Facing Anomalies Head-On: Network Traffic Anomaly Detection via Uncertainty-Inspired Inter-Sample Differences

Xinglin Lian
Xidian University
Xi'an, China
kenshin.lian24@gmail.com

Chengtai Cao
City University of Hong Kong
Hong Kong, China
chengtcao2-c@my.cityu.edu.hk

Yan Liu
University of Electronic Science and Technology of China
Chengdu, China
yan.liu@std.uestc.edu.cn

Xovee Xu
University of Electronic Science and Technology of China
Chengdu, China
xovee.xu@gmail.com

Yu Zheng*
Xidian University
Xi'an, China
yzheng@xidian.edu.cn

Fan Zhou*
University of Electronic Science and Technology of China
Chengdu, China
fan.zhou@uestc.edu.cn

Abstract

Network traffic anomaly detection is pivotal in cybersecurity, especially as data volume grows and security requirement intensifies. This study addresses critical limitations in existing reconstruction-based methods, which quantify anomalies relying on intra-sample differences and struggle to detect drifted anomalies. In response, we propose a novel approach, the Uncertainty-Inspired Inter-Sample Differences (UnDiff) method, which leverages model uncertainty to enhance anomaly detection capabilities, particularly in scenarios involving anomaly drift. By employing evidential learning, the UnDiff model gathers evidence to minimize uncertainty in normal network traffic, enhancing its ability to differentiate between normal and anomalous traffic. To overcome the limitations of intra-sample difference quantification in reconstruction-based methods, we propose a novel anomaly score based on inter-sample uncertainty deviation that directly quantifies the anomaly degree. Benefiting from a concise model design and parameterized uncertainty quantification, UnDiff achieves high efficiency. Extensive experiments on three benchmarks demonstrate UnDiff's superior performance in detecting both undrifted and drifted anomalies with minimal computational overhead.

CCS Concepts

• **Security and privacy** → **Intrusion detection systems**; • **Information systems** → **Traffic analysis**.

Keywords

Network Traffic Anomaly Detection; Uncertainty Quantification; Drifted Anomaly Detection; Zero-Positive Learning

*Corresponding Authors.

WWW '25, Sydney, NSW, Australia

ACM ISBN 979-8-4007-1274-6/25/04
https://doi.org/10.1145/3696410.3714621

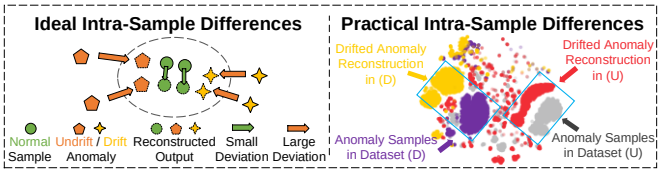
ACM Reference Format:
Xinglin Lian, Chengtai Cao, Yan Liu, Xovee Xu, Yu Zheng, and Fan Zhou. 2025. Facing Anomalies Head-On: Network Traffic Anomaly Detection via Uncertainty-Inspired Inter-Sample Differences. In *Proceedings of the ACM Web Conference 2025 (WWW '25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3696410.3714621

1 Introduction

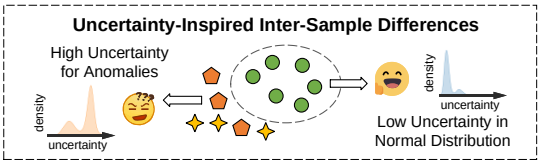
Network traffic anomaly detection, a fundamental component of cybersecurity infrastructure [53], plays a pivotal role in identifying malicious activities across various network environments. As data volumes surge exponentially and security requirements are stringent, precisely identifying anomalous network traffic patterns has emerged as a critical imperative. This capability underpins multiple important applications, including enhancing the stability and reliability of network services [13, 32] and fortifying personal privacy protection mechanisms [25, 30].

Current literature on network traffic anomaly detection predominantly employs a reconstruction-based "zero-positive learning" paradigm [5, 21, 28, 53], which only reconstructs normal network traffic distributions during the training phase, typically leveraging architectures such as auto-encoder models [29]. Subsequently, during the inference phase, common practice for evaluating anomaly degrees of network traffic is to utilize a distance-based metric [3, 15, 34, 48, 53, 54], i.e., samples exhibiting significant distance deviation between their pre- and post-reconstruction representations are identified as anomalous network traffic, while those demonstrating minimal divergence are considered as normal network traffic (cf. left part of Figure 1(a)).

Despite the recent advancements in reconstruction-based methods for network traffic anomaly detection, an intrinsic limitation persists. These approaches fully rely on intra-sample differences of pre- and post-reconstruction from an egocentric perspective while insufficiently leveraging inherent inter-sample differences, i.e., the diverse distribution between normal and anomalous network traffic [21]. This limitation is exacerbated by the potential "identical shortcut" issue in reconstruction models [45]. Instead of capturing differentiated characteristics of normal and anomalous patterns, reconstruction-based methods tend to converge on



(a) The limitations of existing reconstruction-based methods.



(b) The motion of our UnDiff.

Figure 1: Motivation for this work. (a) Existing methods encounter the "identical shortcut" issue, exemplified by the proximity of pre- and post-reconstruction drifted anomalies. (b) Our UnDiff is based on uncertainty-inspired inter-sample differences, facilitating direct anomaly identification.

a set of shortcut parameters that merely replicate the input as output [33, 35]. This limitation becomes particularly salient in detecting *drifted anomalies*, where the distribution of anomalous data evolves over time. The right part of Figure 1(a) visualizes the pre- and post-reconstruction embedding of drifted anomalies using a state-of-the-art reconstruction-based model Trident [53]. Empirical observations indicate that the pre- and post-reconstruction representations exhibit high proximity in the representation space. The intra-sample differences do not satisfy the ideal institution of reconstruction-based methods, thereby significantly impeding the discrimination of drifted anomalies.

To address this limitation, we propose a novel **Un**certainty-Inspired Inter-Sample **Diff**erences model (UnDiff). UnDiff leverages the concept of model uncertainty to enhance the discriminative capacity of anomaly detection systems, with particular emphasis on anomaly drift scenarios. As illustrated in Figure 1(b), the key intuition of our model is rooted in the differential uncertainty characteristics exhibited by normal and anomalous traffic patterns. Normal samples, well-represented in the training data, manifest low model uncertainty. Conversely, anomalous samples, particularly drifted anomalies, induce higher uncertainty due to their deviation from the learned normal patterns [21]. In contrast to existing reconstruction-based methods, our proposed UnDiff addresses the limitation above by directly facilitating inter-sample differences rather than relying on the sub-optimal intra-sample quantification.

Central to our UnDiff is a novel uncertainty learning module that quantifies model detection uncertainty. This module employs an evidential learning approach [6], acquiring evidence from training examples to construct an evidential distribution, facilitating robust uncertainty modeling for normal network traffic. Furthermore, we introduce explicit objectives to minimize uncertainty in normal network traffic during training. These objectives provide a more pronounced separation between normal and anomalous samples in the uncertainty space. To overcome the limitation of intra-sample disparity quantification in reconstruction-based methods, we further propose an innovative uncertainty-inspired anomaly score that adequately leverages inter-sample distributional differences for detecting anomalies. Notably, thanks to the concise design of the uncertainty learning module and the efficient parameterized uncertainty quantification technique, the enhancements we proposed above have negligible additional computational overhead. We conduct extensive experiments on real-world encrypted anomaly traffic datasets and evaluate the performance of UnDiff in both undrifted and drifted anomaly detections. Empirical results verified the effectiveness of our proposed model in detection performance across both scenarios. In summary, our key contributions are threefold:

- We propose a novel uncertainty-based evidential detection framework from an inter-sample difference perspective. Unlike the suboptimal intra-sample difference quantification in existing methods, our approach better utilizes the prior knowledge that anomalies inherently deviate from normal patterns, achieving more effective anomaly detection, particularly in scenarios involving anomaly drift.
- We introduce an innovative uncertainty learning module and a new anomaly score. This module provides an efficient and robust method for capturing sample uncertainty, while the anomaly score effectively quantifies inter-sample differences, significantly enhancing the discriminative capacity of the detection system.
- We conduct comprehensive empirical evaluations on three real-world anomaly network traffic datasets. The results demonstrate the effectiveness of our framework, UnDiff, in detecting both drifted and undrifted anomalies.

