

Berdasarkan sumber referensi yang Anda berikan, **WGAN-GP (Wasserstein GAN with Gradient Penalty)** menangani masalah ketidakseimbangan data pada trafik IoT dengan cara menghasilkan data sintetis (buatan) untuk kelas minoritas (seperti serangan siber atau trafik perangkat tertentu) yang **realistik, beragam, dan stabil secara statistik**.

Berikut adalah mekanisme detail bagaimana WGAN-GP bekerja mengatasi masalah ini, yang dapat Anda gunakan sebagai landasan teori dalam paper Anda:

1. Mengatasi Kelemahan GAN Tradisional dan SMOTE

Metode lama seperti SMOTE hanya melakukan interpolasi linear antar titik data, yang sering kali gagal menangkap pola non-linear yang kompleks pada trafik IoT berdimensi tinggi 1, 2. Sementara itu, GAN standar sering mengalami *mode collapse* (hanya menghasilkan satu variasi data) dan ketidakstabilan pelatihan (vanishing gradients) 3, 4.

WGAN-GP mengatasi ini dengan dua inovasi kunci:

- **Wasserstein Distance:** Mengganti fungsi *loss* standar dengan *Earth Mover's distance*. Ini memberikan gradien yang lebih halus dan stabil, memungkinkan model terus belajar meskipun distribusi data asli dan buatan masih berjauhan 5, 6.
- **Gradient Penalty (GP):** Menggantikan teknik *weight clipping* (pemangkas bobot) yang digunakan di WGAN awal. *Clipping* membatasi kapasitas model, sedangkan GP memberikan penalti pada norma gradien agar tetap mendekati 1 (Lipschitz constraint). Ini memastikan model mampu mempelajari fitur trafik yang kompleks tanpa kehilangan stabilitas 7, 8.

2. Strategi Two-Stage (Hybrid) untuk Kualitas Data

Dalam konteks IIoT (Industrial IoT), referensi **Riaz et al. (2025)** mengusulkan framework **EO-WGAN (Enhanced Optimization WGAN)** yang menggabungkan SMOTE dan WGAN-GP 9, 10:

- **Tahap 1 (Inisialisasi):** Menggunakan SMOTE untuk menghasilkan sampel kasar secara cepat guna menyeimbangkan jumlah data awal.
- **Tahap 2 (Refinement):** Melatih WGAN-GP pada data hasil SMOTE tersebut untuk "memperhalus" distribusi. Generator WGAN belajar memperbaiki *noise* dan artifak linear dari SMOTE, menghasilkan sampel yang memiliki variasi non-linear yang lebih realistik dan sesuai dengan distribusi asli trafik 11, 12.

3. Menjaga Integritas Struktural & Semantik (Penting untuk Dataset 37 Fitur Anda)

Agar data buatan tidak sekadar "angka acak", **Alyoubi et al. (2025)** memperkenalkan pendekatan **Graph-Conditioned WGAN** yang dipandu oleh LLM 13, 14:

- **Graph Conditioning:** Menggunakan graf hubungan fitur (dibuat dari korelasi Pearson/Spearman dan *Mutual Information*) sebagai input kondisi bagi Generator. Ini memaksa WGAN untuk mematuhi ketergantungan antar fitur (misalnya: jika rx_mbps tinggi, maka rx_pps juga harus tinggi) 14.
- **LLM Guidance:** Menggunakan *Large Language Model* untuk menetapkan batasan semantik (misal: aturan protokol jaringan) ke dalam fungsi *loss* WGAN. Ini memastikan data sintetis valid secara protokol (misal: tidak ada *negative latency*) 15.

4. Dampak pada Performa Model

Penggunaan WGAN-GP terbukti meningkatkan performa model klasifikasi pada dataset yang tidak seimbang:

- **Mencegah Bias Mayoritas:** Dengan menyeimbangkan dataset, model (seperti IDS atau RL agent) tidak lagi bias ke kelas mayoritas (trafik normal) 16.
- **Peningkatan Metrik:** Eksperimen menunjukkan peningkatan signifikan pada **Recall** dan **F1-Score** untuk kelas minoritas dibandingkan dengan metode tanpa augmentasi atau SMOTE biasa 17, 18.
- **Visualisasi t-SNE:** Visualisasi menunjukkan bahwa data yang dihasilkan WGAN-GP mengisi ruang distribusi kelas minoritas dengan lebih baik dan tidak tumpang tindih dengan kelas mayoritas, menjaga *decision boundary* yang jelas 19, 20.

Kesimpulan untuk Paper Anda: Anda dapat menyimpulkan bahwa WGAN-GP dipilih karena kemampuannya menghasilkan data trafik sintetis yang **valid secara statistik** (berkat *Wasserstein loss*) dan **stabil** (berkat *Gradient Penalty*), serta mampu menangkap korelasi antar-fitur yang kompleks (jika menggunakan pendekatan *Graph-Conditioned*), yang sangat krusial untuk melatih agen Reinforcement Learning agar *robust* terhadap kondisi jaringan yang jarang terjadi.