual-specific information extracted from the auxiliary view. As confirmed in Figure 3, the proposed Anchor–Residual branch substantially reduces redundancy compared to the naive Anchor–Auxiliary branch. Unlike existing approaches that directly sum scanning results [Qu *et al.*, 2024], this residual-based formulation explicitly suppresses shared-view redundancy, preventing representation degradation. Moreover, it preserves complementary multi-view cues, as reflected by the moderate Auxiliary–Residual branch similarity.

$$\boldsymbol{h}^{l+1} = \text{FFN}(\boldsymbol{h}^l_{\text{anc}} + \text{LN}(\boldsymbol{h}^l_{\text{res}})), \tag{11}$$

where FFN denotes a feed-forward network implemented with two Conv1D layers [Ge *et al.*, 2025; Chen *et al.*, 2026a].

### 3.3 Training

Our objective of zero-positive training is to learn normal bidirectional prediction patterns. To this end, BiPred adopts cosine similarity as the objective $\mathcal{L}$ to enforce prediction consistency. Specifically, given an input $\boldsymbol{x} = \text{Concat}[\boldsymbol{x}_{\text{his}}, \boldsymbol{x}_{\text{fut}}]$ and prediction $\hat{\boldsymbol{x}} = \text{Concat}[\hat{\boldsymbol{x}}_{\text{his}}, \hat{\boldsymbol{x}}_{\text{fut}}]$, the objective $\mathcal{L}$ encourages the predicted history $\hat{\boldsymbol{x}}_{\text{his}}$ (conditioned on $\boldsymbol{x}_{\text{fut}}$) to match the original history $\boldsymbol{x}_{\text{his}}$, and the predicted future $\hat{\boldsymbol{x}}_{\text{fut}}$ (conditioned on $\boldsymbol{x}_{\text{his}}$) to match the original future $\boldsymbol{x}_{\text{fut}}$:

$$\mathcal{L} = \sum_{i=1}^{N} (1 - \frac{\boldsymbol{x}_i^{\boldsymbol{T}}\, \hat{\boldsymbol{x}}_i}{\|\boldsymbol{x}_i\| \|\hat{\boldsymbol{x}}_i\|}). \tag{12}$$

### 3.4 Anomaly Scoring

Consistent with our design philosophy, the model produces accurate bidirectional predictions for normal traffic after zero-positive training. During inference, previously unseen anomalous traffic tends to violate the learned contextual dependencies, resulting in degraded prediction quality. Such deviations are reflected by reduced similarity between the predicted output and the original input. Accordingly, we define the anomaly score as the cosine dissimilarity [Fu *et al.*, 2024] between the prediction $\hat{\boldsymbol{x}}$ and the original traffic sample $\boldsymbol{x}$:

$$s = 1 - \frac{\boldsymbol{x}^T\, \hat{\boldsymbol{x}}}{\|\boldsymbol{x}\| \|\hat{\boldsymbol{x}}\|}. \tag{13}$$

## 4 Experiments

### 4.1 Experimental Setup

**Datasets.** We evaluate on three widely used network traffic anomaly detection datasets. (1) *CIC-IoT2023*: an intrusion detection dataset for Internet of Things network topologies [Neto *et al.*, 2023]. (2) *DoHBrw2020*: a malicious HTTPS traffic detection dataset [MontazeriShatoori *et al.*, 2020]. (3) *ISCX-Tor2016*: an anonymous network traffic detection dataset [Lashkari *et al.*, 2017]. We sample 10,000 normal flows for training and use a balanced test set with 5,000 normal and 5,000 anomalous flows.

**Evaluation Metrics.** We assess model performance using three standard metrics: Area Under Curve (AUC), Accuracy (ACC), and F1-Score (F1), following recent studies [Zheng *et al.*, 2023; Lian *et al.*, 2025a].

**Baselines.** We compare our method with representative approaches from three categories. (1) *Network traffic* anomaly detection methods: DAGMM [Zong *et al.*, 2018], ARCADE [Lunardi *et al.*, 2023], MFR [Lian *et al.*, 2025b], UnDiff [Lian *et al.*, 2025a] and ConMD [Lian *et al.*, 2026b]. (2) *Transformer-based* anomaly detection methods: Anomaly Transformer [Xu *et al.*, 2022], TSLANet [Eldele *et al.*, 2024], and iTransformer [Liu *et al.*, 2024]. (3) *Mamba-based* anomaly detection methods: CountMamba [Deng *et al.*, 2025] and MambaAD [Qin *et al.*, 2025].

