Exploring Feasibility of Diffusion-Based Traffic Generation for Small-Flow Attacks

Akira Tsujikawa*, Junji Takemasa*, and Yuki Koizumi*
*Graduate School of Information Science and Technology, The University of Osaka

***Abstract*—Diffusion-based traffic generation, where bit-level packet features are learned as image representations, enables augmentation of network traffic datasets. This paper empirically investigates whether diffusion models can reproduce protocol semantics in generated traffic. We find that sequence structures of TCP flags and packet directions are not preserved. To mitigate this issue, we design and evaluate an enhanced image representation, which spatially expands important packet features and explicitly correlates directions and flags.**

***Index Terms*—Network traffic generation, slow attack, diffusion model**

I. Introduction

Network attack traffic data are essential for building network security applications such as intrusion detection systems, but public datasets remain scarce due to privacy concerns. While the CICIDS dataset [1] has been widely used, it often suffers from data quality issues such as mislabeling [2].

Synthetic traffic generation has emerged as a promising solution. GAN-based methods such as NetShare [3] reproduce flow-level statistics such as flow volume and duration, but cannot capture packet-level features such as header values. NetDiffusion [4], a representative diffusion model-based method, addresses this limitation by learning and synthesizing bit-level packet representations in an image form.

However, NetDiffusion has limitations in capturing temporal structures of packet sequences. It fails to preserve protocol semantics such as TCP flag ordering and packet directions, instead relying on post-processing heuristics. This limitation is critical for small-flow attacks, where packet sequences define attack behavior. For example, Slowloris traffic is characterized by repeated exchanges of PSH-ACK packets from the client and ACK packets from the server.

This paper empirically analyzes these limitations in a diffusion-based method for small-flow attacks. Our analysis identifies two key issues: (1) temporal structures are lost due to image downsampling, and (2) the relationship between packet directions and other fields such as TCP flags is not preserved due to independent post-processing. To address these issues, we enhance an image representation by designing two mitigations: (1) spatially expanding important features such as TCP flags, and (2) explicitly encoding packet direction into the representation. Our evaluation shows improved fidelity of protocol semantics in generated flows.

This work was supported by KAKENHI Grant-in-Aid for challenging Exploratory Research 25K22803 and JSPS KAKENHI Grant Number 26K02901.

ISBN 978-3-903176-82-9 © 2026 IFIP

II. Background and Motivation

This section introduces diffusion-based traffic generation and the target attack, followed by our motivation.

A. NetDiffusion

NetDiffusion uses image representations of packet sequences in individual flows to train diffusion models for traffic generation. After generating packet sequences with the diffusion model, it enforces protocol semantics through post-processing.

An image representation is generated from the packet sequence of a flow using nPrint [5]. nPrint encodes each packet as a ternary vector, where absent fields are assigned -1, while present fields are represented as binary values. Each packet vector consists of IP, TCP, UDP and ICMP headers. A flow is then represented as a matrix by stacking packet vectors, forming a 2D representation where rows correspond to packets and columns to header features.

Post processing enforces protocol semantics by introducing TCP handshake behavior and packet direction. A TCP 3-way handshake sequence is prepended to each generated flow, and packet directions are assigned based on transition patterns learned from real traces. Specifically, a two-state Markov model is learned from real traces, where each state corresponds to packet direction (client-to-server or server-to-client), and directions are assigned to generated packets according to the learned transition probabilities.

B. Slow Attacks

Slow attacks are a class of network attacks that exploit protocol semantics through low-rate, long-lived connections rather than high-volume traffic. They are difficult to model using coarse-grained statistical features such as flow traffic volume or duration. Instead, their characteristics lie in the temporal structure of packet sequences within individual flows, including packet ordering, protocol-specific sequences, and directional interactions between end hosts.

A representative example is the Slowloris attack, where an attacker establishes multiple HTTP connections to a victim server and keeps them alive to exhaust its connection pool, as shown in Figure 1. After the 3-way handshake, the client keeps the connection by periodically sending partial HTTP requests at intervals slightly shorter than the server timeout. These requests are sent in TCP packets with PSH and ACK flags so that they are directly transferred to HTTP server process without being aggregated in TCP buffer. The server responds

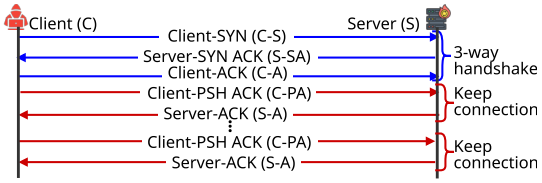

Fig. 1. Packet sequence of Slowloris flow, where X-Y's X and Y show direction (i.e., whether origin is Client or Server) and TCP flag, respectively.

with ACK packets, resulting in repeated exchanges of client-side PSH-ACK (C-PA) and server-side ACK (S-A) packets. This attack serves as a representative case, since it captures temporal structures that characterize small-flow attacks.

