Intelligent Human Systems Integration (IHSI), Vol. 200, 2026, 199–206
https://doi.org/10.54941/ahfe1007079

Synthetic Network Metric Generation via Conditional DDPM With Categorical and Continuous Log-Metric Conditions

Sangwon Oh[1], Hoyong Ryu[2], Jaehyung Park[1], and Jinsul Kim[1]

[1]Chonnam National University, Gwangju, South Korea
[2]Electronics and Telecommunications Research Institute, Daejeon, South Korea

ABSTRACT

Recent network systems increasingly rely on synthetic data for tasks such as anomaly detection, performance analysis, and digital-twin-based evaluation. However, most existing generators focus solely on metric time-series and overlook the contextual information embedded in operational logs. As a result, they fail to reproduce the joint behavior that emerges when metric fluctuations are closely linked to event-driven operational states. To address this limitation, we develop a conditional denoising diffusion probabilistic model (DDPM) that generates metric sequences using both categorical and continuous conditions derived from metrics and logs. These heterogeneous conditions are transformed into a unified vector and injected into the diffusion process, enabling the model to capture dependencies between system events and metric dynamics. Experiments on real network traces demonstrate that our conditional diffusion models—based on U-Net, CSDI, and SSSD architectures—substantially outperform unconditional diffusion baselines and show strong fidelity and downstream utility. These findings indicate that context-aware diffusion modeling provides a robust foundation for synthetic metric generation in AIOps and digital-twin environments where access to real operational data is limited.

Keywords: Computer network, Deep learning, Synthetic generation model, Diffusion model

INTRODUCTION

Modern networked systems generate rich operational data streams composed of performance metrics, logs, and service traces. These data are essential for applications such as anomaly detection, capacity planning, and digital-twin-based simulation. However, collecting sufficiently diverse real-world traces is challenging due to privacy constraints, operational risks, and the rarity of abnormal conditions. This challenge motivates research on synthetic data generation for network intelligence systems (Mariani et al., 2023).

Most prior works generate metric-only sequences without considering log events, despite the fact that real network behavior is inherently multimodal: service operations, configuration changes, failure events, and workload surges often manifest as coordinated patterns across logs and metrics (He et al., 2022). Ignoring this coupling leads to synthetic data that lack temporal coherence and event-driven structure.

Received November 10, 2025; Revised December 2, 2025; Accepted December 17, 2025; Available online February 1, 2026

Recent advances in conditional generative modeling highlight the value of incorporating contextual information into sequence generation. However, many existing approaches rely on GAN-based frameworks that exhibit instability or mode collapse when handling heterogeneous and high-dimensional conditions (Esteban et al., 2017). In contrast, denoising diffusion models (DDPMs) provide stable training and strong expressive power for multimodal distributions (Ho et al., 2020; Song et al., 2021), making them promising candidates for network data generation.

In this paper, we explore conditional diffusion-based methods for synthesizing metric sequences conditioned on operational context. Extending DDPMs with structured categorical and continuous conditions, we investigate different backbone architectures—U-Net, CSDI, and SSSD—that offer varying capabilities in temporal modeling. Through empirical analysis on RCAEval and LEMMA-RCA datasets, we demonstrate that effective contextual conditioning substantially enhances realism and downstream utility. Our contributions are as follows:

1. We formalize context-aware metric synthesis by integrating heterogeneous operational attributes into the diffusion process.
2. We develop a conditional DDPM framework compatible with both Transformer- and S4-based backbones.
3. We show empirically that architectural choices in conditioning injection significantly affect the fidelity and utility of generated sequences.

Related Works

Research on synthetic time-series generation, multimodal network observability, and conditional generative modeling has grown significantly in recent years. In networked systems, logs and metrics have traditionally been analyzed together to support anomaly detection, operational diagnosis, and performance understanding. Early log-analysis models such as DeepLog and LogPai demonstrated that sequential patterns in application logs are tightly coupled with system states, and follow-up studies showed that combining logs with metrics leads to more accurate diagnosis, since event-driven transitions often precede metric deviations (Xu et al., 2019; He et al., 2022). These works highlight the importance of multi-source observability but do not address the challenge of generating synthetic data that captures such joint behaviors.