**Implementation Details.** All experiments are conducted on an NVIDIA GeForce RTX 4090 GPU. We set the number of residual Mamba blocks $L$ to 3 and the token embedding dimension $D$ to 64. For traffic representation, we set $F = 6$ packets per flow and $P = 400$ bytes per packet. The Adam optimizer is used with a learning rate of $1 \times 10^{-4}$ and a batch size of 128. We conduct five independent runs with different random seeds and report the average results.

### 4.2 Main Results

We compare BiPred with ten strong baselines to validate its effectiveness. The results are presented in Table 1. Based on these results, we make the following observations:

**(O1)**: BiPred consistently outperforms all baselines across datasets and evaluation metrics by clear margins. In particular, BiPred achieves AUC improvements of 3.16% on CIC-IoT2023 and 4.21% on DoHBrw2020. These gains stem from the proposed prediction-based anomaly detection

| Methods | CIC-IoT2023 | | | DoHBrw2020 | | | ISCX-Tor2016 | | |
|---|---|---|---|---|---|---|---|---|---|
| | AUC(%) | ACC(%) | F1(%) | AUC(%) | ACC(%) | F1(%) | AUC(%) | ACC(%) | F1(%) |
| DAGMM | 66.40 | 61.38 | 67.22 | 77.41 | 75.18 | 75.07 | 73.28 | 74.30 | 73.82 |
| ARCADE | 72.09 | 77.32 | 74.94 | 76.79 | 79.91 | 80.07 | 92.35 | 90.15 | 90.50 |
| MFR | 82.68 | 83.03 | 81.56 | 70.51 | 70.23 | 73.44 | 93.68 | 91.74 | 91.94 |
| UnDiff | 83.25 | 82.69 | 81.93 | 81.92 | 79.93 | 80.10 | 94.01 | 91.86 | 91.61 |
| ConMD | 83.76 | 83.11 | 81.80 | 83.44 | 80.14 | 82.09 | 94.35 | 91.79 | 91.23 |
| Anomaly Transformer | 64.63 | 67.15 | 71.86 | 78.86 | 78.55 | 78.87 | 92.27 | 90.19 | 90.52 |
| TSLANet | 81.64 | 76.62 | 80.04 | 77.04 | 68.93 | 73.10 | 93.63 | 90.97 | 90.71 |
| iTransformer | 81.32 | 82.07 | 81.95 | 82.44 | 78.66 | 78.26 | 82.65 | 81.34 | 83.97 |
| MambaAD | 82.29 | 77.70 | 80.02 | 82.79 | 79.36 | 81.85 | 92.54 | 85.21 | 85.48 |
| CountMamba | 75.38 | 68.37 | 74.10 | 80.45 | 78.68 | 81.89 | 92.90 | 89.32 | 89.80 |
| **BiPred** | **86.92** | **84.94** | **83.74** | **87.65** | **83.52** | **84.11** | **96.54** | **93.05** | **93.12** |

Table 1: Performance comparisons (%) on three public datasets. The best results are in **bold** font and the second underlined.

| Variant | IoT2023 | DoHBrw2020 | Tor2016 |
|---|---|---|---|
| w/o Prediction | 83.93 | 84.35 | 94.21 |
| w/o Residual | 85.72 | 86.50 | 95.82 |
| w/o Multi-View | 85.40 | 86.29 | 95.63 |
| w/o Mamba | 83.64 | 81.83 | 94.01 |
| **BiPred** | **86.92** | **87.65** | **96.54** |

Table 2: Ablation results on main components (AUC).

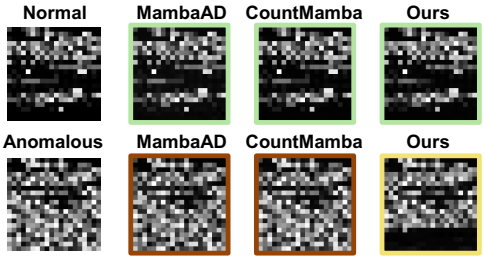


Figure 4: Visual comparison with reconstruction-based paradigms.

paradigm. By utilizing bidirectional-prediction, BiPred fundamentally eliminates the "identical shortcut" issue and establishes a contextual prediction pipeline. Furthermore, our Residual Scanning Mamba effectively mitigates inter-view redundancy, encoding complementary contextual information for more stable predictions.