2 Related Work

2.1 Network Traffic Anomaly Detection

Anomaly detection, particularly zero-positive learning anomaly detection, has gained extensive attention. In this paradigm, only normal data are available during training, and samples that deviate from the learned model behavior are identified as anomalies during inference. Existing methods can be broadly categorized into three groups: distillation-based, statistics-based, and normalizing flow-based approaches [31]. Distillation-based methods focus on intra-sample differences, utilizing a student-teacher architecture to compare the distilled disparities [50, 51]. Conversely, the statistic-based [4, 11] and normalizing flow-based methods [19, 37] aim to learn a mapping from an input domain to a low-dimensional distribution. These approaches quantify inter-sample differences by analyzing deviations in the low-dimensional distribution. However, these methods, primarily designed for natural images, often encounter significant limitations when applied to traffic data. This is due to the unique characteristics of traffic images, such as redundant high-frequency information and disordered texture [29, 54].

Current network traffic anomaly detection methods mainly follow a reconstruction-based paradigm. These methods reconstruct the normal traffic during training and employ intra-sample differences (i.e., disparities between pre- and post-reconstruction) to identify anomalies. A notable example is GANomaly [3], a prominent reconstruction framework that utilizes a discriminator network to improve normal sample modeling. MANomaly [48] introduce a dual autoencoder adversarial training strategy to enhance representation learning, while ARCADE [34] employ WGAN-GP optimization for more effective adversarial training. MFR [29] and MFAD [54] identify a critical "identical shortcut" issue in traffic reconstruction

and utilize low-pass filtering to mitigate this problem. Trident [53] incorporates a U-Net structure to retain more detailed reconstruction information. Most anomaly detection methods for traffic data focus on enhancing the reconstruction quality of normal samples. However, these approaches often evade the "identical shortcut" issue inherent in reconstruction-based models. To overcome this limitation, we propose a novel paradigm based on inter-sample differences. In contrast to the suboptimal intra-sample differences employed by existing methods, we leverage the prior knowledge that anomalous samples inherently deviate from normal samples, achieving a more effective anomaly identification.

2.2 Uncertainty Learning

As deep learning models find increasingly widespread application across diverse domains, accuracy is no longer the only criterion for evaluation. In fields where safety is paramount, there is an urgent need for more trustworthy neural networks. Reliable uncertainty quantification emerges as a critical aspect in this context, as it measures the model's confidence in its output.

As elucidated in the literature [1, 9, 18, 26, 36], two primary categories of uncertainties are associated with neural networks: data uncertainty and model uncertainty. Data uncertainty arises from noise or randomness in the input and can be reduced to zero with sufficient training examples. For model uncertainty, Bayesian learning-based networks provide a mathematically grounded framework, albeit prohibitively expensive to implement and infer. Alternatively, Monte Carlo Dropout [17] approximates Bayesian inference on model parameters. Furthermore, leveraging the ensemble learning paradigm, Deep Ensemble [27] integrates multiple models for uncertainty estimation. To analyze data uncertainty, a unified Bayesian learning-based method [26] has been proposed to directly map input data to estimations of both data and model uncertainties. Uncertainty learning has also received attention in the field of anomaly detection, with approaches such as Bayesian learning [22] and its variational approximations [20, 23, 24, 46].

Recently, *evidential learning* has emerged as a promising uncertainty quantification approach [22, 39, 41]. This method enables uncertainty estimation in a single model and forward pass with parameterized distributions. In this approach, a neural network outputs the hyperparameters of an evidential distribution, allowing the model to estimate both model and data uncertainties without requiring sampling, thus enhancing the efficiency of uncertainty quantification [6]. However, most existing works on evidential learning are designed for supervised learning in computer vision [22] and necessitate large volumes of labeled data to estimate the uncertainty distribution. This requirement does not fit the typical anomaly detection setting. Therefore, in this study, we explore the application of evidential learning for quantifying the anomaly degree of network traffic in a zero-positive learning context.

3 Methodology

This section details our proposed uncertainty-inspired inter-sample difference method, UnDiff. We describe the research problem and introduce a novel research scenario, anomaly drift. Subsequently, we explicate the requisite data processing modules. We then detail our proposed uncertainty learning module, designed to learn the uncertainty space, thereby facilitating the comparison of inter-sample distribution differences. The schematic representation of our methodological pipeline is illustrated in Figure 2.

3.1 Problem Statement

Network Traffic Anomaly Detection. This work investigates the zero-positive learning anomaly detection problem in the context of network traffic analysis. Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ denote a set of N normal samples, where $\mathbf{x}_i \in \mathbb{R}^d$ is a d-dimensional data instance. The objective of detection models is to learn the distributional patterns of normal samples during training. For inference, the model assigns an anomaly score to each test sample $\mathbf{x}_{\text{test}} \in \mathcal{X}_{\text{test}}$, where $\mathcal{X}_{\text{test}}$ represents the set of test samples. The magnitude of the anomaly score is positively correlated with the likelihood of a sample being identified as anomalous.

Drifted Anomaly Detection. The dynamic nature of network activities frequently leads to divergence in the distribution of testing data, a phenomenon known as concept drift [15]. This drift often results in the performance degradation of anomaly detection systems [21, 54]. Existing research on concept drift in anomaly detection primarily focuses on two scenarios: *whole drift*, where both normal and anomalous data experience drift [7, 8, 44, 53] and *normal drift*, where only normal data undergoes drift [21]. However, this study addresses a more realistic scenario: *anomaly drift*, wherein only anomalous data experience drift. This scenario is particularly relevant because, in real-world applications, normal network traffic patterns typically exhibit relative stability, whereas anomalous network traffic patterns often change due to the evolution of attack strategies. Consequently, our research emphasizes the generalization capability of the anomaly detection model when confronted with drifts in the distribution of anomalous data.

3.2 Data Preprocessing

Network traffic fundamentally manifests as a flow format comprising an ordered sequence of packets. In contrast to statistical features designed based on manual heuristics [34, 48], we directly utilize the original traffic packet information for network traffic modeling. This approach circumvents the introduction of bias associated with manually crafted features. A critical consideration in the data processing of network flows is the appropriate representation method, as it significantly influences the detection accuracy and computational overhead. In this study, we employ a Multi-Channel Traffic Image Construction strategy for traffic flow representation. This approach allows for a more comprehensive and nuanced capture of the multidimensional nature of network traffic.

