A Framework for AI-Based Large-Scale Time-Series Anomaly Detection in 5G Core Networks

Yeon-Jea Cho
Future Network Research Laboratory
KT Corporation
Seoul, Republic of Korea
yeonjea.cho@kt.com

Hong-Jae Lee
Future Network Research Laboratory
KT Corporation
Seoul, Republic of Korea
hong_jae.lee@kt.com

A-Sol Choi
Future Network Research Laboratory
KT Corporation
Seoul, Republic of Korea
a-sol.choi@kt,com

Do-Young Kwak
Future Network Research Laboratory
KT Corporation
Seoul, Republic of Korea
dy.kwak@kt.com

Jemin Chung
Future Network Research Laboratory
KT Corporation
Seoul, Republic of Korea
jemin.chung@kt.com

***Abstract*— This paper presents a framework for large-scale anomaly detection on time-series key performance indicators (KPIs) in 5G core networks. While state-of-the-art time-series anomaly detection models can achieve high detection accuracy, their early detection capability is often limited because they need to accumulate sufficient time-series evidence within an inference window to identify diverse anomaly patterns. Furthermore, it is challenging to determine a customized anomaly score threshold for each of the tens of thousands of diverse KPI streams since each KPI exhibits a unique score distribution. To address these challenges, we propose a framework that enhances early detection capability through a lightweight point-level anomaly amplification technique and automatically determines thresholds for each KPI's anomaly scores based on their distributional characteristics. These two contributions enable timely and reliable detection across large-scale KPI streams. We also carried out a proof-of-concept (PoC) verification of the proposed approach in our real-world 5G core network, demonstrating its applicability to large-scale KPI anomaly detection with regular summary reports.**

Keywords—Large-scale time-series anomaly detection, 5G core network, point-level anomaly amplification, distribution-aware auto-thresholding

I. INTRODUCTION

As cellular infrastructure has evolved from 4G (LTE) to cloud-native 5G cores, the operational requirements and conditions have changed dramatically, and early discussions on 6G are already underway [1], [2]. In current commercial 5G deployments, network functions (NFs) in the core network generate tens of thousands of time-series key performance indicators (KPIs) at several-minute intervals, and these will need to be monitored in near real time. This capability will remain essential for the evolution toward AI-native networks. In this regard, manual inspection of the time-series trends and patterns in such large-scale KPI datasets is infeasible for human operators, making timely anomaly detection infeasible as well.

Recent studies have explored machine-learning and deep-learning approaches in the field of time-series anomaly detection. Although models such as LSTM autoencoders, GAN-based detectors, and the Anomaly Transformer have been widely studied for time-series anomaly detection [3]-[5], they are generally studied under small-scale settings and do not consider the requirement of anomaly detection across thousands to tens of thousands of KPI streams.

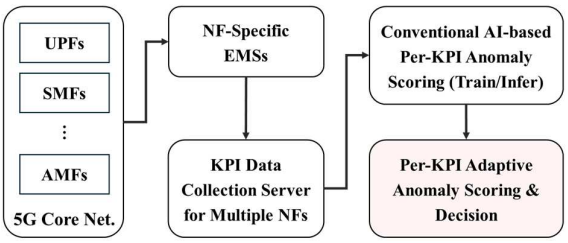

Fig. 1. Sysyem architecture for large-scale KPI anomaly detection in 5G core networks.

However, to be applicable in large-scale, per-KPI monitoring scenarios, these methods face several limitations that become more significant in real-world 5G core networks, as discussed in Section II. Despite progress in this research field [6]-[10], no prior work has directly addressed the challenges of real-time, large-scale KPI anomaly detection in 5G core networks. To address this gap, we propose a framework that enables timely and automated tuning of anomaly score thresholds across tens of thousands of various types of KPI streams.

The remainder of this paper is organized as follows. Section II describes the system architecture and defines the problem addressed in this work. Section III presents proposed large-scale Time-Series Anomaly Detection (TSAD) framework, including the Point-Level Anomaly Amplification (PAA) technique for enhancing detection responsiveness and the Distribution-Aware Auto-Thresholding (DAAT) method for adaptive threshold calibration across diverse KPI streams. Section IV provides proof-of-concept (PoC) results from our real-world 5G core network. Section V concludes the paper.