**(O2)**: Existing NTAD methods, such as ConMD and UnDiff, attempt to alleviate the "identical shortcut" issue through feature extraction or alternative anomaly scoring strategies. However, their modeling pipelines remain constrained within the reconstruction paradigm, where shortcut learning is still inherently present. In addition, most of these methods rely on CNN-based architectures, which reduces their sensitivity for long-range contextual anomalies.

**(O3)**: Transformer-based approaches, including Anomaly Transformer, TSLANet, and iTransformer, perform worse than several specialized NTAD methods. This is because these models are primarily designed for time-series data, where anomalies are characterized by trend and fluctuations. In contrast, network traffic is often payload-encrypted, which obscures exploitable time-series characteristics. Furthermore, these approaches also adopt a reconstruction-based pipeline, leaving them vulnerable to the "identical shortcut" problem. Similarly, recent Mamba-based methods such as MambaAD and CountMamba show limited effectiveness.

## 4.3 Ablation Study

We evaluate four ablated variants by removing key components. The results are reported in Table 2, and the main observations are summarized as follows:

**Effect of Prediction Pipeline:** Replacing the prediction-based pipeline with a conventional reconstruction pipeline (w/o Prediction) leads to a substantial performance drop. This clearly validates the effectiveness of our paradigm. By eliminating the "identical shortcut" and enforcing cross-contextual dependency modeling, the prediction pipeline enables more discriminative anomaly awareness.

**Effect of Residual Mamba:** Removing the proposed Residual Scanning Mamba module (w/o Residual) results in noticeable degradation. Combined with the statistical results in Figure 3, this demonstrates that Residual Mamba effectively mitigates inter-view redundancy through residual fusion. By highlighting complementary view-specific information, it strengthens multi-view scanning representations.

**Effect of Multi-View Scanning:** Discarding multi-view scanning (w/o Multi-View) degrades performance, indicating that a single scanning direction is insufficient to capture the intrinsic multi-view characteristics of network traffic. Notably, single-view scanning (w/o Multi-View) achieves performance comparable to multi-view scanning without residual fusion (w/o Residual), further confirming the necessity of residual fusion for releasing multi-view information.

**Effect of Mamba Architecture:** Replacing the Mamba backbone with CNNs (w/o Mamba) causes significant performance deterioration. This confirms that contextual modeling is critical for the prediction framework. CNN-based architectures lack the capability to capture long-range contextual dependencies required for accurate bidirectional prediction, severely limiting the effectiveness of the paradigm.

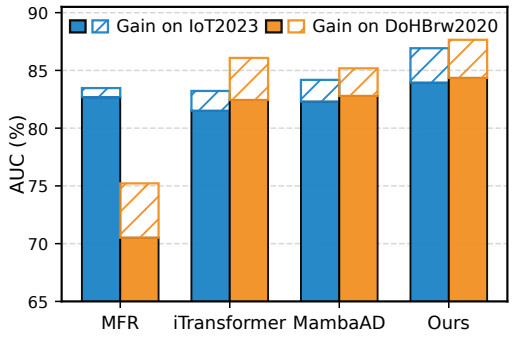


Figure 5: Paradigm scalability study. Solid bars show performance under the reconstruction paradigm, while dashed bars indicate gains achieved with the prediction paradigm.

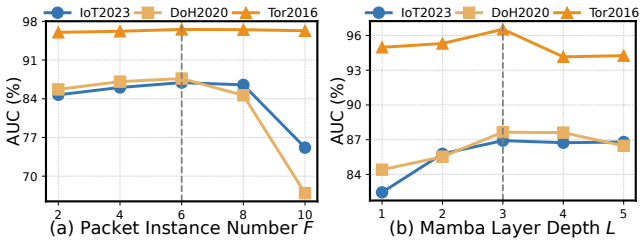


Figure 6: Hyperparameter sensitivity analysis across three datasets.