Multi-Channel Traffic Image Construction. While image-based single packet processing has been widely adopted in network traffic anomaly detection [14, 54], the potential of flow-level image construction remains largely unexplored. Drawing inspiration from video anomaly detection methodology [49], which addresses spatiotemporal representation tasks, we propose a novel approach to network flow representation. In our method, we formulate each network flow as a multi-channel image analogous to a video frame sequence. The specific channel order is determined by the packets' chronological arrival, preserving the flow's temporal dimension. This approach offers two significant advantages: (i) Dimensional

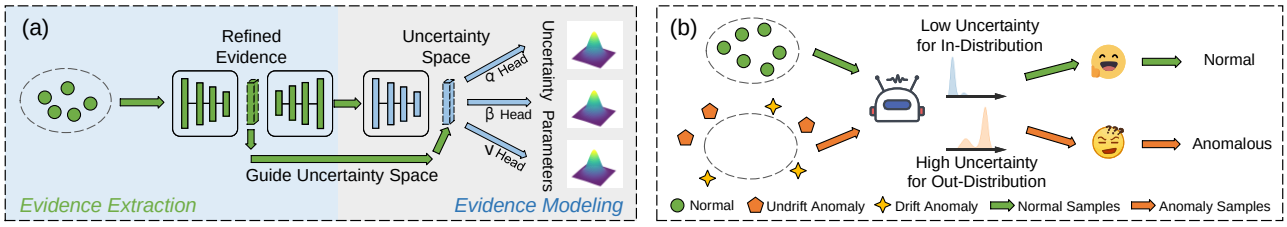


Figure 2: Overall framework of UnDiff. (a) Training: Uncertainty-Inspired Modeling Process. UnDiff first extracts informative representations from normal network traffic. These refined representations are then utilized to instruct uncertainty parameters for uncertainty representation. (b) Inference: Anomaly Metric Process. UnDiff directly outputs model uncertainty to quantify anomalies by assessing inter-sample differences via a distribution prior that differentiates normal and anomalous traffic.

Efficiency: By extending the representation along the channel dimension rather than width and height, we reduce the generation of subsequent high-dimensional feature maps. This design choice ensures enhanced inference speed. (ii) Informative Representation Preservation: Crucially, this approach adequately preserves informative representations as the contextual relationships between packets (represented as multi-channel images), reflecting the spatiotemporal characteristics of network flows. This temporal and spatial information preservation is critical for capturing the nuanced patterns that may indicate anomalies. For each traffic image, we implement a low-pass filtering process to mitigate noise. This step is necessitated by the unique characteristics of traffic images, which, in contrast to natural images, exhibit a chaotic and textureless state [29, 54]. This phenomenon arises from the abundance of high-frequency components inherent in network traffic data. However, these high-frequency components often manifest as detrimental noise, impeding the model's ability to generalize effectively due to the excessive complexity of the information.

3.3 Proposed UnDiff

As shown in Figure 2(a), our UnDiff contains two main components: an evidence extractor to extract evidence and a novel uncertainty learning module to construct an uncertainty space for subsequent anomaly quantification and detection.

3.3.1 Evidence Extractor. The autoencoder is designed to process original input traffic $\mathbf{x}$, encoding it into a latent evidence representation $\mathbf{z}$, and subsequently decoding it to produce a reconstruction of the input $\hat{\mathbf{x}}$. Given that the input is a multi-channel image, we employ Convolutional Neural Networks (CNNs) as the encoder, following established practices in the literature [34, 54]. It is imperative to note that network traffic inherently comprises a series of packet sequences with distinct spatiotemporal characteristics [29, 52]. However, previous autoencoder-based reconstruction methods have primarily focused on enhancing reconstruction quality, often neglecting the crucial spatiotemporal relationships inherent in the network packets. To address this limitation, we introduce a spatiotemporal aware channel-spatial based attention mechanism, specifically the Convolutional Block Attention Module (CBAM) [43], into our autoencoder architecture. This approach enables us to assign higher importance to significant channel images (temporal features) and spatially relevant regions (spatial features), thereby facilitating the extraction of evidence for critical patterns in multi-channel images.

3.3.2 Uncertainty Learning Module. The predominant zero-positive learning paradigm for anomaly network traffic typically frames this task as a reconstruction problem, optimizing the similarity loss of the original input $\mathbf{x}$ and the reconstructed output $\hat{\mathbf{x}}$. The intra-sample differences between pre- and post-reconstruction from an egocentric perspective are utilized to quantify the anomaly degrees. However, this paradigm exhibits suboptimal performance due to two primary limitations. Firstly, comparing the differences between samples before and after reconstruction does not directly address the fundamental nature of the problem: anomalous traffic is inherently defined relative to normal traffic patterns. Secondly, the classical "identical shortcut" problem inherent in autoencoder architectures significantly impacts the intra-sample differences of anomalous samples, particularly leading to performance degradation in scenarios involving anomaly drift.

To address these issues concurrently, we propose a novel uncertainty learning module designed to construct an uncertainty space, facilitating direct inter-sample comparisons to detect anomalous network traffic. This module is based on estimating the detection uncertainty, explicitly focusing on model uncertainty, also known as epistemic uncertainty. Model uncertainty quantifies the uncertainty in estimating model parameters given the training data, effectively measuring the degree of congruity between the model and the data [1]. We posit that this model uncertainty score is intrinsically linked to anomalous patterns and can be leveraged to identify anomalies effectively. The fundamental intuition underpinning our methodology is rooted in the differential uncertainty characteristics exhibited by normal and anomalous traffic patterns [21]. Normal samples, well-represented in the training data, typically manifest low model uncertainty. Conversely, anomalous traffic, particularly in the context of drifted anomalies, induces higher uncertainty due to its deviation from the learned normal patterns.

The uncertainty learning module comprises an encoder and a group of uncertainty parameter heads. The encoder, which shares its architectural design with the preceding encoder of the evidence extractor, is based on the reconstruction output $\hat{\mathbf{x}}$. It processes the reconstruction $\hat{\mathbf{x}}$ as input and generates an uncertainty representation γ, quantifying the model's detection uncertainty. The uncertainty parameter heads, implemented as linear layers, translate the uncertainty representation γ into their corresponding uncertainty parameters. This transformation facilitates effective uncertainty modeling. Through this mechanism, we explicitly incorporate evidential learning to quantify evidence distribution of normal network traffic. In

contrast to Bayesian Neural Networks (BNNs) [26], which place priors on network weights, our evidential-based approach sets priors directly over the likelihood function. This methodology achieves a more computationally efficient uncertainty quantification.

Network traffic data frequently follows Gaussian distributions during standard analysis scenarios [16]. Thus, we consider the uncertainty representations $\mathbf{z}$ extracted from the preceding autoencoder, which encapsulates the evidential information about normal network traffic, to conform to independent homogeneous distributions from a Gaussian distribution. These distributions are characterized by their mean and variance (μ, σ^2). These parameters to be quantified, μ and σ^2, are intrinsically linked to the model uncertainty that is the focus of our investigation [1]. To estimate these parameters, we employ a hierarchical Bayesian approach. Specifically, we utilize a Gaussian prior to estimate the mean value and place an Inverse-Gamma prior on the variance. This choice of priors is motivated by their conjugate relationship with the Gaussian likelihood, facilitating closed-form posterior updates. The hierarchical model can be expressed as follows:

$$\mathbf{z} \sim \mathcal{N}(\mu, \sigma^2) \qquad \mu \sim \mathcal{N}(\gamma, \sigma^2 v^{-1}) \qquad \sigma^2 \sim \Gamma^{-1}(\alpha, \beta), \quad (1)$$