II. SYSTEM ARCHITECTURE AND PROBLEM STATEMENT

Figure 1 illustrates the system architecture for large-scale KPI anomaly detection in 5G core networks, including the end-to-end data flow from the periodic collection of KPI data to AI-based anomaly analysis. The figure includes a representative configuration of 5G core network functions, where key components such as AMF, SMF, UPF, and PCF are virtualized and deployed as individual network functions (NFs). Each NF is managed by its corresponding element management system (EMS), which periodically collects and manages KPI data associated with that NF. These KPIs are

stored in file format at each EMS and then aggregated by a centralized server, where they are used for further analysis and serve as input to the AI-based anomaly detection model. The total number of KPIs can easily reach tens or even hundreds of thousands. Each KPI is associated with a specific network entity and measurement purpose such as an NF instance, message type, or a protocol interface between NF pairs resulting in a vast set of uniquely defined metrics.

In this regard, the '*Conventional AI-based Per-KPI Anomaly Scoring (Train/Infer)*' block, as shown in Fig. 1, is responsible for generating baseline anomaly scores within our framework. It periodically collects time-series data for each KPI and trains an independent deep learning model per KPI stream. During operation, these trained models infer initial anomaly scores in real time. In our implementation, we utilize the Anomaly Transformer [5], a state-of-the-art unsupervised time-series anomaly detection model. During training, the model learns temporal dependencies through self-attention while optimizing both reconstruction error and association discrepancy losses. At inference, anomaly scores are derived by jointly evaluating reconstruction error and association discrepancy. However, even such state-of-the-art models face two key limitations in real-world 5G core networks:

1) Detection-latency bottleneck: Window-based inference pipelines often detect an incident only after a sufficient number of anomalous samples have accumulated within the sliding window, which in turn delays the initial alert when operators require the earliest warning.

2) Threshold-engineering bottleneck: Each KPI typically follows its own anomaly score distribution, which is often non-Gaussian. Consequently, manual tuning of thresholds across tens of thousands of KPI streams is infeasible, and naive approaches such as applying a Z-score rule uniformly across all KPIs can lead to excessive false alarms for some KPIs while failing to anomalies in others.

To overcome these limitations, our framework incorporates the '*Per-KPI Adaptive Anomaly Scoring & Decision*' block, as depicted in Fig. 1. This block applies two core techniques: point-level anomaly amplification to improve responsiveness by enhancing early-stage anomaly signals, and a distribution-aware auto-thresholding mechanism that dynamically calibrates detection thresholds based on each KPI's anomaly score distribution. These components work together to achieve timely and reliable anomaly detection across large-scale KPI streams in 5G core networks.

III. THE PROPOSED LARGE-SCALE TSAD FRAMEWORK

The proposed framework integrates two key techniques: point-level anomaly amplification (PAA) and distribution-aware auto-thresholding (DAAT). These two components jointly enable scalable per-KPI time-series anomaly detection in real-world 5G core networks. An overview of the proposed framework, which incorporates PAA and DAAT to perform large-scale time-series anomaly detection, is presented in Figure 2. The following subsections describe each technique in detail.

A. Point-Level Anomaly Amplication (PAA)

As introduced in Section II, to address the detection latency issue, we incorporate a Point-Level Anomaly Amplification

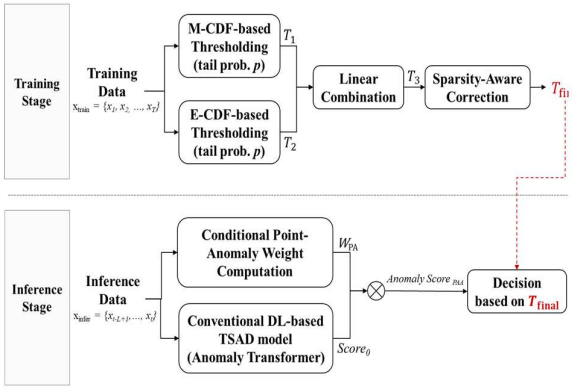

Fig. 2. Overview of the proposed large-scale TSAD framework

(PAA) mechanism, which enhances the sensitivity of anomaly scores to rapid local changes within the inference window. The PAA mechanism is inspired by local Z-score-based approaches, which have been used for anomaly detection in time-series data [10]. However, in this paper, we leverage this concept in a novel way by combining it with deep learning (DL)-based anomaly detectors to efficiently enhance responsiveness to early anomaly signals.

