Late Breaking Results: A Diffusion-Based Framework for Configurable and Realistic Multi-Storage Trace Generation

Seohyun Kim[1], Junyoung Lee[1], Jongho Park[1]
Jinhyung Koo[2] Sungjin Lee[2] and Yeseong Kim[1]
[1]DGIST, [2]POSTECH
{selenium, lolcy3205, psyractal1123, yeseongkim}@dgist.ac.kr, {jhk5361, sungkin.lee}@postech.ac.kr

2025 62nd ACM/IEEE Design Automation Conference (DAC) | 979-8-3315-0304-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/DAC63849.2025.11132678

***Abstract*—We propose DiTTO, a novel diffusion-based framework for generating realistic, precisely configurable, and diverse multi-device storage traces. Leveraging advanced diffusion techniques, DiTTO enables the synthesis of high-fidelity continuous traces that capture temporal dynamics and inter-device dependencies with user-defined configurations. Our experimental results demonstrate that DiTTO can generate traces with high fidelity and diversity while aligning closely with guided configurations with only 8% errors.**

I. Introduction

Understanding and optimizing the performance of distributed storage systems require detailed workload analysis in various environments such as RAID arrays and CephFS. Realistic workload traces, which capture read and write operations across multiple devices, are crucial for identifying performance bottlenecks and evaluating optimization strategies such as caching policies and load balancing. However, collecting real-world traces is costly and often impractical due to high instrumentation overheads and privacy concerns. This lack of accessible traces hinders the development of data-driven storage optimizations and limits the evaluation of emerging techniques.

To mitigate this, synthetic trace generation tools such as SPEC Storage allows for controlled workload simulation. While useful, these tools rely on predefined templates and fail to capture the complex, evolving behavior of real-world workloads. Machine learning-based methods [1] attempt to improve realism but are still constrained by excessive feature engineering by human experts to identify application-specific characteristics.

In this paper, we propose DiTTO (**Di**ffusion-based **T**race generation and **T**emporal **O**utpainting), a *diffusion-based* generative framework for producing realistic and configurable multi-device storage traces. Unlike conventional generative approaches that rely on predefined patterns or static sampling, our framework provides *precise, user-controlled*, and *arbitrary-length* trace generation, enabling workloads to be synthesized according to quantifiable constraints such as read/write ratios, access burstiness, and temporal dependencies.

DiTTO transforms raw access logs, including timestamps, operation types, and device identifiers, into structured multi-channel representations, enabling diffusion models to capture complex temporal dependencies and cross-device correlations. We introduce a workload conditioning mechanism that explicitly embeds quantifiable user-defined parameters into the generative process, providing fine-grained control over trace characteristics. Unlike conventional text-based conditioning in image diffusion models [2], our approach ensures explicit and

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No.2022-0-00991, 1T-1C DRAM Array Based High-Bandwidth, Ultra-High Efficiency Processing-in-Memory Accelerator). This work was also supported by the Technology Innovation Program (RS-2024-00445759, Development of Navigation Technology Utilizing Visual Information Based on Vision-Language Models for Understanding Dynamic Environments in Non-Learned Spaces) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea)

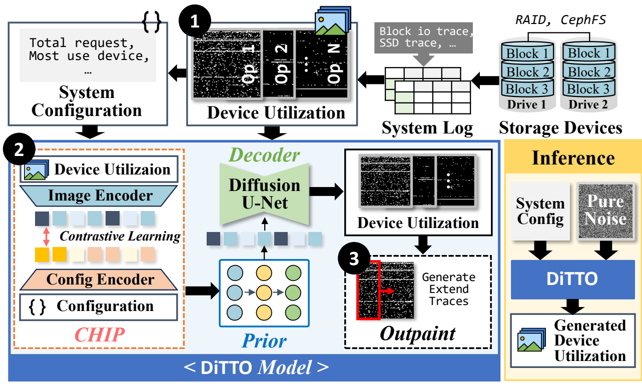

Fig. 1. Overview of DiTTO

constrained output control. Additionally, we design a sparsity-aware training approach to ensure that long periods of inactivity and bursty access patterns are accurately modeled. Finally, by leveraging outpainting techniques, a generative technique that extends content while preserving contextual coherence, DiTTO produces traces of arbitrary length while maintaining contextual consistency. In this paper, we make the following contributions.