where $\Gamma(\cdot)$ denotes the Gamma function, γ represents the uncertainty space to be estimated, $v > 0$, $\alpha > 1$ and $\beta > 0$. We aim to estimate a posterior distribution $q(\mu, \sigma^2|\mathbf{z})$. Following the approach described in work [6], we employ a factorization of the estimated distribution such that $q(\mu, \sigma^2) = q(\mu)q(\sigma^2)$. This factorization allows for a tractable approximation of the posterior distribution. Our approximation takes the form of the Gaussian conjugate prior, specifically the Normal Inverse-Gamma (NIG) distribution:

$$p(\{\mu, \sigma^2\} \mid \Omega) = \frac{\beta^\alpha \sqrt{v}}{\Gamma(\alpha)\sqrt{2\pi\sigma^2}} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left\{-\frac{2\beta + v(\gamma-\mu)^2}{2\sigma^2}\right\}, \quad (2)$$

where $\Omega = \{\gamma, v, \alpha, \beta\}$ denotes the set of uncertainty parameters we aim to estimate. Given a NIG distribution parameterized by Ω, we can compute the uncertainty space and model uncertainty:

$$\underbrace{\mathbb{E}[\mu] = \gamma}_{\text{uncertainty space}} \qquad \underbrace{\text{Var}[\mu] = \frac{\beta}{v(\alpha-1)}}_{\text{model uncertainty}}. \quad (3)$$

This mathematical formulation delineates the theoretical framework underpinning our approach to uncertainty quantification. The evidential learning paradigm we have introduced essentially constitutes an uncertainty estimation methodology based on the likelihood function. This approach involves training a neural network to output the hyperparameters for fitting an evidential distribution.

Next, we outline our method for obtaining evidential parameters. Our training process is designed to optimize a dual-objective function that simultaneously addresses two critical aspects: (i) increasing model evidence to support the training samples, which in this context represent normal network traffic patterns, and (ii) reducing evidence when uncertainty space exhibits inconsistencies or inaccuracies. Objective (i) can be conceptualized as a mechanism for adapting our data to the evidential model, while objective (ii) serves to enforce a prior that mitigates inaccurate evidence and amplifies uncertainty where appropriate.

Objective (i): Maximizing the Normal Evidence. In accordance with Bayesian probability theory, the "model evidence" is defined as the likelihood of an observation, given the evidential distribution parameters Ω. This is computed by marginalizing over the likelihood parameters (μ, σ^2):

$$p(\mathbf{z} \mid \Omega) = \int_{\sigma^2=0}^{\infty} \int_{\mu=-\infty}^{\infty} p\left(\mathbf{z} \mid \mu, \sigma^2\right) p\left(\mu, \sigma^2 \mid \Omega\right) \mathrm{d}\mu \mathrm{d}\sigma^2. \quad (4)$$

The direct fitting of the evidential model parameters Ω to this likelihood distribution presents significant computational challenges. However, by applying a Normal Inverse-Gamma (NIG) evidential prior to the Gaussian likelihood function, we can derive an analytical solution, as demonstrated in work [6]:

$$p(\mathbf{z} \mid \Omega) = \text{St}\left(\mathbf{z}; \gamma, \frac{\beta(1+v)}{v\alpha}, 2\alpha\right), \quad (5)$$

where $\text{St}(\cdot; \mu_{\text{St}}, \sigma^2_{\text{St}}, v_{\text{St}})$ denotes the Student's t-distribution evaluated at location parameter μ_{St}, scale parameter σ^2_{St}, and degrees of freedom v_{St}. To optimize the model's representation of normal network traffic, we maximize the logarithm of the model evidence, which is equivalent to minimizing its negative. This objective guides the uncertainty parameter heads to output the parameters of a NIG distribution that best fits the distribution of normal network traffic. Formally, we define the training objective $\mathcal{L}^{\text{NLL}}$ for maximizing the normal evidence as:

$$\begin{aligned}\mathcal{L}^{\text{NLL}} = &\frac{1}{2}\log\left(\frac{\pi}{v}\right) - \alpha\log(\omega) + \log\left(\frac{\Gamma(\alpha)}{\Gamma\left(\alpha+\frac{1}{2}\right)}\right) \\ &+ \left(\alpha+\frac{1}{2}\right)\log\left((\mathbf{z}-\gamma)^2 v + \omega\right),\end{aligned} \quad (6)$$

where $\omega = 2\beta(1+v)$.

Objective (ii): Minimizing Evidence on Errors. In addition to maximizing the evidence for normal patterns, we incorporate a regularization term that imposes a high uncertainty prior to penalize incorrect evidence in the uncertainty space. The fundamental principle underlying this regularization is that it should attenuate the weight of evidence where the uncertainty space deviates significantly from the true evidence while having minimal impact on evidence predictions that closely align with the instructive evidence $\mathbf{z}$. To achieve this, we formulate an evidence regularizer [6] $\mathcal{L}^{\text{R}}$ as:

$$\mathcal{L}^{\text{R}} = |\mathbf{z} - \gamma| \cdot (2v + \alpha). \quad (7)$$

3.3.3 Training. Our training loss function comprises three principal components: $\mathcal{L}^{\text{NLL}}$, $\mathcal{L}^{\text{R}}$, and $\mathcal{L}^{\text{Rec}}$:

$$\mathcal{L} = \mathcal{L}^{\text{Rec}} \cdot \lambda_{\text{Rec}} + \mathcal{L}^{\text{NLL}} \cdot \lambda_{\text{NLL}} + \mathcal{L}^{\text{R}} \cdot \lambda_{\text{R}}, \quad (8)$$

where $\lambda_\cdot$ is the hyperparameter to control the contribution of each component. $\mathcal{L}^{\text{Rec}}$ is the reconstruction loss for the autoencoder:.

$$\mathcal{L}^{\text{Rec}} = ||\mathbf{x} - \hat{\mathbf{x}}||_1, \quad (9)$$

where $||\cdot||_1$ denotes the L1 norm. The inclusion of this term ensures the preservation of the autoencoder's fundamental reconstruction capability, enabling the generation of meaningful latent representations. These representations serve as effective evidence instructors for the subsequent uncertainty quantification.

Table 1: Performance comparisons (%) for undrifted anomaly detection on the DataCon2020, CIC-IDS2017, and USTC-TFC2016 datasets. The best results are in bold, and the runner-up results are underlined.

Model	DataCon2020			CIC-IDS2017			USTC-TFC2016		
	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1
PaDim	61.01±0.5	55.07±0.2	67.23±0.2	57.74±0.4	55.06±0.1	68.16±0.1	98.79±0.0	96.93±0.1	96.85±0.1
DFM	83.03±0.3	78.67±0.3	78.85±0.3	69.91±0.2	63.89±0.3	66.87±0.1	94.94±0.3	93.03±0.3	92.63±0.3
DFKDE	72.85±0.5	64.37±0.3	71.86±0.2	67.67±0.6	63.62±0.3	68.61±0.2	91.63±2.4	93.38±0.2	93.79±0.2
Fastflow	69.98±0.5	63.93±0.3	71.65±0.3	78.25±0.3	74.82±0.4	76.39±0.4	99.14±0.0	95.60±0.2	95.42±0.2
Cflow	68.69±1.6	64.14±1.0	72.40±0.8	66.42±0.9	69.46±0.6	69.41±0.5	97.22±0.2	96.76±0.2	96.68±0.2
STFPM	82.37±0.6	80.44±2.1	80.93±1.7	85.89±1.7	80.00±2.1	81.50±1.1	91.63±2.4	89.02±1.3	89.71±1.3
ReverDis	74.53±2.4	68.80±3.1	75.30±1.1	82.22±0.3	77.87±0.4	76.62±0.4	98.05±0.5	95.21±0.9	95.07±0.9
MMR	80.60±2.1	78.85±2.1	79.64±1.3	85.87±1.2	74.36±1.1	74.40±1.0	99.44±0.0	96.15±0.2	96.04±0.2
GANomaly	81.50±1.0	79.40±2.1	79.95±1.6	82.75±4.7	80.85±1.7	81.21±0.9	95.36±1.0	91.27±2.9	91.07±3.2
ARCADE	81.98±4.1	81.48±2.0	80.31±3.4	84.85±2.6	80.15±1.6	82.78±1.0	88.62±2.2	93.13±0.1	93.57±0.1
MFAD	83.16±1.9	76.28±2.4	78.59±1.1	86.02±0.8	81.66±1.9	83.67±1.7	99.73±0.0	97.45±0.4	97.43±0.4
Trident	63.89±0.5	73.67±0.3	78.37±0.3	82.99±0.1	77.42±0.2	75.17±0.2	96.19±0.2	89.86±0.3	89.47±0.3
UnDiff	**86.93±0.3**	**83.16±0.2**	**82.78±0.2**	**88.88±0.4**	**83.31±0.4**	**83.72±0.4**	**99.90±0.0**	**99.47±0.2**	**99.47±0.2**