This mechanism is lightweight and requires no modification or retraining of the underlying detection model, making it highly practical for large-scale deployments. As a result, the system can react more promptly to sudden deviations in KPI behavior, enabling faster detection of point anomalies by enhancing each raw anomaly score with a localized point-anomaly weight. This weight is computed during the inference stage in four steps as follows.

1) Adaptive Fractional Differentiation:

For each KPI, given the most recent sequence $\{x_{t-L+1}, \ldots, x_t\}$ of length L, we define $f(x)$ as a function that characterizes the temporal trend within this window. To remove low-frequency trend components while retaining sharp local changes, we compute a fractionally differenced sequence defined as:

$$f^n(x) = \mathrm{c0} \cdot f^0(x) + \mathrm{c1} \cdot f^1(x) + \mathrm{c2} \cdot f^2(x), \qquad (1)$$

where $f^n(x)$ denotes the n-th order discrete derivative of $f(x)$. The coefficients $\{c_0, c_1, c_2\}$ are selected such that each lies between 0 and 1, and are normalized to sum to 1, which enables a flexible approximation of the desired differentiation order n^*. Unlike conventional differentiation where n is restricted to integers, here n^* is generalized to any real value between 0 and 2, allowing fractional-order derivatives that capture intermediate behaviors between smoothing ($n^* \approx 0$) and sharp differencing ($n^* \approx 2$). Specifically, the coefficients are determined according to the following cases:

- If $0 \le n^* \le 1$, we set $c_0 = 1 - n^*$, $c_1 = n^*$, $c_2 = 0$.
- If $1 < n^* \le 2$, we set $c_0 = 0$, $c_1 = 2 - n^*$, $c_2 = n^* - 1$.

These conditions ensure that the weighted average of adjacent integer-order derivatives approximates the target fractional order n^*, while maintaining algebraic simplicity and computational efficiency.

2) Local Standardized Point-Anomaly Score :

For each KPI, the mean μ_t and standard deviation σ_t are computed from the frantionally differened function $f^n(x)$ over the most recent L-1 samples, excluding the current sample x_t. The local standardized point-anomaly score is then defined as

$$z_t = |(\mu_t - x_t) / \sigma_t|. \quad (2)$$

Both μ_t and σ_t are updated at each inference step, ensuring that z_t reflects only the most recent window of observations.

3) Conditional Weighting:

The raw anomaly score from the backbone model is scaled by a weight W_{PA}, which is set to $K \cdot z_t$ when $z_t > z_{th}$, and to 1 otherwise. The parameter K controls the strength of amplication, while z_{th} defines the minimum deviation required to trigger it.

4) Amplified Anomaly Score:

The final anomaly score used in the proposed framework is computed as the product of the raw backbone score and the point-anomaly weight:

$$Anomaly\ Score_{PAA} = W_{PA} \cdot Score_0, \quad (3)$$

where $Score_0$ denotes the original anomaly score produced by the backbone (e.g., Anomaly Transformer), and W_{PA} is the conditional weight defined in Step 3. This weight amplifies the anomaly score when the deviation exceeds a threshold z_{th}, and equals 1 otherwise.

Amplification is applied selectively, which enhances the sensitivity to point anomalies. Although a slight increase in false alarms is theoretically possible, this can be effectively controlled through tunable parameters such as the amplification factor K, the fractional differencing order n^*, and the threshold z_{th}. This yields a more flexible and effective scoring mechanism compared to raw backbone outputs.