Generative modeling for multivariate time-series has largely been driven by adversarial architectures. TimeGAN (Yoon et al., 2019) and RC-GAN variants adapt GAN objectives to temporal settings through recurrent generators and discriminators. While effective for short sequences, GAN-based methods often suffer from mode collapse and instability when dealing with heterogeneous conditions, long-term dependencies, or rare-event patterns typical of network telemetry (Esteban et al., 2017). Extensions using conditional GANs introduced auxiliary labels or covariates, but these methods still struggle with capturing non-stationary temporal structures or generating data aligned with complex operational contexts.

Beyond GANs, autoregressive and variational models have been explored to improve stability. Transformer-based models have shown strong expressiveness for long sequences, especially when combined with covariate encoding or probabilistic forecasting (Lim et al., 2021; Zerveas et al., 2021). Variational approaches such as VRNNs or Gaussian latent models add stochasticity but face challenges modeling multimodal operational patterns. Despite progress, these approaches do not fully bridge the gap between discrete log events and continuous metric evolution, limiting their ability to model network dynamics at scale.

Diffusion-based generative models emerged as a promising alternative due to their stable likelihood-based training and capacity to capture rich multimodal distributions (Ho et al., 2020; Song et al., 2021). Diffusion models have achieved state-of-the-art results in audio synthesis (Kong et al., 2020), probabilistic forecasting (Rasheed et al., 2023), and general time-series generation. Conditional extensions have introduced mechanisms for integrating covariates, labels, or structural constraints, enabling context-guided generation in domains such as climate modeling, healthcare, and sensor streams. However, most existing diffusion models assume structured inputs and do not directly address the integration of log-derived features or operational context from distributed systems.

Recent architectures such as CSDI (Tashiro et al., 2021) and SSSD (Gu et al., 2021; Gupta et al., 2022) push diffusion modeling toward more complex temporal structures. CSDI adopts a Transformer encoder–decoder that enables fine-grained per-timestep conditioning and effective handling of irregular or partially observed sequences. State-space models such as S4 and subsequent SSSD frameworks introduce efficient kernels that propagate information across long horizons, offering advantages for systems where faults or workload changes propagate gradually across components. Despite these advancements, the application of these architectures to network metric–log synthesis remains underexplored.

In summary, prior work separately advances log–metric analysis, generative time-series modeling, and diffusion-based learning, but few studies combine these threads to synthesize metric sequences grounded in operational context. This gap motivates our proposed conditional diffusion framework, which unifies categorical and continuous log-derived signals with diffusion-based temporal modeling to generate context-aware synthetic network metrics.

Proposed Method

Network telemetry is inherently multimodal: performance metrics evolve continuously while logs capture discrete operational events such as service invocations, configuration updates, and fault occurrences. Because real system behavior emerges from the interaction between these modalities, a generative model must incorporate both to produce realistic synthetic data. Building on this insight, the proposed method transforms diverse log-derived attributes into a structured conditioning vector and injects it into a denoising diffusion probabilistic model (DDPM). The framework enables the generation of metric sequences that respond coherently to operational context.

We first convert operational metadata into conditioning features. Attributes such as system type, dataset identifiers, service names, operation types, fault categories, severity levels, and response codes serve as categorical conditions representing discrete system states. These attributes are encoded using either one-hot representations or learnable embeddings depending on cardinality. In parallel, continuous conditions—including CPU and memory usage, latency and throughput statistics, KPI scores, load intensity, and time-of-day encodings—capture dynamic behavior and are normalized using z-score, min–max scaling, or log-transformation. By concatenating categorical and continuous elements, we form a unified conditional vector that summarizes the operational context of each metric segment.