3.3.4 Inference. The anomaly detection process fundamentally relies on an anomaly score to quantify the degree of deviation from normality. Given that our model is trained exclusively on normal network traffic, the proposed UnDiff naturally assigns low uncertainty to patterns consistent with normal network behavior. Our approach is motivated by the well-established principle that there exists a distributional divergence between normal and anomalous network traffic, encompassing both undrifted and drifted anomalies [8, 21, 54]. Leveraging this insight, we adopt an inter-sample differences method, utilizing model uncertainty as a direct proxy for anomaly scoring. This approach is underpinned by the widely accepted notion in uncertainty learning that deviant samples inherently induce higher model uncertainty [22]. As depicted in Figure 2(b), our method yields an effective and computationally efficient uncertainty-inspired anomaly score. This score is characterized by its ability to generate high uncertainty values for anomalous samples (i.e., out-of-distribution instances relative to the training set) while maintaining low uncertainty for normal samples (i.e., in-distribution instances relative to the training set). In contrast to traditional reconstruction-based anomaly quantification methods, which we categorize as intra-sample difference approaches, UnDiff capitalizes on the intrinsic distributional divergence between normal and anomalous network traffic. Formally, we define our anomaly score as follows:

$$\text{Anomaly Score} = \text{Var}[\mu] = \frac{\beta}{v(\alpha - 1)}. \tag{10}$$

4 Experiments

4.1 Experimental Setting

Dataset. We use three publicly available network traffic anomaly detection datasets for evaluation: (i) *DataCon2020* [10] is an encrypted network traffic dataset comprising normal and malicious traffic, with the latter consisting of encrypted malware communications; (ii) *CIC-IDS2017* [38] is a network intrusion detection dataset that includes seven common attacks; (iii) *USTC-TFC2016* [42] is malware traffic detection dataset with malicious traffic from public sources and normal traffic from eight application types. For consistent evaluation, we randomly sample 10,000 normal network flows for training and 5,000 normal plus 5,000 anomalous flows for testing across all datasets.

Baselines. We evaluate UnDiff with 12 state-of-the-art baselines, categorized into two groups as follows: (i) *Network Traffic Anomaly Detection*: *GANomaly* [3], *ARCADE* [34], *MFAD* [54], and *Trident* [53]; (ii) *Other Advanced Anomaly Detection*: *PaDim* [11], *DFM* [2], *DFKDE* [4], *FastFlow* [47], *CFlow* [19], *STFPM* [40], *ReverDis* [12], and *MMR* [51].

Evaluation Metrics. In alignment with recent models in network traffic anomaly detection [34, 54], we employ three commonly used metrics: AUC, Accuracy (ACC), and F1-Score (F1). Practical detection accuracy is defined as the performance achieved under the optimal F1-Score value.

Drifted Anomaly. We assess model's robustness to anomaly drift by conducting cross-dataset evaluations. Specifically, we train one model on one dataset and evaluate this model's performance on anomalous samples from other datasets. This approach allows us to investigate model's generalization capability and resilience to potential concept shifts in network traffic, thereby assessing model's efficacy in detecting drifted anomalies in real-world environments.

Implementation Details. All experiments are conducted on an NVIDIA GeForce RTX 3090 GPU. We use the Adam optimizer with learning rates of $1e^{-4}$, $1e^{-3}$, $1e^{-6}$ for DataCon2020, CIC-IDS2017, and USTC-TFC2016, respectively. Loss coefficient (λ_{Rec}, λ_{NLL}, λ_{R}) are set as (1, $1e^{-2}$, $1e^{-4}$), (1, $5e^{-2}$, $5e^{-5}$) and (1, 1, $1e^{-2}$), while the low-pass filter uses a cutoff radius of 5. Training proceeds with a batch size of 128 for a maximum of 50 epochs, with early stopping implemented to mitigate overfitting. To ensure statistical robustness, we perform five independent runs with different random seeds, reporting mean results with standard deviations. To facilitate reproducibility, the model code for our UnDiff is available at https://github.com/ikun0124/UnDiff and will be made public.

4.2 Anomaly Detection on Benchmark

To assess UnDiff's efficacy in typical anomaly traffic detection scenarios (i.e., undrifted anomalies), we conducted a comprehensive comparison of our model against 12 competitive baselines on three datasets. The results, as presented in Table 1, demonstrate

Table 2: Performance comparisons (%) for drifted anomaly detection on the DataCon2020, CIC-IDS2017, and USTC-TFC2016 datasets. The abbreviations are explained as follows: D: DataCon2020, I: CIC-IDS2017, and U: USTC-TFC2016.

Model	D->I			D->U			I->D			I->U			U->D			U->I		
	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1
GANomaly	50.73	50.99	66.86	60.96	75.82	79.68	56.58	72.35	75.76	74.92	72.32	77.49	90.30	90.70	91.25	96.01	88.67	87.95
ARCADE	49.99	49.99	66.66	57.87	74.83	78.74	67.47	79.24	80.94	49.60	59.74	69.60	97.87	93.44	93.84	88.61	90.67	91.44
MFAD	62.28	51.09	67.01	79.50	72.76	77.90	77.90	70.56	74.60	82.18	75.84	78.02	98.20	95.27	95.43	98.62	93.92	94.19
Trident	55.11	49.90	66.66	66.08	78.99	81.77	51.20	57.90	70.07	67.47	66.24	72.57	97.47	98.28	98.31	98.83	97.57	97.62
UnDiff-AE	64.48	70.57	77.10	71.08	76.97	80.73	81.48	79.21	80.47	61.88	62.27	71.91	99.57	98.92	98.92	98.75	92.80	93.27
UnDiff	**84.10**	**76.14**	**80.46**	**96.08**	**87.95**	**88.66**	**93.70**	**88.65**	**88.57**	**91.18**	**86.43**	**86.69**	**99.83**	**99.59**	**99.59**	**99.76**	**98.07**	**98.07**

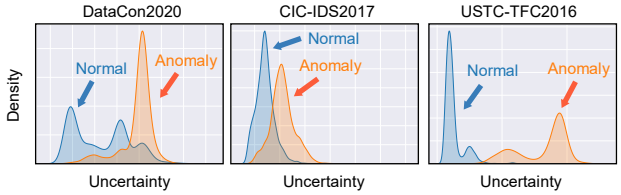


Figure 3: Statistics of the uncertainty-based anomaly scores for UnDiff under the undrifted anomaly scenario.

that our model consistently outperforms baselines across all three datasets. Notably, on the DataCon2020 and CIC-IDS2017 datasets, UnDiff exhibits significant performance improvements over the best-performing baseline MFAD, with enhancements of 3.7% and 2.8% in AUC, respectively. The underlying strength of our method lies in its innovative utilization of uncertainty measures to directly quantify inter-sample differences, thereby facilitating more accurate discrimination of anomalous network traffic patterns.