B. Distribution-Aware Auto-Thresholding (DAAT)

While PAA enhances anomaly scoring by amplifying early signals of point anomalies, it is also necessary for reliable anomaly detection that anomaly score thresholds are properly calibrated for each KPI. To address this challenge, we introduce the Distribution-Aware Auto-Thresholding (DAAT) module. In existing approaches, such as prior work [10], applying a simple Z-score rule uniformly to the anomaly scores of individual KPIs still fails to account for the fact that each KPI often follows a distinct anomaly score distribution.

DAAT automatically computes reliable decision thresholds by analyzing the anomaly score distribution of each KPI stream. It adapts to diverse score characteristics across different KPIs and ensures threshold stability even for KPI streams with near-constant temporal behavior. This is achieved by combining empirical and parametric views of the score distribution and by applying a sparsity-aware correction to prevent the threshold from becoming unnecessarily large in the case of near-constant KPI streams. The detailed procedure is described in the following steps.

1) Empirical Score Sampling :

During the training phase, time-series KPI data is provided as input. An unsupervised anomaly detection model is trained separately for each KPI to learn its normal behavior patterns. The trained model is then applied to the same training data to generate anomaly score samples across the entire training window. We denote the input time-series as $x = \{x_1, x_2, \ldots, x_T\}$. Applying the trained model yields a corresponding sequence of anomaly scores, $\boldsymbol{s} = \{s_1, s_2, \ldots, s_T\}$. The set $\boldsymbol{s}$ serves as the empirical sample for deriving the score distribution in the following steps.

2) Measured CDF (M-CDF) :

To represent the empirical distribution of the anomaly scores, we construct a cumulative distribution function (CDF) directly from the histogram. This histogram-based CDF, denoted by $F_M(z)$, is

$$F_M(z) = \frac{1}{T}\sum_{t=0}^{T} \mathbf{1}\{s_t \le z\}, \quad (4)$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Based on this empirical CDF, we define the M-CDF-based threshold T_1 as the $(1-p)$-quantile of this distribution:

$$T_1 = F_M^{-1}(1-p), \quad (5)$$

where p is a user-defined tail probability that determines how strictly the threshold filters extreme anomaly scores.

M-CDF directly reflects the observed score distribution, making it effective when anomaly scores exhibit complex characteristics, such as clustering in specific ranges or intermittent sharp fluctuations. By deriving the threshold from the empirical histogram, the method can accurately capture such characteristics. However, it may also be sensitive to temporary fluctuations in the distribution caused by a small number of unusual anomaly score samples, which can result in excessively high thresholds and increase the risk of missed detections.

3) Estimated-CDF (E-CDF) :

To complement the empirical estimation from M-CDF, the same set of anomaly scores is fitted to a family of parametric probability distribution functions (PDFs), including Normal, Log-normal, Exponential, Gamma, Beta. Each candidate distribution is fitted by minimizing the root mean square error (RMSE) with respect to the empirical score histogram. The distribution with the lowest RMSE yields the estimated CDF, denoted $F_E(z)$, along with its optimal parameters.

Based on this estimated CDF, we define the E-CDF-based threshold T_2 as the $(1-p)$-quantile:

$$T_2 = F_E^{-1}(1-p), \quad (6)$$

where p is a user-defined tail probability (e.g., 10^{-4} to 10^{-3}) as in Step 2, controlling the sensitivity of anomaly detection in the E-CDF-based thresholding.

E-CDF determines the threshold by selecting the best-fitted distribution from a set of well-established probability models, enabling it to capture broader trends in the data while being less sensitive to extreme outliers. This approach generally produces a more stable threshold. However, when anomaly scores are heavily concentrated in specific ranges or show intermittent sharp fluctuations, E-CDF may fail to capture such localized characteristics, potentially leading to excessively low thresholds and a higher risk of false alarm.

