Poster: Diffusion-Driven Spatio-Temporal Modeling of Cellular Traffic Generation

Xiaowen Xu
The Hong Kong University of Science and Technology (Guangzhou)

Xiaosi Liu
Shenzhen University

Zhidan Liu*
The Hong Kong University of Science and Technology (Guangzhou)

Zhenjiang Li
City University of Hong Kong

Kaishun Wu
The Hong Kong University of Science and Technology (Guangzhou)

Abstract

Accurate modeling of traffic demand through cellular traffic generation is crucial for optimizing base station deployment. We thus present STOUTER, a **S**patio-**T**emporal diffusi**O**n model for cell**U**lar **T**raffic gen**ER**ation. To effectively capture spatial and temporal dynamics, we pretrain both a temporal graph and a base-station graph, and introduce a Spatio-Temporal Feature Fusion Module (STFFM). On five datasets from two regions, STOUTER reduces Jensen–Shannon Divergence by 52.8% over prior methods, generating distributions that closely match real traffic and aiding downstream planning tasks.

ACM Reference Format:

Xiaowen Xu, Xiaosi Liu, Zhidan Liu, Zhenjiang Li, and Kaishun Wu. 2025. Poster: Diffusion-Driven Spatio-Temporal Modeling of Cellular Traffic Generation. In *The 31st Annual International Conference on Mobile Computing and Networking (ACM MOBICOM '25), November 4–8, 2025, Hong Kong, China.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3680207.3765675

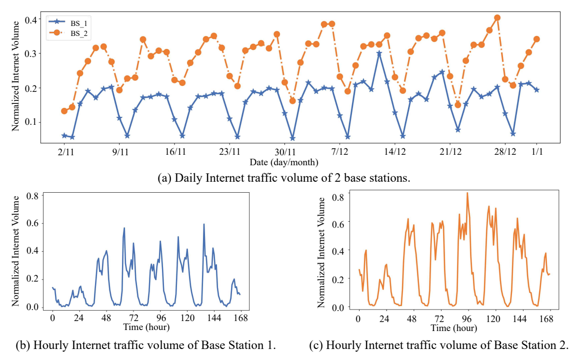

Figure 1: Statistics on (a) daily Internet traffic and (b, c) hourly Internet traffic for two typical base stations, where 'BS' means Base Station.

1 Introduction

Cellular networks serve as essential infrastructure for emerging applications such as smart cities and autonomous systems. With the rapid adoption of 5G and beyond, growing traffic demands call for efficient network resource planning and deployment. However, existing traffic prediction methods face key limitations: they rely heavily on large-scale historical data and often lack access to real-time operator information, hindering scalability and practical usage. Moreover, suboptimal base station deployment strategies frequently overlook actual demand distributions. While existing data generation approaches [2, 6], such as generative adversarial networks (GANs) or autoregressive models, have shown promise, they often lack scalability, diversity, and the ability to model real-world uncertainty.

To address these challenges, we propose STOUTER[1], a spatio-temporal diffusion model that generates realistic cellular traffic by learning both spatial heterogeneity and temporal periodicity, while simulating uncertainty via a denoising diffusion framework. Our approach enables scalable and realistic traffic generation to support data-driven network optimization.

*Corresponding author.

[1]This poster is based on our previous work [4].

ACM MOBICOM '25, Hong Kong, China

ACM ISBN 979-8-4007-1129-9/25/11
https://doi.org/10.1145/3680207.3765675

2 Spatio-temporal Graph Modeling

By visualizing Internet traffic data from two typical base stations in the Milan dataset (Figure 1), we identify three key patterns in cellular traffic that realistic generation models must consider.

Pattern 1: Temporal Periodicity. Traffic exhibits both long-term (e.g., lower weekend vs. higher weekday volumes) and short-term (daily peaks and troughs by hour) cycles.

Pattern 2: Spatial Heterogeneity. Different base stations show distinct local usage levels—densely populated areas carry more load—yet share global trends, necessitating region-specific modeling without losing overall correlations.

Pattern 3: Traffic Uncertainty. Even at the same station and time, volumes fluctuate unpredictably due to varying user behaviors and application demands, so generated data must reflect this inherent randomness rather than producing fixed values.

To capture the temporal periodicity of cellular traffic, we model hourly statistics for a week (Sunday to Saturday) and divide a single day into 24 graph nodes, with directed hourly edges connecting consecutive hours and daily edges linking the same hour across days of the week. A Graph Isomorphism Network (GIN) -based autoencoder is adopted to learn node embeddings that capture both short-term (hourly) and long-term (weekly) traffic cycles.

