

1 Notation

t	时间索引 (hour h 、day d)
b	区块索引 (block index)
BaseFee_b	区块 b 的 base fee (Gwei)
GasUsed_b	区块 b 的总 gas 用量 (gas)
Burn_b	区块 b 被销毁的 ETH 数量 (ETH)
$Tx_{i,b}$	区块 b 中第 i 笔交易
$\text{gasUsed}_{i,b}$	交易 $Tx_{i,b}$ 消耗的 gas
p_t	ETH 在时间 t 的价格 (USD/ETH)
L2Vol_t	Layer-2 总交易量或活跃度
Activity_t	链上活跃度指标 (Tx count、active address、NFT volume)
M_t	市场情绪指标 (正负向情绪)
S_t	质押率 (staked ETH / circulating ETH)
ε_t	扰动项

2 EIP-1559

单笔交易燃烧量:

$$\text{burn}_{i,b} = \text{BaseFee}_b \cdot \text{gasUsed}_{i,b}. \quad (1)$$

区块级燃烧总量:

$$\text{Burn}_b = \sum_{i \in \text{Tx in } b} \text{burn}_{i,b} = \text{BaseFee}_b \cdot \text{GasUsed}_b. \quad (2)$$

BaseFee 升降规则:

$$\text{BaseFee}_b = \text{BaseFee}_{b-1} \cdot \left(1 + \frac{1}{8} \left(\frac{\text{GasUsed}_{b-1}}{\text{GasTarget}} - 1 \right) \right). \quad (3)$$

3 Labels

$$y_t^{(1)} = \text{Burn}_t, \quad (\text{ETH}) \quad (4)$$

$$y_t^{(2)} = \text{Burn}_t \cdot p_t, \quad (\$) \quad (5)$$

$$y_t^{(3)} = \text{Burn}_t - \text{issuance}_t, \quad (\text{Deflation}) \quad (6)$$

$$y_t^{(4)} = \Delta \text{BaseFee}_t = \text{BaseFee}_t - \text{BaseFee}_{t-1}, \quad (7)$$

$$y_t^{(5)} = \mathbf{1}\{\text{Burn}_t > \gamma\}. \quad (\text{Detection}) \quad (8)$$

4 Input Features

输入向量:

$$X_t = \begin{bmatrix} \text{BaseFee}_{t-1}, \text{GasUsed}_{t-1}, \text{TxCount}_t, \text{ActiveAddr}_t, \text{L2Vol}_t, \\ \text{NFTVol}_t, \text{TVL}_{\text{DeFi},t}, p_t, \Delta p_t, M_t, S_t, \text{weekday}, \text{hour}, \dots \end{bmatrix} \quad (9)$$

区块数据可以把 X_b 设置为滑动窗口

5 深度学习任务

5.1 时序预测

预测未来 h 步:

$$\hat{y}_{t+1:t+h} = \text{Model}(X_{t-w+1:t}). \quad (10)$$

5.2 因果建模

估计干预效应:

$$P(\text{Burn}_t \mid do(\text{L2Vol}_t = l)). \quad (11)$$

6 损失函数

$$L_{\text{reg}} = \text{MSE}(y_t, \hat{y}_t), \quad (12)$$

$$L_{\text{multi}} = \lambda_1 L_{\text{Burn}} + \lambda_2 L_{\text{BaseFee}} + \dots. \quad (13)$$

7 模拟数据生成（数据增强）

$$\text{GasUsed}_t = \alpha_0 + \alpha_1 \text{Activity}_t - \alpha_2 \text{L2Vol}_t \quad (14)$$

$$\text{BaseFee}_t = \text{BaseFee}_{t-1} \left(1 + \frac{1}{8} \left(\frac{\text{GasUsed}_{t-1}}{\text{GasTarget}} - 1 \right) \right) + \eta_t^B, \quad (15)$$

$$\text{Burn}_t = \text{BaseFee}_t \cdot \text{GasUsed}_t + \varepsilon_t. \quad (16)$$

8 Problem setup: environment and decision objective

令时间/块索引为 t (或区块高度 b)，原 EIP-1559 下每个区块的协议计算得到的基准费为 $\widehat{\text{BaseFee}}_t^{\text{proto}}$ 。我们拟训练一个参数化函数 (神经网络)

$$\pi_\theta : s_t \mapsto a_t,$$

使其输出用于替代或调整协议 base fee 的动作 a_t (例如: 直接给出 $\widehat{\text{BaseFee}}_t$, 或给出乘子 m_t 使得 $\widehat{\text{BaseFee}}_t = m_t \cdot \widehat{\text{BaseFee}}_t^{\text{proto}}$)。

系统状态 (观测) 定义为

$$s_t = [\widehat{\text{BaseFee}}_{t-1}^{\text{proto}}, \text{GasUsed}_{t-1}, \text{TxCount}_{t-1}, \text{L2Vol}_{t-1}, p_{t-1}, M_{t-1}, \dots].$$