## 4.4 Paradigm Comparison

Figure 4 empirically demonstrates the superiority of our prediction-based paradigm. In reconstruction-based methods, the "identical shortcut" is easily triggered, causing models to learn a trivial identity mapping regardless of input normality [Lian *et al.*, 2025a]. As observed in MambaAD and CountMamba, anomalous traffic (red rectangles) can be reconstructed with fidelity comparable to normal samples (green rectangles), violating the core assumption of anomaly detection. In contrast, our prediction-based framework fundamentally avoids this shortcut. Normal samples (green rectangle) yield outputs closely matching the input, whereas anomalous traffic (yellow rectangle) produces significant prediction deviations. This clearer anomaly separability qualitatively validates the advances and effectiveness of our paradigm.

## 4.5 Paradigm Scalability

To assess paradigm scalability, we replace the reconstruction-based pipelines in several representative baselines with our prediction-based pipeline. The results, shown in Figure 5, demonstrate that our BiPred exhibits strong scalability. All replaced baselines achieve substantial performance gains, confirming the generality and effectiveness of the bidirectional prediction paradigm. These observations indicate that the benefits of shortcut elimination and enhanced contextual modeling can be naturally transferred to existing methods.

## 4.6 Hyperparameter Analysis

Figure 6 presents the sensitivity analysis of two key hyperparameters in BiPred. The observations are as follows:

**(O1)**: The parameter $F$ controls the amount of flow-view contextual information used for prediction. A relatively small $F$ is sufficient to establish effective flow-level context and an intrinsic predictive pipeline. In contrast, excessively large $F$ degrades detection performance.

**(O2)**: The parameter $L$ determines the contextual representation capacity of the Mamba encoder in the prediction pipeline. Our results show that performance stabilizes for $L \geq 3$, so we set $L = 3$ for both prediction branches to trade off contextual representation and computational efficiency.

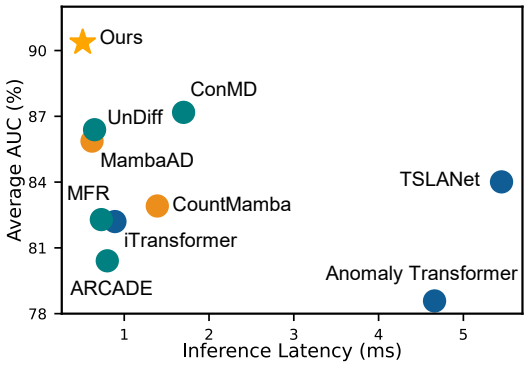


Figure 7: Inference overhead comparison with baseline methods.

## 4.7 Overhead Analysis

Detection overhead is a critical consideration for the practical deployment of NTAD systems. Figure 7 compares the average AUC across three datasets and the inference latency of competitive approaches. While CNN-based methods have linear complexity $\mathcal{O}((F \cdot P)D)$, they rely on extensive feature engineering, which introduces non-negligible overhead. Transformer-based approaches suffer from quadratic complexity $\mathcal{O}((F \cdot P)^2 D)$, leading to substantial computational costs. In contrast, Mamba-based architectures exhibit linear complexity $\mathcal{O}((F \cdot P)D)$, providing significantly higher efficiency. This advantage enables the practical deployment of our bidirectional-prediction paradigm and multi-view scanning mechanism. Overall, compared with all baselines, our method achieves both superior detection performance and millisecond-level inference efficiency.

# 5 Conclusions

In this work, we propose BiPred, the first bidirectional prediction-based paradigm for the NTAD task. From a conceptual paradigm perspective, we move beyond conventional reconstruction-based frameworks by eliminating the inherent risk of the "identical shortcut" and reformulating anomaly detection as a contextual prediction problem. This shift establishes explicit bidirectional contextual dependencies and enables more sensitive and reliable anomaly detection. From a methodological perspective, we introduce a novel Multi-View Residual Scanning Mamba architecture to explicitly address inter-view redundancy in multi-view scanning. The proposed design preserves complementary multi-view information and effectively supports contextual prediction. Extensive experimental results demonstrate the effectiveness of the BiPred. Overall, this work establishes a prediction-based paradigm for NTAD, opening new opportunities for future research.

## Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 62572097 and No. 62176043).