To effectively differentiate cellular traffic characteristics across base stations, we construct the base station graph, where nodes are base stations, with undirected edges between any two stations within $1km$. Each node's initial features count surrounding Points of Interest (POIs). A graph convolutional network (GCN) -based autoencoder is applied to encode spatial heterogeneity across the network.

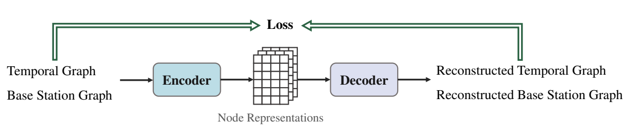

Figure 2: Spatio-temporal graph autoencoder.

These two graph encoders are pretrained and utilized to obtain latent representations of the corresponding graph nodes, preserving essential spatio-temporal features. As shown in Figure 2, both the temporal and base-station graphs are trained via an encoder-decoder framework to produce node embeddings. We optimize the encoders GIN and GCN by minimizing the scaled cosine error (SCE) loss function, which ensures that the learned node representations retain essential structural and feature information:

$$\mathcal{L}_{pre} = \frac{1}{|\mathcal{V}|} \sum_{v_i \in \mathcal{V}, h_i \in \mathcal{H}, z_i \in \mathcal{Z}} \left(1 - \frac{h_i^T z_i}{\|h_i\| \cdot \|z_i\|}\right)^{\gamma}, \quad (1)$$

where the scaling factor $\gamma > 1$, v_i represents the final node embedding from $\mathcal{V}_t$ or $\mathcal{V}_s$, h_i denotes the original node feature from $\mathcal{H}_t$ or $\mathcal{H}_s$, and z_i is the restored node feature from the decoder.

3 Diffusion-based Traffic Generation

Once the temporal embeddings $\mathcal{F}_t$ and spatial embeddings $\mathcal{F}_s$ are learned, STOUTER can be trained and used to generate traffic data via a denoising diffusion process guided at every step by these priors. Figure 3 illustrates the framework of STOUTER, which employs a denoising network that integrates spatio-temporal information to iteratively refine generated data and produce large-scale cellular traffic data.

The diffusion model simulates traffic uncertainty through a two-phase process: forward process and reverse denoising. Real traffic undergoes a forward diffusion that incrementally adds Gaussian noise, then a reverse denoising via STUnet refines samples from pure noise back into realistic traffic sequences. STUnet's U-Net backbone is enriched by a Spatio-Temporal Feature Fusion Module (STFFM), which injects both pretrained temporal and spatial embeddings at each denoising step. $\mathcal{F}_t$ and $\mathcal{F}_s$ are concatenated and projected with a fully-connected layer, and timestep embedding encoded by sinusoidal position is then added to obtain a fused context steering the denoising toward realistic spatial heterogeneity and temporal cycles.

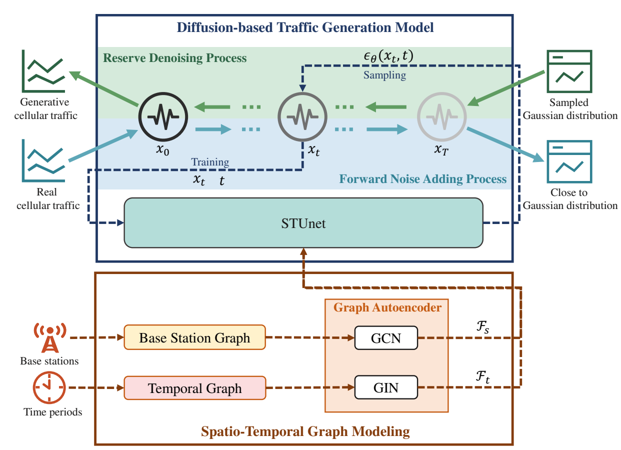

Figure 3: Framework of STOUTER.

4 Performance Evaluation

We evaluate STOUTER on two real-world Call Detailed Records (CDR) datasets: Milan and Trentino[1], each covering five

Table 1: Performance comparisons of our proposed STOUTER and baseline methods using the metrics of MAE, RMSE, and JSD ($\times 10^{-4}$). Detailed results for Internet, Call, and SMS datasets from Milan and Trentino. The optimal results are highlighted in bold and the suboptimal results are underlined.