C. Motivation

The goal of our traffic generation is to reproduce protocol semantics with high fidelity, while preserving diversity in other packet-level features such as packet sizes. The original NetDiffusion study has demonstrated that diffusion models are effective at generating diverse packet-level features such as packet size distribution. However, diffusion models do not capture protocol semantics. Instead, they are enforced through post-processing heuristics. For example, packet directions are assigned after generation, and TCP flag sequences such as those of the 3-way handshake are explicitly injected.

This limitation is critical for slow attack traffic generation, where temporal structures characterize attack behaviors. This paper aims to identify the root cause of this limitation and explore simple yet effective representations to improve temporal structure preservation.

III. Empirical Analysis

This section analyzes whether NetDiffusion can reproduce sequences of TCP flags and packet directions through a case study of Slowloris traffic. Our analysis reveals key limitations of NetDiffusion for slow attack traffic generation.

A. Analysis Method

We evaluate how well generated flows preserve sequences of TCP flags and packet directions in Slowloris traffic. In Fig. 1, a flow consists of repeated request-response pairs, where a client PSH-ACK packet is followed by a server ACK packet.

To quantify temporal structure preservation, we count consecutive non-overlapping valid (PSH-ACK, ACK) pairs in each flow. In the original Slowloris trace, each flow contains seven such pairs; a higher count indicates better preservation.

We further evaluate protocol semantics with direction by counting consecutive non-overlapping valid (C-PSH-ACK, S-ACK) pairs, where C and S denote client and server.

B. Experiment Setting

We generate Slowloris traces in a local testbed where an attacker client connects to a victim server via a router. The server runs Nginx, and the attacker generates Slowloris traffic using `slowhttptest`. Each flow consists of 20 packets: 3 for the TCP 3-way handshake, 14 for connection maintenance (7 pairs of C-PA and S-A), and 3 for termination.

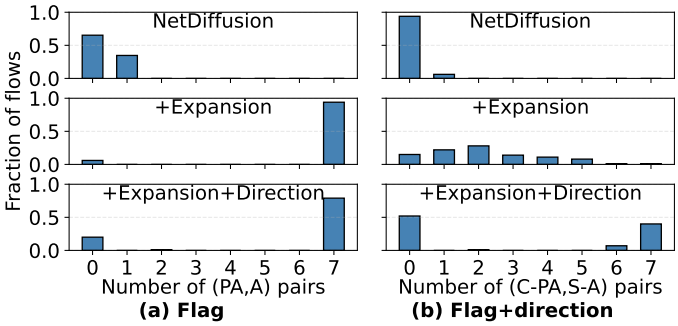

Fig. 2. Distributions of valid packet pairs in generated flows (PA: PSH ACK, A: ACK, C: Client, S: Server). Each flow in the original Slowloris trace contains 7 pairs. NetDiffusion produces almost no (PA, A) or (C-PA, S-A) pairs, failing to preserve temporal structures, while our enhancement (+Expansion+Direction) increases valid pairs.

Following NetDiffusion, we train a diffusion model on the generated Slowloris traces. Stable Diffusion 3 (SD3) is used as the base model and fine-tuned with LoRA. We use 660 training images and 13,200 training steps. The trained model generates packet sequences, after which TCP handshake packets and packet directions are added via post-processing (as in Section 2.1). In total, 100 flows are generated for analysis.

C. Results

In this section, we show that NetDiffusion fails to preserve sequences of TCP flags and packet directions. Through a series of observations, we identify key limitations of NetDiffusion.

Observation 1: Temporal structures of TCP flags are not preserved. Figure 2(a) shows the distribution of valid (PSH-ACK, ACK) pairs in NetDiffusion generated flows. A large fraction of flows contain zero valid pairs, indicating that temporal structures of TCP flag sequences are not reproduced. Specifically, 65% of flows contain no valid pairs, and the remaining 35% contain only one valid pair.

Observation 2: Relationships between directions and flags are not preserved. Figure 2(b) shows the distribution of valid (C-PSH-ACK, S-ACK) pairs. When packet direction is considered, the fraction of flows with zero valid pairs further increases to 94%, indicating that bidirectional interaction patterns are not well captured.

Observation 3: Generated sequences contain invalid packet patterns. To further understand these limitations, we analyze frequently occurring packet sequences using 3-gram statistics over pairs of TCP flags and packet directions. Table I shows the top five most frequent patterns.

In the original trace, request–response patterns dominate (i.e., C-PA $\rightarrow$ S-A $\rightarrow$ C-PA and S-A $\rightarrow$ C-PA $\rightarrow$ S-A, corresponding to the first and second patterns, respectively). In contrast, flows generated by NetDiffusion frequently contain sequences of packets with identical flags (e.g., repeated PSH packets in the first and second patterns) or packets without flags (in the third and fourth patterns), indicating a loss of TCP flag sequence structure. Furthermore, PSH packets are

TABLE I
TOP 5 3-GRAM PATTERNS OVER PAIRS OF TCP FLAGS AND DIRECTIONS IN SLOWLORIS TRACE AND TRACE GENERATED BY NETDIFFUSION.