The generative backbone is based on denoising diffusion probabilistic models (Ho et al., 2020), which synthesize data by reversing a forward corruption process. In the forward process, Gaussian noise is incrementally added to a clean metric sequence, resulting in a fully noisy sample. The reverse process learns to reconstruct the clean data by predicting the noise injected at each step, enabling stable training and strong representational capacity. Given a clean sample x_0, noise schedule β_t and timestep t the diffusion processes are:

$$q\left(x_1 \middle| x_{t-1}\right) = \mathcal{N}\left(\sqrt{1-\beta_t}\,x_{t-1}, \beta_t I\right)$$

$$p_0\left(x_{t-1} \middle| x_t, c\right) = \mathcal{N}\left(\mu_\theta\left(x_t, t, c\right), \sigma_t^2 I\right)$$

$$\mu_\theta\left(x_t, t, c\right) = \frac{1}{\sqrt{1-\beta_t}}\left(x_t - \beta_t \varepsilon_\theta\left(x_t, t, c\right)\right)$$

The neural network $\in_\theta(\cdot)$ is conditioned on operational context c, allowing the reverse trajectory to be shaped by the system's discrete and continuous states.

To explore the influence of architectural choices on context-aware sequence generation, we adopt three class of diffusion backbones.

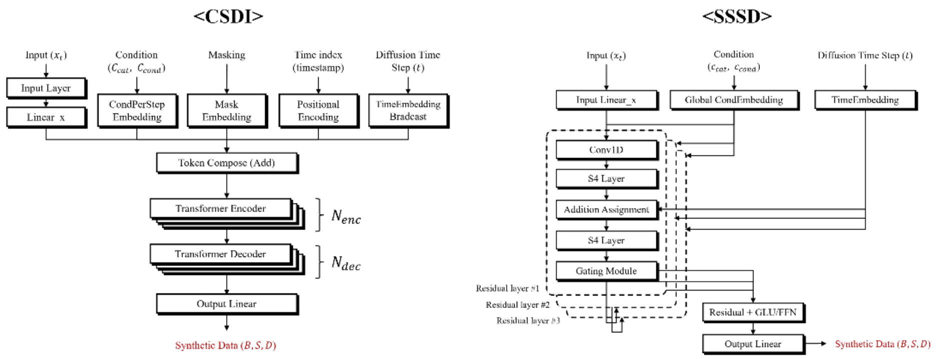

Figure 1: Architectures of CSDI and SSSD.

DDPM(U-Net) provides a baseline approach in which conditional vectors are broadcast to all layers of a 1D U-Net and optionally fused via FiLM modulation. This design captures local temporal structure but has limited ability to model long-range dependencies.

DDPM(CSDI) introduces Transformer-based encoder–decoder layers that embed conditions at each timestep. Through self-attention, positional encoding, and masking mechanisms, CSDI achieves fine-grained alignment between context and metric evolution, making it effective for irregular or context-sensitive sequences.

DDPM(SSSD) leverages structured state-space models (S4), which propagate information across long horizons efficiently. By injecting conditions at the sequence level and modulating transitions through FiLM or gating, this architecture supports global context propagation, capturing the slow-moving operational drifts and fault-induced transitions characteristic of network telemetry. The three backbones thus offer complementary inductive biases, enabling a comprehensive comparison of how architectural design impacts conditional diffusion modeling.

Overall, the proposed framework unifies heterogeneous operational conditions with diffusion-driven generative modeling, enabling the synthesis of metric sequences that exhibit both temporal realism and context consistency—properties essential for downstream tasks in AIOps and digital-twin environments.

To better model temporal dependencies and condition integration, we employ two DDPM architectures suited for time-series generation: CSDI and SSSD.

Experiments

Experiments were performed to evaluate both the fidelity and the practical utility of the proposed method across two real-world datasets: RCAEval and LEMMA-RCA. These datasets contain multivariate metric traces aligned with log-derived contextual attributes, making them well suited for assessing how generative models respond to operational conditions. RCAEval includes microservice traces with injected faults, diverse service interactions, and workload variations, while LEMMA-RCA expands the domain to include IT cloud platforms and OT industrial systems, capturing a broader spectrum of system behaviors.