Methods	Milan								
	Internet			Call			SMS		
	MAE	RMSE	JSD	MAE	RMSE	JSD	MAE	RMSE	JSD
TCN-GAN	0.4183	0.4693	2.6895	0.1357	0.1598	1.0161	0.2714	0.4335	9.2472
VAE	<u>0.1057</u>	<u>0.1444</u>	<u>0.8017</u>	<u>0.0818</u>	<u>0.1071</u>	<u>0.4765</u>	<u>0.0612</u>	<u>0.0735</u>	0.2652
DiffWave	0.1329	0.1707	0.9199	0.0860	0.1086	0.5573	0.0721	0.0837	<u>0.2213</u>
STOUTER	**0.0781**	**0.1088**	**0.4843**	**0.0697**	**0.0925**	**0.3963**	**0.0583**	**0.0715**	**0.1997**
Methods	Trentino								
	Internet			Call			SMS		
	MAE	RMSE	JSD	MAE	RMSE	JSD	MAE	RMSE	JSD
TCN-GAN	0.4786	0.5312	2.9059	0.1267	0.1488	0.8921	0.0977	0.1103	0.3402
VAE	<u>0.1044</u>	<u>0.1440</u>	<u>0.7938</u>	0.0810	0.1018	<u>0.3520</u>	<u>0.0721</u>	<u>0.0896</u>	**0.3369**
DiffWave	0.0986	0.1348	0.8333	<u>0.0809</u>	<u>0.1005</u>	0.5050	0.1540	0.1668	0.4479
STOUTER	**0.0827**	**0.1139**	**0.5196**	**0.0719**	**0.0915**	**0.3076**	**0.0613**	**0.0808**	<u>0.3380</u>

traffic types (Internet, Call, SMS, Incoming-Call, Outgoing-Call or aggregated Call/SMS). Three representative generation methods are selected as baselines: a TCN-GAN[2], a convolutional VAE[5], and a DiffWave-style diffusion model[3] with a WaveNet backbone. Models are assessed by Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Jensen-Shannon Divergence (JSD) between generated and real traffic distributions.

Table 1 presents the results of the performance evaluation of our model compared to baseline models in multiple datasets from Milan and Trentino. STOUTER achieves the lowest error and divergence in nearly all settings. On the Milan Internet dataset, it reduces MAE by 26.1%, RMSE by 24.6%, and JSD by 39.5% compared to the best baseline. Similar trends are observed in Trentino, where STOUTER lowers JSD by 34.5% on Internet data. These results confirm STOUTER's ability to generate realistic, distributionally aligned cellular traffic at scale.

5 Conclusion

In this work, we propose STOUTER, a spatio-temporal fusion diffusion model for cellular traffic generation. By embedding spatio-temporal relationships into the diffusion process, it produces realistic, high-quality synthetic traffic. We validate STOUTER on large-scale real-world datasets, showing significant improvements across multiple metrics over existing generative models, and demonstrate its effectiveness for downstream tasks such as network optimization, traffic prediction, and resource allocation.

Acknowledgments

This work was supported in part by National Natural Science Foundations of China under Grant 62172284 and the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things under Grant 2023B1212010007.

References

[1] Gianni Barlacchi, Marco De Nadai, Roberto Larcher, Antonio Casella, Cristiana Chitic, Giovanni Torrisi, Fabrizio Antonelli, Alessandro Vespignani, Alex Pentland, and Bruno Lepri. 2015. A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Scientific data* 2, 1 (2015), 1–15.

[2] Shuodi Hui, Huandong Wang, Tong Li, Xinghao Yang, Xing Wang, Junlan Feng, Lin Zhu, Chao Deng, Pan Hui, Depeng Jin, et al. 2023. Large-scale urban cellular traffic generation via knowledge-enhanced GANs with multi-periodic patterns. In *ACM SIGKDD*. 4195–4206.

[3] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. 2021. DiffWave: A Versatile Diffusion Model for Audio Synthesis. arXiv:2009.09761

[4] Xiaosi Liu, Xiaowen Xu, Zhidan Liu, Zhenjiang Li, and Kaishun Wu. 2025. Spatio-Temporal Diffusion Model for Cellular Traffic Generation. *IEEE Transactions on Mobile Computing* (2025), 1–15.

[5] Stanislau Semeniuta, Aliaksei Severyn, and Erhardt Barth. 2017. A hybrid convolutional variational autoencoder for text generation. In *EMNLP*. 627–637.

[6] Shiyuan Zhang, Tong Li, Shuodi Hui, Guangyu Li, Yanping Liang, Li Yu, Depeng Jin, and Yong Li. 2023. Deep transfer learning for city-scale cellular traffic generation through urban knowledge graph. In *ACM SIGKDD*. 4842–4851.