动作空间:

$$a_t \in \mathcal{A} \subseteq \mathbb{R}_+, \quad \widehat{\text{BaseFee}}_t = \begin{cases} a_t, & (\text{direct}) \\ a_t \cdot \widehat{\text{BaseFee}}_t^{\text{proto}}, & (\text{multiplier}) \end{cases}$$

注意: 为保证安全和协议兼容性, 可把动作限制为区间 $a_t \in [\alpha_{\min}, \alpha_{\max}]$ (例如 $\alpha \in [0.5, 2.0]$)。

9 五个目标的量化（指标定义）

我们定义五个即时目标（即时奖励/代价项），在每个区块 t 基于 $\widehat{\text{BaseFee}}_t$ 与区块内交易行为计算。

9.1 (1) Social welfare (社会福利)

将用户（交易发起者）和验证者（或质押者）视为参与者。令区块 t 内共有交易集合 \mathcal{T}_t ，每笔交易 $i \in \mathcal{T}_t$ 的交易价值（对用户的粗略估计）为 $v_{i,t}$ ，其实际支付给链的费用（以 ETH 计）为

$$\text{fee}_{i,t} = \widehat{\text{BaseFee}}_t \cdot \text{gasUsed}_{i,t} + \text{priorityFee}_{i,t}.$$

用户剩余（surplus）：

$$u_{i,t} = v_{i,t} - \text{fee}_{i,t}.$$

验证者得到的收入（reward）为优先费总和（Tip）：

$$R_t^{\text{val}} = \sum_{i \in \mathcal{T}_t} \text{priorityFee}_{i,t} \cdot \frac{\text{gasUsed}_{i,t}}{\text{gasUsed_total}_t}.$$

定义社会福利（可以含验证者）为：

$$\text{SW}_t = \sum_{i \in \mathcal{T}_t} u_{i,t} + \beta R_t^{\text{val}},$$

其中 $\beta \geq 0$ 控制是否把验证者收入计入社会福利。常选 $\beta = 0$ （仅用户福利）或 $\beta = 1$ （社会整体福利）。

为了把优化目标转为 ** 最大化 ** 的即时奖励，我们令

$$r_t^{(\text{sw})} = \text{SW}_t.$$

9.2 (2) Stability (稳定性)

稳定性衡量 base fee 的波动性（短期内不希望大幅震荡）。我们可以定义即时惩罚为 base fee 的一阶差分平方：

$$r_t^{(\text{stab})} = -\left(\widehat{\text{BaseFee}}_t - \widehat{\text{BaseFee}}_{t-1}\right)^2.$$

或者归一化版本：

$$r_t^{(\text{stab})} = -\left(\frac{\widehat{\text{BaseFee}}_t - \widehat{\text{BaseFee}}_{t-1}}{\widehat{\text{BaseFee}}_{t-1} + \epsilon}\right)^2.$$

9.3 (3) Fairness (公平性)

公平性在此可定义为“不同类型/规模用户承担费用的分布平等性”。若把交易者分组（按 gas 量、钱包类别或交易类型）为 G 组，组 g 在区块 t 支付总费用为 $F_{g,t}$ ，总交易数或权重为 $w_{g,t}$ ，我们用基于 *Gini coefficient* 的指标或逆向的基尼（高值为公平）：

$$\text{Gini}_t = 1 - \frac{\sum_{g=1}^G \sum_{h=1}^G \left| \frac{F_{g,t}}{w_{g,t}} - \frac{F_{h,t}}{w_{h,t}} \right|}{2G \cdot \bar{f}_t},$$

其中 \bar{f}_t 为每组平均费用（做适当归一化）。为方便，定义即时公平奖励为

$$r_t^{(\text{fair})} = \text{Gini}_t \quad (\text{或 } r_t^{(\text{fair})} = -\text{Gini_loss}_t).$$

（注：你也可以直接用费用的标准差或 CV（coefficient of variation）取负作为公平惩罚。）

9.4 (4) Arbitrage (套利 / MEV)

套利量（例如 MEV 提取或跨市场套利）可以用区块内被提取的 MEV 收益 MEV_t 或可执行的无摩擦套利量来量化。我们希望降低可被抢占/套利的机会。即时惩罚：

$$r_t^{(\text{arb})} = -MEV_t.$$

若无法直接度量 MEV，可以用“价格滑点/交易失败率/回滚次数”的代理指标。

9.5 (5) Collusion (合谋 / 骨干节点集中度)

衡量验证者/出块者在手续费或小费分配上的集中度。令 $q_{j,t}$ 为验证者 j 在区块 t 的累计小费份额（或在时间窗口 W 内的份额），用 Herfindahl–Hirschman Index (HHI)：

$$\text{HHI}_t = \sum_j q_{j,t}^2.$$

更大的 HHI 表示更强的集中/合谋风险，故即时奖励可设为负：

$$r_t^{(\text{coll})} = -\text{HHI}_t.$$

10 奖励向量与总体目标

构造即时奖励向量：

$$\mathbf{r}_t = (r_t^{(\text{sw})}, r_t^{(\text{stab})}, r_t^{(\text{fair})}, r_t^{(\text{arb})}, r_t^{(\text{coll})})^\top.$$

