

Berdasarkan sumber-sumber yang Anda berikan, berikut adalah penjelasan teknis bagaimana WGAN-GP mengatasi data statis dan analisis perbandingan antara menggunakan WGAN saja vs. kombinasi SMOTE+WGAN (Framework EO-WGAN) untuk skenario *burst* trafik Anda.

1. Bagaimana WGAN-GP Mengatasi Data Stabil/Statis?

Masalah utama data statis adalah model generatif biasa (Vanilla GAN) sering mengalami *mode collapse*, di mana model hanya menghasilkan satu variasi data yang "aman" dan gagal menciptakan variasi ekstrem (seperti *burst* atau serangan) karena gradiennya menghilang (*vanishing gradient*) saat distribusi data asli dan buatan tidak saling tumpang tindih 1, 2.

WGAN-GP mengatasi ini dengan dua mekanisme kunci:

- **Wasserstein Distance (Earth Mover's Distance):** Berbeda dengan GAN biasa yang menggunakan Jensen-Shannon divergence, WGAN menggunakan jarak Wasserstein. Ini memberikan gradien yang **lebih halus dan kontinu** bagi generator untuk belajar, bahkan ketika data asli sangat statis dan data buatan masih jauh dari target. Artinya, model tetap bisa belajar cara membuat data "ekstrem" meskipun data tersebut jarang muncul di dataset asli 3-5.
- **Gradient Penalty (GP):** Mekanisme ini menggantikan *weight clipping* (pemangkasan bobot) yang kasar. GP memaksa norma gradien tetap mendekati 1 (aturan 1-Lipschitz), yang menjamin stabilitas pelatihan tanpa membatasi kapasitas model untuk mempelajari fitur-fitur kompleks dan non-linear dari trafik jaringan 4, 6, 7.

Agar Data Tidak "Ngasal" (Graph-Conditioned Guidance): Agar data *burst* yang dihasilkan masuk akal secara fisik (misal: jika *throughput* naik, *packet rate* juga harus naik), Anda tidak bisa hanya mengandalkan WGAN polos. Sumber terbaru 8-10 menyarankan integrasi **Graph-Conditioned WGAN**:

- Buatlah graf hubungan fitur menggunakan korelasi Pearson dan Spearman dari data asli Anda.
- Gunakan graf ini sebagai "panduan" (condition) bagi generator WGAN. Ini memaksa generator untuk mematuhi ketergantungan antar-fitur (misal: hubungan antara *rx_mbps* dan *util_rx_pct*) sehingga data sintetis tetap valid secara protokol jaringan 10, 11.

2. Keunggulan Kombinasi SMOTE + WGAN (EO-WGAN)

Untuk skenario *burst* yang sangat jarang terjadi (minoritas ekstrem), referensi 12, 13 sangat menyarankan pendekatan hibrida yang disebut **EO-WGAN (Enhanced Optimization WGAN)** daripada menggunakan satu metode saja.

Cara Kerjanya:

1. **Tahap 1 (SMOTE):** Menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) untuk melakukan interpolasi linear kasar antar sampel *burst* yang ada. Tujuannya bukan untuk hasil akhir, tapi untuk "memperbanyak" bahan data agar dataset awal lebih seimbang sebelum masuk ke GAN 14.
2. **Tahap 2 (WGAN Refinement):** WGAN dilatih menggunakan data hasil augmentasi SMOTE tersebut. WGAN bertugas untuk "memperhalus" dan memperbaiki distribusi data. WGAN belajar pola non-linear yang kompleks yang tidak bisa ditangkap oleh

SMOTE (yang hanya garis lurus), menjadikan data *burst* lebih realistik dan bervariasi 13, 15.

Keunggulan Kombinasi:

- **Stabilitas Awal:** SMOTE menyediakan distribusi awal yang lebih seimbang, sehingga WGAN tidak perlu belajar dari nol (cold start) pada data yang sangat tidak seimbang, yang sering menyebabkan ketidakstabilan 16.
- **Menangkap Non-Linearitas:** SMOTE saja cenderung menghasilkan data yang kaku (linear) dan *noisy* di perbatasan kelas. WGAN memperbaiki ini dengan menangkap distribusi probabilitas asli yang kompleks 17, 18.
- **Visualisasi t-SNE:** Hasil visualisasi menunjukkan bahwa data *burst* yang dihasilkan EO-WGAN mengisi ruang fitur dengan lebih alami dan membentuk klaster yang jelas dibandingkan SMOTE biasa yang cenderung menyebar acak 18.

3. Mana yang Lebih Baik: Satu Metode atau Kombinasi?

Untuk skenario Anda (membuat variasi kondisi port seperti *overload* dan *burst*), **kombinasi (EO-WGAN) terbukti lebih baik secara statistik.**

Bukti dari *Ablation Study* (Studi Penghapusan Komponen) di referensi 19, 20:

- **Full EO-WGAN (SMOTE + WGAN):** Mencapai F1-Score **93.2%** dan Akurasi **94.6%**.
- **Hanya WGAN (Tanpa SMOTE):** F1-Score turun menjadi **86.7%**. Tanpa "pancingan" awal dari SMOTE, WGAN kesulitan menangkap variasi kelas minoritas (*burst*) dengan sempurna.
- **Hanya SMOTE:** F1-Score hanya **81.4%**. Ini membuktikan bahwa SMOTE saja tidak cukup berkualitas untuk data jaringan yang kompleks.

Kesimpulan & Rekomendasi: Gunakan pendekatan **EO-WGAN (SMOTE + WGAN-GP)**.

1. Gunakan SMOTE untuk melipatgandakan data sampel "*burst*" atau "*overload*" yang sedikit itu secara kasar.
2. Latih WGAN-GP (idealnya dengan panduan *Graph/Correlation Matrix* dari 8) menggunakan data tersebut untuk menghasilkan variasi yang realistik dan "tidak ngasal".
3. Strategi ini akan memberikan dataset latih terbaik untuk model RL Anda agar bisa menangani skenario tak terduga.