	Trace	NetDiffusion
1	C-PA → S-A → C-PA (33.28%)	C-P → S-P → C-P (5.85%)
2	S-A → C-PA → S-A (33.28%)	S-P → C-P → S-P (5.73%)
3	C-S → S-SA → C-A (5.56%)	C- → S- → C- (5.54%)
4	S-SA → C-A → C-PA (5.56%)	S- → C- → S- (5.44%)
5	C-A → C-PA → S-A (5.56%)	C-RP → S-RP → C-RP (4.60%)

often generated from the server, although they should always originate from the client in the Slowloris trace. This violates the consistency between packet direction and TCP flags.

D. Lessons Learned

From these observations, we identify two key limitations in NetDiffusion traffic generation.

Sequence structure loss by downsampling: The 3-gram results imply that temporal structures are lost due to downsampling in the diffusion model. In SD3, input images are compressed by a factor of 8 in each dimension through a Variational Autoencoder (VAE). As a result, adjacent packets are mapped into a single latent representation, leading to loss of fine-grained temporal dependencies. The packet sequence (20 packets) is compressed by a factor of 8 along the packet dimension, resulting in a latent height of 20/8 (i.e., 2.5).

Direction independently modeled from flags: As described in Section 2.1, packet directions are assigned through post-processing independently of other fields such as TCP flags. This breaks the relationship between direction and flags, resulting in unrealistic patterns not observed in real traffic, such as PSH packets being generated from the server. Therefore, packet directions and protocol-related fields should be jointly modeled to preserve protocol semantics.

IV. ENHANCED IMAGE REPRESENTATION

Our analysis reveals that (1) diffusion models fail to preserve packet sequences due to downsampling, and (2) independently modeling packet direction and TCP flags degrades protocol semantics. To address these issues, we enhance the image representation based on two principles: making important features robust to downsampling and explicitly encoding the relationship between direction and flags.

A. Design

Spatial expansion: We spatially expand protocol-related features, such as TCP flags, within the image representation, to preserve the structure of these critical features even after the downsampling process. Considering the constraints of SD3 (i.e., a downsampling factor of 8), we expand critical features from 1×1 pixels to 8×8 pixels. This increases input resolution but remains within the 1024×1024 resolution of SD3.

Direction embedding: Although IP addresses are not directly used due to privacy concerns in the original NetDiffusion study, we treat packet direction as a meaningful feature and explicitly embed it into the image. Specifically, the direction column is placed at the right edge of the image and is effectively isolated from other features by adjacent inactive regions (i.e., fields filled with -1). This spatial separation reduces interference from neighboring active features.

B. Evaluation Result

We compare the proposed enhancements with NetDiffusion using the following two configurations. *+Expansion* applies only spatial expansion, while packet directions are assigned through post-processing as in NetDiffusion. *+Expansion+Direction* applies both spatial expansion and direction embedding, enabling traffic generation without relying on post-processing. After generation, random client and server IP addresses are assigned. Figure 2 shows the results.

With +Expansion, the temporal structure of TCP flags (i.e., the number of (PA, A) pairs) is significantly improved, indicating that spatial expansion makes flag information more robust to downsampling. 94% of generated flows completely reproduce 7 (PA,A) pairs. However, since packet direction is still assigned independently through post-processing, the consistency between direction and flags (i.e., the number of (C-PA, S-A) pairs) is only marginally improved.

+Expansion+Direction improves not only the temporal structure of flags but also the consistency between direction and flags, as both are jointly included in the image representation. As a result, 40% of the generated flows contain all seven valid (C-PA, S-A) pairs, indicating that the model can reproduce the bidirectional interaction pattern. However, in both approaches, some flows fail to reproduce valid pairs.

Additionally, in our approach, TCP 3-way handshakes are also generated by the diffusion model without post-processing. In +Expansion+Direction, 65% of the generated flows exhibit valid 3-way handshakes with consistent direction and flags.

V. CONCLUSION

This paper investigates the applicability of diffusion-based traffic generation for small-flow attacks. We show that it fails to preserve sequence structures of TCP flags and packet directions due to image downsampling and independent direction post-processing. We design enhanced image representations based on spatial feature expansion and explicit correlation between directions and flags, and show their effectiveness.

REFERENCES

[1] Canadian Institute for Cybersecurity, University of New Brunswick, "Intrusion detection evaluation dataset (CIC-IDS2017)," accessed: April 9, 2026. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2017.html

[2] G. Engelen, V. Rimmer, and W. Joosen, "Troubleshooting an intrusion detection dataset: The CICIDS2017 case study," in *Proceedings of IEEE Symposium on Security and Privacy Workshops (SPW)*, 2021.

[3] Y. Yin, Z. Lin, M. Jin, G. Fanti, and V. Sekar, "Practical GAN-based synthetic IP header trace generation using NetShare," in *Proceedings of ACM SIGCOMM Conference*, 2022.

[4] X. Jiang, S. Liu, A. Gember-Jacobson, A. N. Bhagoji, P. Schmitt, F. Bronzino, and N. Feamster, "Netdiffusion: Network data augmentation through protocol-constrained traffic generation," in *Proceedings of ACM on Measurement and Analysis of Computing Systems*, 2024.

[5] J. Holland, P. Schmitt, N. Feamster, and P. Mittal, "New directions in automated traffic analysis," in *Proceedings of ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2021.