最终我们希望学到策略 π_θ ，在长期折扣回报意义上达成多目标最优（或近似 Pareto 最优）：

$$\max_{\theta} \mathbb{E} \left[\sum_{t=0}^T \gamma^t \mathbf{w} \cdot \mathbf{r}_t \right]$$

上式为标量化（scalarization）形式， $\mathbf{w} \in \mathbb{R}_+^5$ 为权重向量（可手工设定或学习）；若不想预先指定权重，可采用 Pareto 优化或多目标策略搜索方法以获得一组权衡解。

11 多目标优化策略（候选方法）

11.1 1. 加权和 (Weighted Sum)

直接使用固定或动态权重 \mathbf{w} ：

$$r_t^{\text{scalar}} = \mathbf{w} \cdot \mathbf{r}_t.$$

然后用任意单目标 RL 算法（PPO, SAC, A2C 等）优化 $\mathbb{E}[\sum \gamma^t r_t^{\text{scalar}}]$ 。

11.2 2. 鞍-Chebyshev 标量化 (Chebyshev / Min-Max)

得到对所有目标均衡优化的解：

$$r_t^{\text{Cheb}} = -\max_k \frac{|G_k(\theta) - z_k|}{\lambda_k},$$

其中 $G_k(\theta)$ 为目标 k 的长期回报估计， z_k 为理想点， λ_k 为标准化参数。

11.3 3. 多策略 / 帕累托搜索

训练多个策略，使用进化/多目标策略优化（例如 NSGA-II、MOEA）或多策略 RL (Pareto PPO) 寻找 Pareto 前沿。

11.4 4. 约束型 RL (Constrained Optimization)

将其中若干目标作为约束（例如把 fairness 和 collusion 作为硬约束），并最大化主要目标（如 social welfare）：

$$\begin{aligned} \max_{\theta} \mathbb{E} & \left[\sum_t \gamma^t r_t^{(\text{sw})} \right] \\ \text{s.t. } \mathbb{E} & \left[\sum_t \gamma^t g_i(s_t, a_t) \right] \leq c_i, \quad i \in \{\text{stab, fair, arb, coll}\}. \end{aligned}$$

这种方法可用 Lagrangian relaxation (LCRL) 或 Constrained PPO 实现。

12 训练与仿真环境设计

为训练 RL 代理，需要一个可交互的环境 \mathcal{E} ，它在给定动作 $\widehat{\text{BaseFee}}_t$ 下，模拟区块内交易行为、用户优先费出价、L2 迁移、MEV 抽取、验证者分配等，输出下一个状态 s_{t+1} 和奖励 \mathbf{r}_t 。环境可由三层组成：

1. **用户下单/策略层**: 按照某个行为模型（带价格敏感性的 Poisson 到达或基于历史数据的生成器）产生交易，包含价值 $v_{i,t}$ 、gas 需求、priorityFee 决策模型（可以是策略性：优先费 = f(等待成本, urgency)）。
2. **L2 与跨链层**: 模拟部分交易转移到 L2，或 L2 批量提交到 L1 (影响 L1 的 gasUsed)。
3. **矿工/验证者层**: 模拟优先费的提取、MEV 搜索者行为以及出块者分配（用于计算 HHI 和 MEV）。

你可以基于简化的 mechanistic model (参见前文给出的生成器方程) 搭建一个仿真器，或用历史链上数据回放并在每个区块替换 base fee 来评估指标 (off-policy evaluation / importance sampling)。

13 算法实现建议

- 若采用标量化 (weighted sum): 用 PPO (或 SAC) 训练策略；损失为标准策略梯度损失 + value loss。
- 若采用约束 RL: 使用 Lagrangian PPO: 在每次迭代更新 Lagrange multiplier λ_i 以满足约束。
- 若追求帕累托前沿: 训练多组权重或用进化方法 (MOEA) 在策略参数空间做搜索。
- 在训练初期用一个较稳健的 baseline (例如 $\widehat{\text{BaseFee}}_t = \text{BaseFee}_t^{\text{proto}}$) 来 warm-start 环境与估计器。

14 评价与基线

- 基线策略：原生协议 $\text{BaseFee}_t^{\text{proto}}$ ，以及简单比例器 $m \cdot \text{BaseFee}_t^{\text{proto}}$ (多项 m)。
- 评价指标：长期折扣回报 $\mathbb{E}[\sum \gamma^t r_t^{(k)}]$ (五个目标分别计算)、Pareto front、稳健性 (在不同用户行为/交易压力下的表现)、off-policy policy evaluation 的经济回测 (PnL/滑点/failed tx rate)。
- 可视化：绘制 Pareto front (SW vs Stability vs Fairness 等)，并给出 HHI/MEV 等合规风险指标随时间变化图。

15 安全与约束注意

将策略推向真实网络前应确保：

1. 策略不会生成极端 base fee (用动作裁剪与 penalty 强制)。
2. 在模拟器上做全面鲁棒性测试 (极端拥堵、L2 爆发、DeFi 大型 swap、闪电贷)。
3. 考虑实现为建议层 (off-chain oracle 提议) 而非直接修改协议——实验结果可用于提案或协议改进建议。