While statistics-based methods such as PaDim, DFM, and DFKDE, as well as normalizing flow-based approaches like FastFlow and CFlow, attempt to compute distribution deviations by exploiting inter-sample differences, their comparative spaces lack the discriminative power of our uncertainty space. Our approach, built upon the informative reconstruction of latent variables and guided by evidential learning, constructs a more robust and discerning comparative framework. Moreover, distillation-based methods in anomaly detection, including STFPM, ReverDis, and MMR, are constrained by the limitation in effective feature extraction. In contrast, approaches specific to anomaly network traffic detection, while not requiring additional feature extractors, quantify anomalies through intra-sample reconstruction differences. However, these methods, including GANomaly, ARCADE, MFAD, and Trident, suffer from the "identical shortcut" issue, which may significantly compromise the intra-sample differences of anomalies, leading to suboptimal performance. Our uncertainty-inspired framework addresses these limitations by effectively leveraging distributional differences between normal and anomalous samples. By quantifying anomalies from an inter-sample differences perspective, UnDiff provides a more nuanced and robust approach to anomaly detection.

To further corroborate the feasibility of our UnDiff framework, we present a detailed analysis of the anomaly score distributions in Figure 3. The graphical representation reveals a marked bimodal distribution, with a clear separation between the scores associated with normal and anomalous samples. This pronounced divergence in score distributions provides compelling evidence for the discriminative power of our uncertainty-inspired anomaly metric. The clear detachment between normal and anomalous samples also

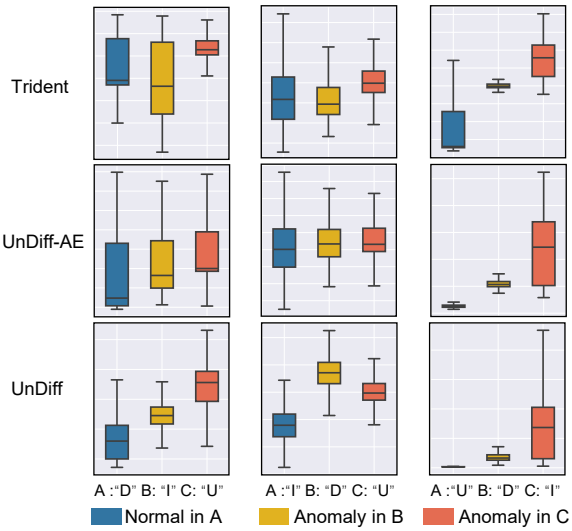


Figure 4: Anomaly score distribution for Trident, UnDiff-AE, and UnDiff under the anomaly drift scenario.

underscores the method's ability to generate highly informative indicators, facilitating more accurate and reliable anomaly detection.

4.3 Drifted Anomaly Detection

To assess the efficacy in addressing drifted anomalies, we compare our UnDiff with state-of-the-art network traffic anomaly detection baselines and a variant of our method – UnDiff-AE, which employs a pure auto-encoder architecture without uncertainty learning. As evidenced in Table 2, these approaches have suboptimal performance, particularly in the drifted experiments from DataCon2020 (D) to CIC-IDS2017 (I), D to USTC-TFC2016 (U), I to D, and I to U. These empirical observations highlight the critical necessity for robust drifted anomaly detection methodologies. The primary limitation of these baselines stems from their reliance on an intra-sample difference paradigm, which is inherently susceptible to the "identical shortcut" issue prevalent in reconstruction-based models. Therefore, the divergence in anomaly scores between normal and anomalous samples is suppressed and obfuscated. We visualize the detailed anomaly scores for Trident and UnDiff-AE in Figure 4 to elucidate this phenomenon. The anomaly score distribution for Trident exhibits significant overlap between normal and anomalous samples, with anomalous samples occasionally scoring lower than normal samples. This observation indicates that the "identical shortcut" issue profoundly compromises the efficacy of intra-sample differences in detecting drifted anomalies. In contrast,

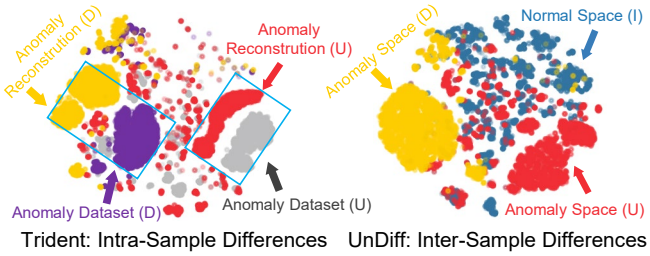


Figure 5: The t-SNE visualization comparison between intra-sample and inter-sample differences.

UnDiff achieves apparent distinction in the distribution between normal and drifted anomaly samples, thereby validating the effectiveness of our inter-sample differences approach. Notably, the comparative analysis with UnDiff-AE shows that the substantial improvement in UnDiff's performance is predominantly attributable to the uncertainty-inspired inter-sample differences rather than the fundamental auto-encoder architecture.

4.4 Qualitative Study

We now elucidate the underlying mechanisms contributing to UnDiff's enhanced performance by two *t*-SNE visualizations. As illustrated in Figure 5, we observe a degree of confusion between the pre- and post-reconstruction embeddings of anomalous samples on Trident, manifested as certain overlaps and approximate profiles with minimal distance. We posit that this phenomenon arises from the "identical shortcut" issue, an inherent limitation in reconstruction-based approaches. This limitation leads to well-reconstructed representations even for anomalous samples, a phenomenon that contradicts the fundamental detection motivation of reconstruction methods. Consequently, this results in indistinguishable intra-sample differences between normal and anomalous traffic patterns, compromising the efficacy of traditional approaches. In contrast, UnDiff is based on a novel inter-sample differences perspective, effectively leveraging the axiom that anomalous samples inherently deviate from normal samples in the feature space. The representation within our uncertainty space demonstrates the feasibility and effectiveness of uncertainty-inspired modeling and detection. This approach makes the discrimination by exploiting inter-sample differences, thereby overcoming the limitations inherent in intra-sample comparison methods.

4.5 Ablation Study

To evaluate the contributions of each component in our UnDiff, we conduct an ablation study comprising four variants. These variants are constructed by removing one of the key components in UnDiff: the reconstruction loss $\mathcal{L}^{\text{Rec}}$ (w/o $\mathcal{L}^{\text{Rec}}$), the regularization loss $\mathcal{L}^{\text{R}}$ (w/o $\mathcal{L}^{\text{R}}$), the uncertainty-based anomaly score (w/o AS), and both the uncertainty-based modeling and anomaly score (w/o T&AS). As illustrated in Table 3, the removal of $\mathcal{L}^{\text{Rec}}$ results in substantial performance degradation, underscoring the critical role of reconstruction loss in ensuring a refined representation of normal network traffic. Furthermore, we observe a notable decline in performance upon removal of $\mathcal{L}^{\text{R}}$, indicating its efficacy as a regularization constraint in preventing the formation of erroneous evidence spaces during the uncertainty quantification process. While removing the uncertainty-based anomaly score (w/o AS) and both uncertainty-based modeling and anomaly score (w/o T&AS) resulted in performance degradation, our complete UnDiff model demonstrates optimal performance in drifted anomaly detection. This suggests that the uncertainty-based modeling and inter-sample difference detection components effectively leverage prior differences between normal and anomalous samples, mitigating the inherent limitations of purely reconstruction-based methods.