## References

[Cai *et al.*, 2026] Jiati Cai, Yuxun Zhao, Yan Liu, Ting Zhong, and Fan Zhou. Scope-diff: A spatiotemporal comprehensive perception diffusion framework for video anomaly detection. *Neurocomputing*, 678:133157, 2026.

[Chen *et al.*, 2026a] Bin Chen, Hanting Shen, Zhangtao Cheng, Xueting Liu, Ting Zhong, and Fan Zhou. Unveiling cross-modal consistency: Taming inter- and intra-modal noise for robust multi-modal knowledge graph completion. *Information Processing & Management*, 63(2):104472, 2026.

[Chen *et al.*, 2026b] Mengyang Chen, Lingwei Wei, Wei Zhou, and Songlin Hu. An information-theoretic propagation denoising and fusion framework for fake news detection. *arXiv: 2605.02259*, 2026.

[Dang *et al.*, 2025] Zhangxuan Dang, Yu Zheng, Xinglin Lian, Chunlei Peng, Qiuyu Chen, and Xinbo Gao. Semi-supervised learning for anomaly traffic detection via bidirectional normalizing flows. *IEEE Transactions on Network and Service Management*, 22(5):5106–5117, 2025.

[Deng *et al.*, 2025] Xianwen Deng, Ruijie Zhao, Yanhao Wang, Mingwei Zhan, Zhi Xue, and Yijun Wang. Countmamba: A generalized website fingerprinting attack via coarse-grained representation and fine-grained prediction. In *IEEE Symposium on Security and Privacy*, pages 1419–1437, 2025.

[Eldele *et al.*, 2024] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, and Xiaoli Li. Tslanet: Rethinking transformers for time series representation learning. In *International Conference on Machine Learning*, pages 12409–12428, 2024.

[Fu *et al.*, 2024] Jiadong Fu, Jiang Fang, Jiyan Sun, Shangyuan Zhuang, Liru Geng, and Yinlong Liu. Loft: Lora-based efficient and robust fine-tuning framework for adversarial training. In *International Joint Conference on Neural Networks*, pages 1–8, 2024.

[Fu *et al.*, 2026a] Jiadong Fu, Jiang Fang, Jiyan Sun, Laile Xi, Shangyuan Zhuang, Yinlong Liu, and ZhiQiang Lv. Capt: A lightweight continual adversarial pre-training framework for traffic analysis model. In *IEEE International Conference on Acoustics, Speech and Signal ProcessingP*, pages 14367–14371, 2026.

[Fu *et al.*, 2026b] Jiadong Fu, Jiang Fang, Jiyan Sun, Shangyuan Zhuang, Yinlong Liu, and Zhiqiang Lv. Toward efficient distributed network security: A lightweight multitask traffic analysis framework. *IEEE Transactions on Networking*, 34:2271–2286, 2026.

[Ge *et al.*, 2025] Yunfeng Ge, Jiawei Li, Yiji Zhao, Haomin Wen, Zhao Li, Meikang Qiu, Hongyan Li, Ming Jin, and Shirui Pan. T2S: high-resolution time series generation with text-to-series diffusion models. In *International Joint Conference on Artificial Intelligence*, pages 5208–5216, 2025.

[Gu and Dao, 2024] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *Conference on Language Mdeling*, 2024.

[Guo *et al.*, 2024] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European Conference on Computer Vision*, pages 222–241, 2024.

[He *et al.*, 2026] Yaru He, Jiaqi Gao, Daoqi Han, Yueming Lu, and Yaojun Qiao. Hierarchical traffic fingerprint-assisted graph neural networks for encrypted traffic classification in internet of vehicles. *Knowledge-Based Systems*, 343:116027, 2026.

[Lashkari *et al.*, 2017] Arash Habibi Lashkari, Gerard Draper-Gil, Mohammad Saiful Islam Mamun, and Ali A. Ghorbani. Characterization of tor traffic using time based features. In *ICISSP*, pages 253–262. SciTePress, 2017.

[Li *et al.*, 2023] Yuxin Li, Wenchao Chen, Bo Chen, Dongsheng Wang, Long Tian, and Mingyuan Zhou. Prototype-oriented unsupervised anomaly detection for multivariate time series. In *International Conference on Machine Learning*, pages 19407–19424, 2023.

