Berdasarkan literatur yang tersedia, berikut adalah perbedaan mendasar antara **Penalti Berbasis Biner (Hard Penalty)** dan **Penalti Fungsi Sigmoid (Soft/Risk-Sensitive Penalty)** dalam konteks *Reward Function* untuk penegakan SLA (*Service Level Agreement*) pada Reinforcement Learning:

1. Penalti Berbasis Biner (Hard Constraint)

Metode ini menerapkan pendekatan "hitam-putih" terhadap pelanggaran SLA.

- **Mekanisme:** Menggunakan fungsi indikator atau *step function*. Jika metrik kinerja (misalnya latensi) melampaui ambang batas ($Threshold_{SLA}$), agen langsung menerima penalti tetap yang besar. Jika masih di bawah batas, penalti adalah nol (atau agen mendapat reward positif).
- **Karakteristik Matematis:** Bersifat diskrit dan tidak kontinu.
- Contoh: Jika $Latency > 10ms$, maka $Reward = -100$. Jika tidak, $Reward = +1$.
- **Kelemahan:**
- **Informasi Gradien Terbatas:** Agen tidak mengetahui "seberapa buruk" pelanggaran tersebut. Bagi agen, pelanggaran sebesar 0.1 ms diperlakukan sama buruknya dengan pelanggaran 100 ms.
- **Ketidakstabilan:** Dapat menyebabkan *sparse reward* (jarang mendapat sinyal positif) atau osilasi kebijakan karena perubahan nilai reward yang tajam dan tiba-tiba saat melewati ambang batas 1.

2. Penalti Fungsi Sigmoid (Soft Constraint)

Metode ini menggunakan kurva berbentuk 'S' untuk memberikan penalti yang bertahap dan halus.

- **Mekanisme:** Menggunakan fungsi non-linear (seperti logistik atau *hyperbolic tangent*) yang memetakan selisih antara nilai aktual dan target SLA ke dalam rentang penalti kontinu 2.
- **Formula Umum:**$$R_{penalty} = \frac{1}{1 + e^{-k \cdot (Violation - Threshold)}}$$Di mana k adalah parameter kemiringan (*steepness*) dan *Threshold* adalah titik infleksi (batas SLA) 1, 3.
- **Karakteristik Matematis:**
- **Smooth Gradient:** Memberikan sinyal yang dapat diturunkan (*differentiable*). Saat kondisi jaringan mendekati batas bahaya, penalti mulai naik perlahan, memberikan "peringatan dini" kepada agen sebelum pelanggaran fatal terjadi 1.
- **Tunable:** Anda dapat mengatur parameter $c1$ (kemiringan) dan $c2$ (titik tengah) untuk menentukan seberapa agresif penalti diberikan saat mendekati batas SLA 1.
- **Keuntungan:**
- **Sinyal Belajar Lebih Kaya:** Memberikan umpan balik yang lebih halus (smoother learning signal), mencegah masalah *vanishing gradient* yang sering terjadi pada fungsi *step* (biner) 4.
- **Efisiensi Sumber Daya:** Memungkinkan agen belajar untuk memenuhi SLA "secukupnya" tanpa melakukan *over-provisioning* (alokasi sumber daya berlebihan) yang boros energi, karena agen diberi insentif untuk berada di zona aman namun tetap efisien 3, 4.

Ringkasan Perbedaan

Fitur,Penalti Biner (Hard),Penalti Sigmoid (Soft)
Bentuk Fungsi,Diskrit / Step Function,Kontinu / Kurva 'S'
Respon terhadap Pelanggaran,Hukuman instan & tetap,Hukuman bertahap & proporsional
Sensitivitas,Tidak membedakan near-miss dan far-miss,Sensitif terhadap jarak ke ambang batas
Stabilitas Training,Cenderung tidak stabil (gradien tajam),Lebih stabil (gradien halus) 4
Cocok untuk,SLA Keras yang absolut,Optimasi trade-off (mis: Latency vs Energy)
Rekomendasi untuk Paper Anda: Mengingat Anda menggunakan **PPO** (yang bekerja sangat baik dengan ruang aksi kontinu dan fungsi tujuan yang *smooth*), penggunaan **Penalti Sigmoid** seperti yang disarankan dalam referensi 2 dan 3 akan lebih superior dibandingkan penalti biner. Ini akan membantu agen PPO Anda belajar menjaga latensi Port 4 (Healthcare) tetap rendah secara konsisten tanpa guncangan kebijakan yang ekstrem.