Table 3: Ablation studies for drifted and undrifted anomaly detection (AUC). The gray color denotes undrifted detection.

Variant	DataCon2020			CIC-IDS2017			USTC-TFC2016		
	D	I	U	D	I	U	D	I	U
w/o $\mathcal{L}^{\text{Rec}}$	83.44	61.79	80.26	78.46	76.22	59.41	98.95	80.97	98.25
w/o $\mathcal{L}^{\text{R}}$	84.80	71.48	72.50	85.31	86.80	81.12	99.55	99.39	99.78
w/o AS	85.66	82.52	83.25	83.28	87.67	86.15	99.74	98.81	99.84
w/o T&AS	85.55	64.48	71.08	81.48	86.02	61.88	99.51	98.75	99.71
UnDiff	**86.93**	**84.10**	**96.08**	**93.70**	**88.88**	**91.18**	**99.83**	**99.76**	**99.90**

Table 4: Overhead comparison for inference.

	GANomaly	ARCADE	MFAD	Trident	UnDiff
MACs (G)	0.98	0.82	0.99	**0.03**	0.25
#Paras (M)	9.66	6.7	10.07	27.61	**2.55**

4.6 Overhead Evaluation

We conduct an analysis of model efficiency, focusing on multiply-accumulate operations per second (MACs) and the number of model parameters (#Paras) during inference. The results of this analysis are summarized in Table 4. Our UnDiff demonstrates excellent performance with favorable computational overhead compared to alternative baselines. This efficiency can be attributed to strategic design choices, such as the multi-channel image representation, a low-parameter evidence extractor, and a set of concise uncertainty parameter heads. Notably, UnDiff balances performance and computational requirements, rendering it particularly suitable for practical deployment in network traffic anomaly detection scenarios.

5 Conclusion

This study presents a pioneering approach to network traffic anomaly detection by developing an inter-sample differences method based on uncertainty. This novel methodology directly addresses the challenges of anomaly detection while circumventing the "identical shortcut" issue inherent in existing methods that rely on intra-sample differences between pre- and post-reconstruction representations. Our proposed UnDiff effectively leverages the prior knowledge that anomalous samples inherently deviate from normal samples. This enables learning a more discriminative uncertainty space, facilitating optimal detection performance. Comprehensive empirical evaluations across three benchmark datasets demonstrate UnDiff's superior performance in detecting undrifted and drifted anomalies with minimal additional computational overhead.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No.62176043, No.62072077 and No.U22A2097).

References

[1] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul W. Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion* 76 (2021), 243–297.

[2] Nilesh A. Ahuja, Ibrahima J. Ndiour, Trushant Kalyanpur, and Omesh Tickoo. 2019. Probabilistic Modeling of Deep Features for Out-of-Distribution and Adversarial Detection. *arXiv:1909.11786* (2019).

[3] Samet Akcay, Amir Atapour Abarghouei, and Toby P. Breckon. 2018. GANomaly: Semi-supervised Anomaly Detection via Adversarial Training. In *Asian Conference on Computer Vision (ACCV)*. 622–637.

[4] Samet Akcay, Dick Ameln, Ashwin Vaidya, Barath Lakshmanan, Nilesh A. Ahuja, and Ergin Utku Genc. 2022. Anomalib: A Deep Learning Library for Anomaly Detection. In *IEEE International Conference on Image Processing (ICIP)*. 1706–1710.

[5] Mejbah Alam, Justin Gottschlich, Nesime Tatbul, Javier S Turek, Tim Mattson, and Abdullah Muzahid. 2019. A Zero-Positive Learning Approach for Diagnosing Software Performance Regressions. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*.

[6] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. 2020. Deep Evidential Regression. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*. 14927–14937.

[7] Federico Barbero, Feargus Pendlebury, Fabio Pierazzi, and Lorenzo Cavallaro. 2022. Transcending TRANSCEND: Revisiting Malware Classification in the Presence of Concept Drift. In *Symposium on Security and Privacy (S&P)*. IEEE, 805–823.

[8] Saihua Cai, Han Tang, Jinfu Chen, Yikai Hu, and Wuhao Guo. 2025. CDDA-MD: An efficient malicious traffic detection method based on concept drift detection and adaptation technique. *Computers & Security* 148 (2025), 104121.

[9] Chengtai Cao, Xinhong Chen, Jianping Wang, Qun Song, Rui Tan, and Yung-Hui Li. 2024. CCTR: Calibrating Trajectory Prediction for Uncertainty-Aware Motion Planning in Autonomous Driving. In *AAAI Conference on Artificial Intelligence (AAAI)*. 20949–20957.

[10] Community DataCon. 2020. DataCon: open dataset DataCon2020-encrypted malicious traffic dataset direction open dataset. https://datacon.qianxin.com/opendata.

[11] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. 2020. PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization. In *International Conference on Pattern Recognition (ICPR)*. 475–489.

[12] Hanqiu Deng and Xingyu Li. 2022. Anomaly Detection via Reverse Distillation from One-Class Embedding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 9727–9736.

[13] Xianwen Deng, Yijun Wang, and Zhi Xue. 2024. AN-Net: an Anti-Noise Network for Anonymous Traffic Classification. In *The ACM Web Conference (WWW)*. 4417–4428.

[14] Yasir Ali Farrukh, Syed Wali, Irfan Khan, and Nathaniel D. Bastian. 2023. SeNet-I: An approach for detecting network intrusions through serialized network traffic images. *Engineering Applications of Artificial Intelligence* 126, Part D (2023), 107169.

[15] Chuanpu Fu, Qi Li, and Ke Xu. 2023. Detecting Unknown Encrypted Malicious Traffic in Real Time via Flow Interaction Graph Analysis. In *Network and Distributed System Security Symposium (NDSS)*. The Internet Society.

[16] Chuanpu Fu, Qi Li, Ke Xu, and Jianping Wu. 2023. Point Cloud Analysis for ML-Based Malicious Traffic Detection: Reducing Majorities of False Positive Alarms. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 1005–1019.

[17] Yarin Gal and Zoubin Ghahramani. 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *International Conference on Machine Learning (ICML)*. 1050–1059.

[18] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna M. Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiaoxiang Zhu. 2023. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review* 56, S1 (2023), 1513–1589.

[19] Denis A. Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. 2022. CFLOW-AD: Real-Time Unsupervised Anomaly Detection with Localization via Conditional Normalizing Flows. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 1819–1828.

[20] Maryam Habibpour, Hassan Gharoun, Mohammadreza Mehdipour, AmirReza Tajally, Hamzeh Asgharnezhad, Afshar Shamsi Jokandan, Abbas Khosravi, and Saeid Nahavandi. 2023. Uncertainty-aware credit card fraud detection using deep learning. *Engineering Applications of Artificial Intelligence* 123, Part A (2023), 106248.

[21] Dongqi Han, Zhiliang Wang, Wenqi Chen, Kai Wang, Rui Yu, Su Wang, Han Zhang, Zhihua Wang, Minghui Jin, Jiahai Yang, Xingang Shi, and Xia Yin. 2023. Anomaly Detection in the Open World: Normality Shift Detection, Explanation, and Adaptation. In *Network and Distributed System Security Symposium (NDSS)*. The Internet Society.