[Li *et al.*, 2025] Zetao Li, Zheng Hu, Peng Han, Yu Gu, and Shimin Cai. Ssl-stmformer self-supervised learning spatio-temporal entanglement transformer for traffic flow prediction. In *AAAI Conference on Artificial Intelligence*, pages 12130–12138, 2025.

[Lian *et al.*, 2024] Xinglin Lian, Yang Liu, Shanfeng Wang, and Yu Zheng. Payload level anomaly network traffic detection via semi-supervised contrastive learning. In *Trust, Security and Privacy in Computing and Communications*, pages 2559–2566, 2024.

[Lian *et al.*, 2025a] Xinglin Lian, Chengtai Cao, Yan Liu, Xovee Xu, Yu Zheng, and Fan Zhou. Facing anomalies head-on: Network traffic anomaly detection via uncertainty-inspired inter-sample differences. In *the ACM on Web Conference*, pages 3908–3917, 2025.

[Lian *et al.*, 2025b] Xinglin Lian, Yu Zheng, Zhangxuan Dang, Chunlei Peng, and Xinbo Gao. Semi-supervised anomaly traffic detection via multi-frequency reconstruction. *Pattern Recognition*, 161:111215, 2025.

[Lian *et al.*, 2026a] Xinglin Lian, Chengtai Cao, Ting Zhong, Yong Wang, Kai Chen, and Fan Zhou. Decompose to understand, fuse to detect: Frequency-decoupled anomaly detection for encrypted network traffic. *arXiv: 2605.02970*, 2026.

[Lian *et al.*, 2026b] Xinglin Lian, Yu Zheng, Yan Liu, Fan Zhou, Chunlei Peng, and Xinbo Gao. Contextual masking distillation for network traffic anomaly detection. *IEEE Transactions on Information Forensics and Security*, 21:1273–1286, 2026.

[Liu *et al.*, 2024] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long.

itransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 2024.

[Lunardi *et al.*, 2023] Willian Tessaro Lunardi, Martin Andreoni Lopez, and Jean Pierre Giacalone. ARCADE: adversarially regularized convolutional autoencoder for network anomaly detection. *IEEE Transactions on Network and Service Management*, 20(2):1305–1318, 2023.

[MontazeriShatoori *et al.*, 2020] Mohammadreza MontazeriShatoori, Logan Davidson, Gurdip Kaur, and Arash Habibi Lashkari. Detection of doh tunnels using time-series classification of encrypted traffic. In *IEEE DASC/PiCom/CBDCom/CyberSciTech*, pages 63–70, 2020.

[Neto *et al.*, 2023] Euclides Carlos Pinto Neto, Sajjad Dadkhah, Raphael Ferreira, Alireza Zohourian, Rongxing Lu, and Ali A. Ghorbani. Ciciot2023: A real-time dataset and benchmark for large-scale attacks in iot environment. *Sensors*, 23(13):5941, 2023.

[Patro and Agneeswaran, 2025] Badri Narayana Patro and Vijay Srinivas Agneeswaran. Mamba-360: Survey of state space models as transformer alternative for long sequence modelling: Methods, applications, and challenges. *Engineering Applications of Artificial Intelligence*, 159:111279, 2025.

[Qin *et al.*, 2025] Shuxin Qin, Jing Zhu, Aipeng Guo, Yansong Yang, Lu Wang, and Gaofeng Tao. Mambaad: Multivariate time series anomaly detection in iot via multi-view mamba. *Neurocomputing*, 655:131385, 2025.

[Qu *et al.*, 2024] Haohao Qu, Liangbo Ning, Rui An, Wenqi Fan, Tyler Derr, Hui Liu, Xin Xu, and Qing Li. A survey of mamba. *arXiv preprint arXiv:2408.01129*, 2024.

[Shi *et al.*, 2024] Yuheng Shi, Minjing Dong, and Chang Xu. Multi-scale vmamba: Hierarchy in hierarchy visual state space model. In *Annual Conference on Neural Information Processing Systems*, pages 25687–25708, 2024.

[Tran *et al.*, 2025] Hoan My Tran, Damien Lolive, Aghilas Sini, Arnaud Delhay, Pierre-François Marteau, and David Guennec. Multi-level ssl feature gating for audio deepfake detection. In *ACM International Conference on Multimedia*, pages 11766–11775, 2025.