To assess the effect of architectural choices, we compare five models: DDPM(U-Net), DDPM(CSDI), DDPM(CSDI without masking), DDPM(SSSD with FiLM), and DDPM(SSSD with Gating). These models represent varying conditioning mechanisms and temporal modeling capabilities, allowing us to examine how each design impacts the generator's ability to reflect operational context. All models use identical noise schedules and loss functions, isolating the contribution of architectural differences.

Table 1: DDPM experiment result (similarity).

Dataset	Metric	U-Net	CSDI	CSDI w/o Mask	SSSD w FiLM	SSSD w Gating
RCAEval	CS_TSF	0.4102	0.8290	0.5618	0.8471	0.8714
	JSD_TSF	0.2874	0.1428	0.2312	0.1382	0.1350
	FD_TSF	11.5210	6.9104	9.8427	6.7299	6.5462
LEMMA-RCA	CS_TSF	0.3985	0.8164	0.5527	0.8342	0.8539
	JSD_TSF	0.3012	0.1531	0.2449	0.1497	0.1468
	FD_TSF	12.1043	7.1029	10.0238	6.9813	6.8041

Table 2: DDPM experiment result (utility).

Method	Min-max scaling		Standard scaling		Avg
	LSTM	CNN	LSTM	CNN	
U-Net	0.6240	0.6098	0.5670	0.5767	0.5944
CSDI	0.8009	0.8108	0.7753	0.7893	0.7941
CSDI w/o mask	0.7632	0.6535	0.6245	0.5751	0.6541
SSSD w FiLM	0.8009	0.7755	0.7450	0.7532	0.7687
SSSD w Gating	0.8209	0.8068	0.8338	0.8221	0.8209
Original	0.8142	0.8324	0.8509	0.8574	0.8387

We evaluate the fidelity of synthetic sequences using three complementary metrics. CS_TSF measures feature-level cosine similarity and captures local temporal alignment. JSD_TSF evaluates distributional divergence between real and synthetic representations, providing insight into statistical consistency. FD_TSF, a Fréchet-style distance, quantifies global distance in feature space, with lower values indicating higher similarity. Across both datasets, diffusion models with stronger temporal modeling—particularly CSDI and the SSSD variants—significantly outperform the U-Net baseline. The gating-enhanced SSSD architecture achieves the lowest FD_TSF values and the highest similarity scores overall, demonstrating its advantage in propagating contextual information across long sequences. The degradation observed in the CSDI model without masking highlights the importance of attention masking when modeling condition-dependent time-series.

To evaluate downstream utility, we conduct anomaly classification experiments using synthetic data as a drop-in replacement for original traces. LSTM-based and CNN-based classifiers are trained separately using datasets generated by each diffusion model, and their performance is compared against classifiers trained on real data. Two normalization schemes—min–max and standard scaling—are employed to test robustness. The results show strong correlation with fidelity metrics: classifiers trained on SSSD-generated data perform closest to those trained on real data, with SSSD-Gating achieving the highest average accuracy (0.8209). CSDI-generated data also yields competitive accuracy, while U-Net–generated data performs substantially worse across all settings. These observations indicate that precise condition injection and robust temporal modeling are critical for retaining downstream discriminative structure.

Collectively, the experiments demonstrate that context-aware diffusion models—and especially architectures with long-range modeling and gated conditioning—produce synthetic metric sequences that not only resemble real data but also preserve meaningful operational semantics. These properties are crucial when synthetic data is used for AIOps, anomaly detection, data augmentation, or digital-twin simulation, where both fidelity and utility determine effectiveness.