[22] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. 2023. Trusted Multi-View Classification With Dynamic Evidential Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligenc (TPAMI)* 45, 2 (2023), 2551–2566.

[23] Chao Huang, Chengliang Liu, Zheng Zhang, Zhihao Wu, Jie Wen, Qiuping Jiang, and Yong Xu. 2022. Pixel-Level Anomaly Detection via Uncertainty-aware Prototypical Transformer. In *ACM International Conference on Multimedia (MM)*. 521–530.

[24] Chao Huang, Yushu Shi, Bob Zhang, and Ke Lyu. 2024. Uncertainty-aware prototypical learning for anomaly detection in medical images. *Neural Networks* 175 (2024), 106284.

[25] Shuodi Hui, Huandong Wang, Zhenhua Wang, Xinghao Yang, Zhongjin Liu, Depeng Jin, and Yong Li. 2022. Knowledge Enhanced GAN for IoT Traffic Generation. In *The ACM Web Conference (WWW)*. 3336–3346.

[26] Alex Kendall and Yarin Gal. 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?. In *Annual Conference on Neural Information Processing Systems (NIPS)*. 5574–5584.

[27] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. 2017. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Annual Conference on Neural Information Processing Systems (NIPS)*. 6402–6413.

[28] Tae Jun Lee, Justin Gottschlich, Nesime Tatbul, Eric Metcalf, and Stan Zdonik. 2018. Greenhouse: A Zero-Positive Machine Learning System for Time-Series Anomaly Detection. *arXiv:1801.03168* (2018).

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

[30] Xinjie Lin, Gang Xiong, Gaopeng Gou, Zhen Li, Junzheng Shi, and Jing Yu. 2022. ET-BERT: A Contextualized Datagram Representation with Pre-training Transformers for Encrypted Traffic Classification. In *The ACM Web Conference (WWW)*. 633–642.

[31] Jiaqi Liu, Guoyang Xie, Jin-Bao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. 2024. Deep Industrial Image Anomaly Detection: A Survey. *Machine Intelligence Research* 21, 1 (2024), 104–135.

[32] Ximeng Liu, Shizhen Zhao, Yong Cui, and Xinbing Wang. 2024. FIGRET: Fine-Grained Robustness-Enhanced Traffic Engineering. In *ACM SIGCOMM Conference (SIGCOMM)*. 117–135.

[33] Ruiying Lu, Yujie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. 2023. Hierarchical Vector Quantized Transformer for Multi-class Unsupervised Anomaly Detection. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*. 8487–8500.

[34] Willian Tessaro Lunardi, Martin Andreoni Lopez, and Jean Pierre Giacalone. 2023. ARCADE: Adversarially Regularized Convolutional Autoencoder for Network Anomaly Detection. *IEEE Transactions on Network and Service Management (TNSM)* 20, 2 (2023), 1305–1318.

[35] Neelu Madan, Nicolae-Catalin Ristea, Kamal Nasrollahi, Thomas B. Moeslund, and Radu Tudor Ionescu. 2024. CL-MAE: Curriculum-Learned Masked Autoencoders. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2480–2490.

[36] Andrey Malinin and Mark J. F. Gales. 2018. Predictive Uncertainty Estimation via Prior Networks. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*. 7047–7058.

[37] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. 2022. Fully Convolutional Cross-Scale-Flows for Image-based Defect Detection. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 1829–1838.

[38] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani. 2018. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. In *International Conference on Information Systems Security and Privacy (ICISSP)*. 108–116.

[39] Wenxin Tai, Bin Chen, Fan Zhou, Ting Zhong, Goce Trajcevski, Yong Wang, and Kai Chen. 2023. TrustGeo: Uncertainty-Aware Dynamic Graph Learning for Trustworthy IP Geolocation. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. 4862–4871.

[40] Guodong Wang, Shumin Han, Errui Ding, and Di Huang. 2021. Student-Teacher Feature Pyramid Matching for Anomaly Detection. In *British Machine Vision Conference (BMVC)*. 306.

[41] Meng Wang, Tian Lin, Lianyu Wang, Aidi Lin, Ke Zou, Xinxing Xu, Yi Zhou, Yuanyuan Peng, Qingquan Meng, Yiming Qian, et al. 2023. Uncertainty-inspired open set learning for retinal anomaly identification. *Nature Communications* 14, 1 (2023), 6757.

[42] Wei Wang, Ming Zhu, Xuewen Zeng, Xiaozhou Ye, and Yiqiang Sheng. 2017. Malware traffic classification using convolutional neural network for representation learning. In *International Conference on Information Networking (ICOIN)*. 712–717.

[43] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. CBAM: Convolutional Block Attention Module. In *European conference on computer vision (ECCV)*. 3–19.

[44] Limin Yang, Wenbo Guo, Qingying Hao, Arridhana Ciptadi, Ali Ahmadzadeh, Xinyu Xing, and Gang Wang. 2021. CADE: Detecting and Explaining Concept

Drift Samples for Security Applications. In *USENIX Security Symposium (USENIX Security)*. USENIX Association, 2327–2344.

[45] Xincheng Yao, Chongyang Zhang, Ruoqi Li, Jun Sun, and Zhenyu Liu. 2023. One-for-All: Proposal Masked Cross-Class Anomaly Detection. In *AAAI Conference on Artificial Intelligence (AAAI)*. 4792–4800.

[46] Bang Xiang Yong and Alexandra Brintrup. 2022. Bayesian autoencoders with uncertainty quantification: Towards trustworthy anomaly detection. *Expert Systems with Applications* 209 (2022), 118196.

[47] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. 2021. FastFlow: Unsupervised Anomaly Detection and Localization via 2D Normalizing Flows. *arXiv:2111.07677* (2021).

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

[49] Menghao Zhang, Jingyu Wang, Qi Qi, Haifeng Sun, Zirui Zhuang, Pengfei Ren, Ruilong Ma, and Jianxin Liao. 2024. Multi-Scale Video Anomaly Detection by Multi-Grained Spatio-Temporal Representation Learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 17385–17394.

[50] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. 2023. DeSTSeg: Segmentation Guided Denoising Student-Teacher for Anomaly Detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 3914–3923.

[51] Zilong Zhang, Zhibin Zhao, Xingwu Zhang, Chuang Sun, and Xuefeng Chen. 2023. Industrial anomaly detection with domain shift: A real-world dataset and masked multi-scale reconstruction. *Computers in Industry* 151 (2023), 103990.

[52] Ruijie Zhao, Mingwei Zhan, Xianwen Deng, Yanhao Wang, Yijun Wang, Guan Gui, and Zhi Xue. 2023. Yet Another Traffic Classifier: A Masked Autoencoder Based Traffic Transformer with Multi-Level Flow Representation. In *AAAI Conference on Artificial Intelligence (AAAI)*. 5420–5427.

[53] Ziming Zhao, Zhaoxuan Li, Zhuoxue Song, Wenhao Li, and Fan Zhang. 2024. Trident: A Universal Framework for Fine-Grained and Class-Incremental Unknown Traffic Detection. In *The ACM Web Conference (WWW)*. 1608–1619.

[54] Yu Zheng, Xinglin Lian, Zhangxuan Dang, Chunlei Peng, Chao Yang, and Jianfeng Ma. 2023. A Semi-Supervised Anomaly Network Traffic Detection Framework via Multimodal Traffic Information Fusion. In *ACM International Conference on Information and Knowledge Management (CIKM)*. 4455–4459.