[Wang *et al.*, 2024] Tongze Wang, Xiaohui Xie, Wenduo Wang, Chuyi Wang, Youjian Zhao, and Yong Cui. Netmamba: Efficient network traffic classification via pre-training unidirectional mamba. In *International Conference on Network Protocols*, pages 1–11, 2024.

[Xia *et al.*, 2025] Linhan Xia, Mingzhan Yang, Jingjing Wang, Ziwei Yan, Yakun Ren, Guo Yu, and Kai Lei. Mamba4net: Distilled hybrid mamba large language models for networking. In *International Conference on Network Protocols*, pages 1–10, 2025.

[Xiao *et al.*, 2025] Haoke Xiao, Lv Tang, Peng-Tao Jiang, Hao Zhang, Jinwei Chen, and Bo Li. Boosting vision state space model with fractal scanning. In *AAAI Conference on Artificial Intelligence*, pages 8646–8654, 2025.

[Xu *et al.*, 2022] Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly transformer: Time series anomaly detection with association discrepancy. In *International Conference on Learning Representations*, 2022.

[Yang *et al.*, 2025] Luming Yang, Yongjun Wang, Lin Liu, Junjie Huang, Jiangyong Shi, Shaojing Fu, and Shize Guo. unflows: An unsupervised construction scheme of flow spectrum for network traffic detection. *IEEE Transactions on Information Forensics and Security*, 20:3330–3345, 2025.

[Zhang *et al.*, 2022] Lianming Zhang, Xiaowei Xie, Kai Xiao, Wenji Bai, Kui Liu, and Pingping Dong. Manomaly: Mutual adversarial networks for semi-supervised anomaly detection. *Information Sciences*, 611:65–80, 2022.

[Zhang *et al.*, 2024] Xinchen Zhang, Running Zhao, Zhihan Jiang, Zhicong Sun, Yulong Ding, Edith C. H. Ngai, and Shuang-Hua Yang. AOC-IDS: autonomous online framework with contrastive learning for intrusion detection. In *IEEE Conference on Computer Communications*, pages 581–590, 2024.

[Zhang *et al.*, 2025a] Haozhen Zhang, Haodong Yue, Xi Xiao, Le Yu, Qing Li, Zhen Ling, and Ye Zhang. Revolutionizing encrypted traffic classification with mh-net: A multi-view heterogeneous graph model. In *AAAI Conference on Artificial Intelligence*, pages 1048–1056, 2025.

[Zhang *et al.*, 2025b] Qiao Zhang, Mingwen Shao, Xinyuan Chen, Xiang Lv, and Kai Xu. Wave-mambaad: Wavelet-driven state space model for multi-class unsupervised anomaly detection. In *IEEE/CVF International Conference on Computer Vision*, pages 20868–20877, 2025.

[Zhao *et al.*, 2024a] Ruijie Zhao, Mingwei Zhan, Xianwen Deng, Fangqi Li, Yanhao Wang, Yijun Wang, Guan Gui, and Zhi Xue. A novel self-supervised framework based on masked autoencoder for traffic classification. *IEEE/ACM Transactions on Networking*, 32(3):2012–2025, 2024.

[Zhao *et al.*, 2024b] Ziming Zhao, Zhaoxuan Li, Zhuoxue Song, Wenhao Li, and Fan Zhang. Trident: A universal framework for fine-grained and class-incremental unknown traffic detection. In *the ACM on Web Conference*, pages 1608–1619, 2024.

[Zheng *et al.*, 2023] Yu Zheng, Xinglin Lian, Zhangxuan Dang, Chunlei Peng, Chao Yang, and Jianfeng Ma. A semi-supervised anomaly network traffic detection framework via multimodal traffic information fusion. In *International Conference on Information and Knowledge Management*, pages 4455–4459, 2023.

[Zheng *et al.*, 2024] Yu Zheng, Zhangxuan Dang, Xinglin Lian, Chunlei Peng, and Xinbo Gao. Multi-view multi-label network traffic classification based on mlp-mixer neural network. *Computer Networks*, 253:110746, 2024.

[Zong *et al.*, 2018] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Dae-ki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International Conference on Learning Representations*, 2018.