4) Combined Threshold (T_3) :

The two thresholds T_1 (from M-CDF) and T_2 (from E-CDF) are combined into a single threshold via a linear combination:

$$T_3 = c_1 \cdot T_1 + c_2 \cdot T_2, \quad (7)$$

where $c_1 + c_2 = 1$, and $c_1 > 0$, $c_2 > 0$. The default weights are set to $c_1 = 0.30$ and $c_2 = 0.70$, which has been validated as

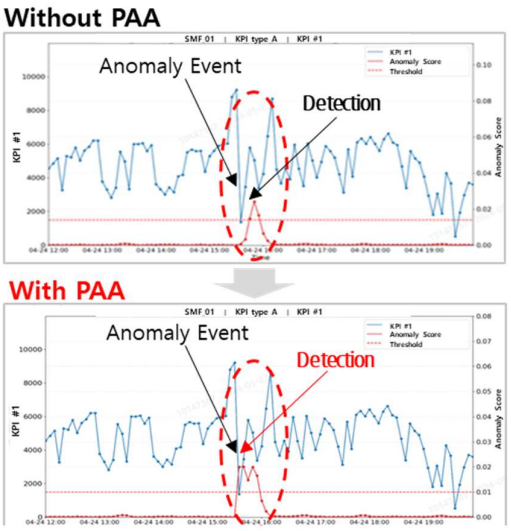

Fig. 3. Comparison of anomaly detection response time Without PAA (top) and With PAA (bottom).

effective in our real-world 5G core network. These coefficients can also be further tuned according to operator feedback or specific deployment requirement.

By combining the M-CDF and E-CDF thresholds, T_3 effectively integrates the detailed distributional characteristics from M-CDF and the stability of E-CDF. This hybrid thresholding approach enables the automatic determination of anomaly score thresholds for each KPI in large-scale time-series KPI analysis, thereby providing a consistent and robust decision criterion that can be directly applied to anomaly detection in 5G core networks.

5) Sparsity-Aware Correction :

To address sparse or low-variance KPI streams where even minor changes can produce abnormally high anomaly scores, we define a sparsity metric ρ as the ratio of the most frequent score value to the total number of samples. The value of ρ approaches 1 for nearly constant KPI streams and decreases as the underlying time-series becomes more dynamic or fluctuating. To mitigate the over-amplification of anomaly scores in high-ρ scenarios, we apply a correction factor $C(\rho)$ that exponentially scales the combined threshold T_3:

$$C(\rho) = \left(\frac{1}{(1-\rho)+\varepsilon}\right)^k, \tag{8}$$

where $k = 1.15$ is a scaling coefficient empirically validated in our real-world 5G core network, which determines the level of exponential adjustment for compensating anomaly score spikes in sparse KPI streams, and ε is a small positive constant introduced to avoid division by zero.

6) Final Threshold (T_{final})

The final threshold for each KPI stream is defined as:

$$T_{\text{final}} = C(\rho) \cdot T_3 . \tag{9}$$

During real-time inference, the model generates an anomaly score s_t for each KPI based on its recent time-series behavior, and an anomaly is detected when

$$s_t > T_{\text{final}}. \tag{10}$$

The final threshold T_{final} integrates all preceding components, providing dynamic adaptation to the distributional