1) We propose a **diffusion-based synthetic trace generation framework** that accurately captures multi-device workload patterns through structured representations. To the best of our knowledge, it is the first work that utilizes the diffusion technique for the storage trace generation.

2) We present a novel **quantifiable workload conditioning mechanism** that ensures quantifiable control over key workload properties such as read/write ratios.

3) We design **sparsity-aware training** and **outpainting-based trace generation** techniques to generate workloads with long inactive periods and extended trace durations.

4) We show that DiTTO can generate high-fidelity storage traces that replicate real-world workload characteristics while preserving trace diversity with less than 8% error.

II. Proposed Technique: DiTTO

We propose DiTTO, a novel diffusion-based framework for generating realistic, configurable, and diverse multi-device storage traces. The framework captures intricate temporal dynamics and inter-device dependencies while allowing users to specify workload characteristics. The overall pipeline of DiTTO is illustrated in Fig. 1, consisting of three main stages: ❶ transforming traces into structured image representations, ❷ encoding numeric workload characteristics for explicit conditioning, and ❸ generating arbitrarily long traces using a diffusion-based model enhanced with outpainting.

Transforming system traces into image-based representations. DiTTO first converts raw per-device access logs into structured multi-channel image representations, enabling the diffusion model to learn workload patterns effectively. Each input trace consists of timestamped read and write operations

Authorized licensed use limited to: University of Texas at Austin. Downloaded on September 18,2025 at 19:31:23 UTC from IEEE Xplore. Restrictions apply.

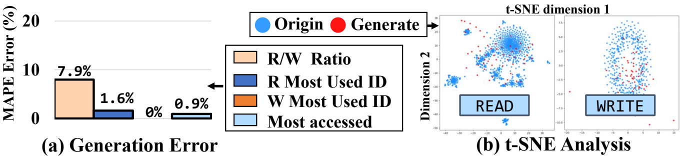

Fig. 2. Generation error and t-SNE

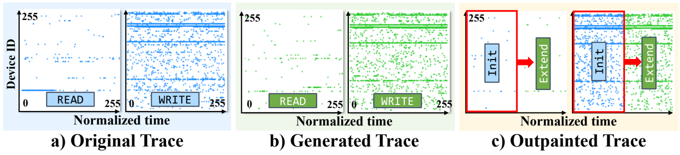

Fig. 3. (a), (b) Comparison between the original and generated traces, (c) Extended trace with collected data and configuration (Outpainting)

mapped to corresponding storage devices. To construct the image representation, we segment time into fixed intervals along the x-axis, while the y-axis corresponds to different storage devices. Each pixel encodes the presence of a read or write event at a specific time interval and device, with separate channels distinguishing operation types. Storage workloads exhibit extreme sparsity, where access events are interspersed with long periods of inactivity. To mitigate this, we introduce local Gaussian noise augmentation, which creates smooth intensity gradients around access events. This prevents excessive sparsity in the image representation, improving the diffusion model to learn contextual relationships across time and devices.

Encoding numeric workload characteristics. A key requirement for synthetic trace generation is the ability to precisely control workload properties including read/write ratios, request rates, and device utilization patterns. To achieve this, we introduce CHIP (Contrastive Hyperconfiguration-Image Pre-training), a novel embedding mechanism that conditions the model on user-defined workload parameters. CHIP operates by mapping *numeric workload configurations* into a latent space that aligns with the image-based representations of system traces. This is achieved via contrastive learning [3], where the model learns to associate similar workload characteristics with similar trace representations. Given a workload configuration, CHIP generates an embedding that guides the diffusion process, ensuring that generated traces adhere to the specified workload properties. Unlike conventional text-based conditioning in image diffusion models, CHIP explicitly enforces *quantifiable constraints*, making it well-suited for structured data synthesis.