CONCLUSION

This study presented a conditional diffusion-based framework for synthesizing network metric time-series using contextual information derived from both metrics and log events. Unlike traditional metric-only generators, the proposed approach incorporates heterogeneous operational attributes—such as system type, service identity, fault category, request patterns, and temporal indicators—by transforming them into structured categorical and continuous conditions. By integrating these conditions into the denoising diffusion process, the model can generate sequences that more faithfully reflect the dynamics of real network environments.

Through extensive evaluation on RCAEval and LEMMA-RCA, we demonstrated that the effectiveness of metric generation strongly depends on the model's ability to inject and propagate contextual information. Transformer-based conditional diffusion (CSDI) enhances local temporal consistency through per-step conditioning, while the SSSD architecture shows superior performance by leveraging structured state-space models that capture long-range dependencies. In particular, the gating-enhanced SSSD variant achieved the highest similarity metrics and the strongest downstream utility, enabling classifiers trained on synthetic data to approximate the performance obtained when training on real data.

These results confirm that context-aware diffusion models are a powerful tool for generating realistic and operationally meaningful synthetic network data. The proposed framework not only supports high-fidelity data generation but also offers practical value for AIOps, anomaly detection, and digital-twin simulation, where reliable synthetic data can mitigate the challenges posed by limited, sensitive, or costly real-world traces. Future work may explore extending this framework to multimodal generation, variable-length sequences, or online adaptation to evolving network conditions, further enhancing the applicability of diffusion models in real-world network intelligence systems.

ACKNOWLEDGMENT

This work was supported by the Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government (MSIT) (Project Name: Development of digital twin-based network failure prevention and operation management automation technology, Project Number: RS-2025-00345030, Contribution Rate, 100%).

REFERENCES

Esteban, C., Hyland, S.L., Rätsch, G. (2017). "Real-valued time-series generation with recurrent conditional GANs." Advances in Neural Information Processing Systems.

Gu, A., Goel, K., Ré, C. (2021). "Efficiently modeling long sequences with structured state spaces." International Conference on Learning Representations.

Gupta, A., Chiang, P., Gulrajani, I., et al. (2022). "S4ND: Modeling long sequences with structured state spaces in N-dimensions." International Conference on Machine Learning.

He, Q., Zhu, Y., Li, Z. (2022). "Log-and-metric based anomaly detection in large-scale distributed systems." IEEE Transactions on Network and Service Management, Vol. 19, No. 1, pp. 362–374

Ho, J., Jain, A., Abbeel, P. (2020). "Denoising diffusion probabilistic models." Advances in Neural Information Processing Systems.

Kong, Z., Ping, W., Huang, J., Zhao, K., Catanzaro, B. (2020). "DiffWave: A versatile diffusion model for audio synthesis." International Conference on Learning Representations.

Lim, B., Arık, S.Ö., Loeff, N., Pfister, T. (2021). "Temporal fusion transformers for interpretable multi-horizon time-series forecasting." International Journal of Forecasting, Vol. 37, No. 4, pp. 1748–1764

Mariani, G., et al. (2023). "Synthetic data for intelligent operations: A survey." ACM Computing Surveys.

Rasheed, F., Wang, S., Ding, Y. (2023). "A survey on diffusion models for time-series forecasting and generation." Engineering Applications of Artificial Intelligence.

Song, Y., Sohl-Dickstein, J., Kingma, D., et al. (2021). "Score-based generative modeling through stochastic differential equations." International Conference on Learning Representations.

Tashiro, Y., Song, J., Song, Y., et al. (2021). "CSDI: Conditional score-based diffusion models for probabilistic time-series imputation." Advances in Neural Information Processing Systems.

Xu, W., Huang, L., Fox, A., et al. (2019). "Detecting large-scale system problems by mining console logs." ACM Symposium on Operating Systems Principles.

Yoon, J., Jarrett, D., van der Schaar, M. (2019). "TimeGAN: Generating time-series via generative adversarial networks." Advances in Neural Information Processing Systems.

Zerveas, G., Jayaraman, S., Lalli, M., et al. (2021). "A transformer-based framework for multivariate time series representation learning." International Conference on Knowledge Discovery & Data Mining.