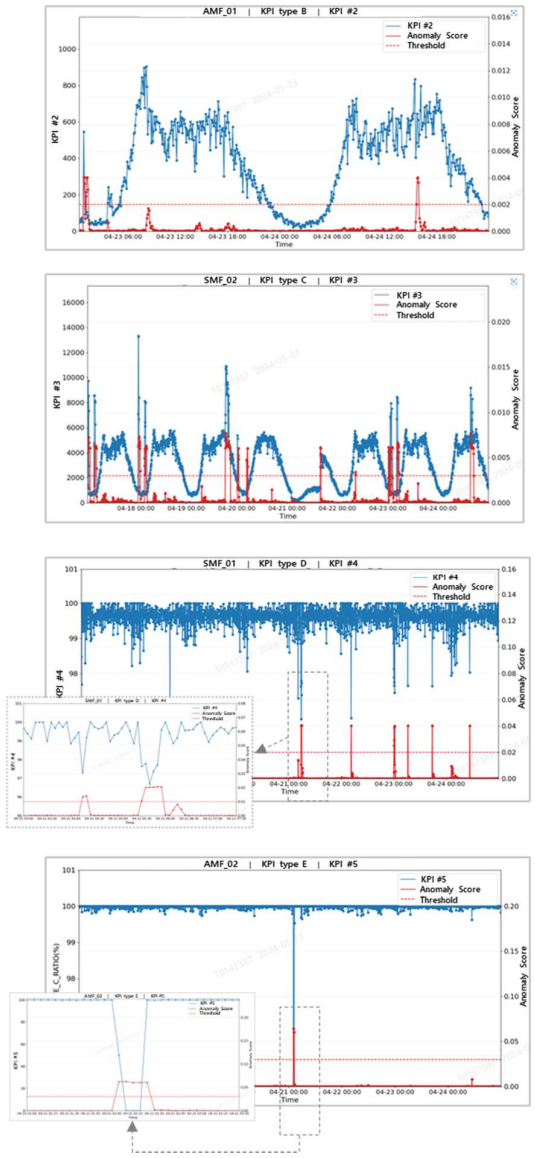

Fig. 4. Sample results for four KPIs after applying the proposed PAA and DAAT methods in our 5G core network.

characteristics of each KPI stream and enabling stable and reliable anomaly detection even for highly skewed or near-constant score distributions.

C. Summary of Framework Contributions

The main technical contributions of this work are summarized as follows:

1) Improved Responsiveness to Point Anomalies:

We introduce a lightweight point-level anomaly amplification mechanism that enhances detection responsiveness by increasing the sensitivity of anomaly scores to local variations. This is achieved without requiring any modification or retraining of the backbone detection model.

2) Automated Threshold Adaptation for Per-KPI Anomaly Scores:

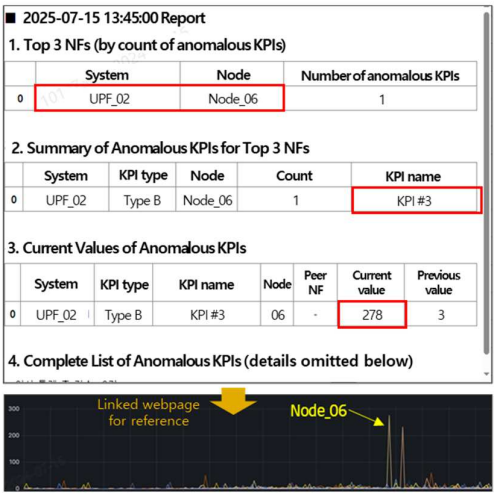

Fig. 5. Example of anomaly detection summary report generated every five minutes using the proposed framework in our real-world 5G core network.

We propose an adaptive thresholding method that automatically calibrates anomaly score thresholds for each KPI by jointly analyzing measured and estimated score distributions. This approach addresses the limitations of conventional Z-score-based thresholding, particularly in large-scale TSAD scenarios.

Overall, these contributions enable practical, reliable, and scalable anomaly detection across tens of thousands of KPI streams in real-world 5G core networks.

IV. PoC VERIFICATION

To validate the feasibility of deploying the proposed framework in a real-world operational environment, we conducted a proof-of-concept (PoC) using real network data from our 5G core network. The PoC involved 10 major network functions (NFs), including AMF, SMF, UPF, and CSCF, covering approximately 37,000 large-scale time-series KPIs. The objective was to verify that the framework, incorporating both anomaly detection responsiveness enhancement and automated per-KPI threshold determination, operates reliably at scale in the real-world network environment. As outlined in Figure 1, KPI data are collected every five minutes from each NF via the EMS and stored in a dedicated database server. The AI-based anomaly detection server processes approximately 37,000 KPI streams within the same five-minute interval, automatically performing time-series anomaly detection using the proposed method and generating a summary report of the results.

Figure 3 illustrates the system performing time-series anomaly detection over a period that includes the time when an anomaly event occurred due to a network maintenance activity in a commercial network. The blue curve, labeled as KPI #1, represents the time series of the initial attempt count within the SMF, the red solid line indicates the anomaly score, and the red dashed line denotes the threshold. The upper graph shows the case before PAA was applied, where anomaly detection was delayed relative to the anomaly occurrence. The lower graph presents the case after PAA application, where the anomaly was detected immediately at the occurrence time. In addition, the baseline anomaly score levels in normal conditions remain sufficiently low and comparable to those before PAA application, confirming that the enhancement does not introduce undesired effects under normal operation.