Generating continuous traces. Since real-world storage workloads span arbitrary durations, DiTTO must generate long sequences while preserving workload characteristics. To achieve this, we incorporate outpainting, a generative technique that extends trace durations while maintaining contextual coherence. The outpainting works by conditioning the diffusion model on the final states of previously generated traces, allowing seamless extension over time. Specifically, when generating a new trace segment, the model samples latent variables from the boundary of the prior output, ensuring that patterns such as periodic bursts, coordinated device accesses, and workload fluctuations remain consistent. This prevents abrupt transitions and maintains statistical fidelity across extended traces.

III. Experimental Results

We implemented DiTTO using Pytorch 2.4 and evaluated the efficacy using Alibaba Block IO trace dataset. One of the key features of DiTTO is its ability to generate realistic traces that closely adhere to user-specified configurations, such as read/write ratios and workload intensity. Figure 2(a) presents the evaluation results for 50 generated traces, reporting the errors in how well the traces follow the user configurations. The results show that DiTTO achieves an average error of less than 8% for read/write ratios and 2% for device utilization patterns, demonstrating its precision in capturing workload characteristics. Additionally, the t-SNE analysis in Figure 2(b) indicates that the generated traces exhibit trends similar to real (Origin) traces while maintaining sufficient diversity. These results confirm that the CHIP model effectively embeds quantifiable characteristics into the trace generation process. For instance, when tasked with generating traces for a read-heavy workload (82.99% reads), DiTTO produces traces with a read ratio of 81.24%, closely aligning with the intended configuration.

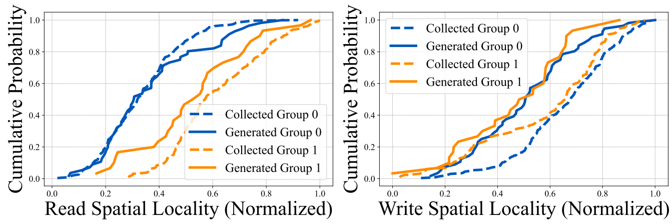

Fig. 4. Distribution of spatial locality for different configuration groups

Figure 3 compares synthetic traces to their real-world counterparts to illustrate the fidelity of DiTTO. The results show that the generated traces (b) closely resemble real-world traces (a) under the same workload configuration, capturing realistic access patterns. Figure 3(c) demonstrates the results of the outpainting technique. The results present that DiTTO extends traces over a longer time horizon while preserving the consistency of previously observed patterns.

To further demonstrate how DiTTO generates distinguishable traces under different conditions, we clustered the collected traces into two groups and then generated traces by sampling the corresponding configurations of each cluster. Figure 4 presents the distribution of real and generated traces in terms of read/write spatial locality. The results indicate that DiTTO successfully generates distinct traces that align with the patterns observed in each cluster, despite spatial locality *not being explicitly included* in the user-configurable parameters. This suggests that DiTTO not only follows explicit workload parameters but also learns and preserves *underlying high-level workload characteristics*, such as spatial locality, by capturing underlying access patterns from real traces.

IV. Conclusion

We propose DiTTO, which generates synthetic traces for distributed storage systems by introducing a diffusion-based framework. By integrating sparsity-aware training, CHIP for fine-grained conditioning, and outpainting for extended durations, DiTTO effectively captures intricate temporal and spatial patterns of real workloads. Our evaluation shows that DiTTO can generate realistic traces aligned with user configurations with the error of less than 8%.

References

[1] Arnab K Paul, et al. Machine learning assisted hpc workload trace generation for leadership scale storage systems. In *Proceedings of the 31st International Symposium on High-Performance Parallel and Distributed Computing*, pages 199–212, 2022.

[2] Aditya Ramesh, et al. Hierarchical Text-Conditional Image Generation with CLIP Latents, 2022.

[3] Alec Radford, et al. Learning transferable visual models from natural language supervision, 2021.

Authorized licensed use limited to: University of Texas at Austin. Downloaded on September 18,2025 at 19:31:23 UTC from IEEE Xplore. Restrictions apply.