The proposed Point-Level Anomaly Amplification (PAA) and Distribution-Aware Auto-Thresholding (DAAT) methods were applied for this PoC to a large set of KPIs in our 5G core network, and their effectiveness in performing time-series anomaly detection was verified in close collaboration with our network operations team. Please refer to Figure 4 for related examples of four KPIs. For confidentiality, the specific KPI names are anonymized and denoted in a generic form (e.g., “KPI type A” and “KPI #1”). The system also provides the operators with an anomaly detection summary report, such as the one shown in Figure 5, every five minutes for the anomalies detected using the proposed method. Through this PoC, we verified that the proposed framework for large-scale time-series anomaly detection can be reliably deployed in a real-world 5G core network.

V. CONCLUSION

This paper presented a practical framework for AI-based large-scale time-series anomaly detection. The framework integrates two key techniques. The first is a Point-Level Anomaly Amplification (PAA) mechanism that enhances detection responsiveness by amplifying early anomaly signals. The second is a Distribution-Aware Auto-Thresholding (DAAT) method that automatically calibrates per-KPI anomaly score thresholds based on their distributional characteristics. We also described the detailed algorithms of each component to provide a clear understanding. Through PoC validation in a real-world 5G core network, covering approximately 37,000 KPIs and generating anomaly detection reports for operators every five minutes, we demonstrated that the proposed framework can be reliably deployed for monitoring tens of thousands of KPIs at scale. In the future, we aim to further enhance the scalability and adaptability of the framework and refine the algorithms to improve practical detection capability, in preparation for the evolution toward AI-native networks.

REFERENCES

[1] 3GPP, *System Architecture for the 5G System (5GS)*, 3GPP TS 23.501 V18.2.0, ETSI, Mar. 2025.

[2] ITU-R, *Framework and Overall Objectives of the Future Development of IMT for 2030 and Beyond (IMT-2030)*, Recommendation ITU-R M.2160-0, 2023.

[3] S. Maleki, S. Maleki, and N. R. Jennings, “Unsupervised anomaly detection with LSTM autoencoders using statistical data-filtering,” *Applied Soft Computing*, vol. 108, Art. no. 107443, 2021.

[4] D. Li, D. Chen, J. Goh, and S. Ng, “MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks,” in *Proc. Int. Conf. Artificial Intelligence (AI)*, 2019, pp. 703–716.

[5] J. Xu, H. Wang, J. Long, and Z. Wang, “Anomaly Transformer: Time series anomaly detection with association discrepancy,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2022.

[6] Y.-J. Cho, Y.-S. Kim, S. Kim, D. Sim, D. Kwak, and J. Lee, “AI-Enabled Wireless KPI Monitoring and Diagnosis System for 5G Cellular Networks,” in *Proc. Int. Conf. Information and Communication Technology Convergence (ICTC)*, Jeju, South Korea, Oct. 2019, pp. 899–901.

[7] R. Chalapathy and S. Chawla, “Deep learning for anomaly detection: A survey,” *arXiv preprint* arXiv:1901.03407, 2019.

[8] W. Ryu, Y.-S. Kim, Y. Jo, S.-Y. Kwon, H.-J. Lee, and D. Kwak, "Anomaly detection and root cause analysis for wireless communication radio unit using Adversarial Auto Encoder," in *Proc. Conf. Korean Institute of Communications and Information Sciences (KICS)*, 2022, pp. 114–115.

[9] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. C. Aggarwal, and M. Salehi, "Deep learning for time series anomaly detection: A survey," *arXiv preprint* arXiv:2211.05244, 2022. (last revised May 2024).

[10] A. S. Yaro, F. Maly, and P. Prazak, "Outlier Detection in Time-Series Receive Signal Strength Observation Using Z-Score Method with Sn Scale Estimator for Indoor Localization," *Applied Sciences*, vol. 13, no. 6, art. 3900, pp. 1–16, Mar. 2023.