Topology-Aware Learning with Guaranteed Feasibility for ML Unit Commitment

sleeping_pizza_, 10 March 2026

Power Grid as a Dynamic Graph

System under contingency c (line l_c removed):

$$\mathcal{G}^{(c)} = \big(V, E^{(c)}, X(t), W\big), \quad A_{ij}^{(c)} = A_{ij} \cdot \big(1 - \mathbf{1}_{(i,j)=l_c}\big)$$

Generator–bus incidence $M \in \{0,1\}^{N\times G}$: $M_{n,g} = 1$ iff generator g sits at bus n. Used to (i) aggregate generator costs onto nodes and (ii) project bus embeddings back to generator decisions.

Node feature: $x_n(t) = \big[P_n^d(t), \bar{c}_n^p, \bar{P}_n^U, \mathbf{1}_{n\in\Lambda_n^G}\big]^\top$ where $P_n^d = d_n - \sum_j p_j^w$ and

$$\bar{c}_n^p = \frac{\sum_g M_{n,g}\Big(\frac{1}{|\Omega_M|}\sum_m c_{g,m}^p\Big)}{\sum_g M_{n,g} + \epsilon}, \quad \bar{P}_n^U = \sum_g M_{n,g} P_g^U$$

(ϵ prevents division by zero on load-only buses.)

Spatial Encoder: Graph Attention Network

Unnormalised attention at layer k between adjacent buses i, j under contingency c:

$$e_{ij}^{(c)} = \mathrm{act}\big(\mathbf{a}^\top [W_n h_i^{(k)} \| W_n h_j^{(k)} \| W_e w_{ij}]\big)$$

$$\alpha_{ij}^{(c)} = \frac{\exp(e_{ij}^{(c)})}{\sum_{k\in\mathcal{N}^{(c)}(i)} \exp(e_{ik}^{(c)})}$$

$$h_n^{(k+1)} = \sigma\Big(\sum_{j\in\mathcal{N}^{(c)}(n)} \alpha_{nj}^{(c)} W_n^{(k)} h_j^{(k)}\Big)$$

Spatial embedding $Q^{(c)}(t) = \mathrm{GAT}^{(K)}(X(t), A^{(c)}) \in \mathbb{R}^{N\times d_h}$. Same weights, different $A^{(c)}$: each contingency yields a *distinct* embedding.

Temporal Encoder: BiLSTM + Projection

Per-bus BiLSTM: $\tilde{Y}_n(t) = \tanh(V_{\rightarrow}\vec{h}_{n,t} + V_{\leftarrow}\overleftarrow{h}_{n,t} + b_Y)$.

Bus → generator projection (via M) and binary output:

$$\tilde{Y}_g(t) = \sum_n M_{n,g}\tilde{Y}_n(t), \qquad \tilde{z}_g^{(c)}(t) = \mathbf{1}[\tilde{Y}_g(t) \geq 0]$$

Topology-conditioned generalisation: same weights serve every contingency.

Contingency-Augmented Offline Training

$$\Delta_{\text{aug}} = \big\{\big(X(t), A^{(c)}, z^{(c)}(\Delta_s)\big) : s\in\mathcal{S},\ c\in\{0\}\cup\Omega_L^{(c)}\big\}$$
$$\mathcal{L} = \tfrac{1}{TGS(1+|\Omega_L^{(c)}|)} \textstyle\sum_{c,s,t,g}\big(Y_{t,g}^{(c)} - \tilde{Y}_{t,g}^{(c)}\big)^2$$

Security-Constrained MILP-UC (SCUC)

For each $c \in \Omega_L^{(c)}$ (line l_c removed), append to base constraints:

$$\textstyle\sum_{g\in\Lambda_n^G} p_g + \sum_j p_j^w - d_n = f_n^{(c)} \qquad \forall n, t$$
$$p_l^{br(c)} = (\delta_i^{(c)} - \delta_j^{(c)})/X_l, \quad |p_l^{br(c)}| \leq F_l^U \qquad \forall l \neq l_c, t$$

Full system $G^{SC}(x) \leq H^{SC}$ stacks $C+1$ copies; $\mathcal{O}(CNT)$ constraints.

Revised Fuzzy MILP Formulations

SCUC cost bounds: $\underline{C} = \min_{x\in\mathbb{R},\, G^{SC}\leq H^{SC}} C(x)$; $\bar{C} = C(x^{\text{all}})$; $\mu^C = \dfrac{\bar{C} - C(x^F)}{\bar{C} - \underline{C}}$.

Soft N-1 thermal membership (*new*, $\kappa > 0$):

$$\mu_l^{(c)}(t) = \begin{cases} 1 & \rho_l^{(c)} \leq 1 \\ \max\big(0, 1 - \kappa(\rho_l^{(c)} - 1)\big) & \text{else} \end{cases}$$
$$\mu^{SC} = \min_{c,l,t} \mu_l^{(c)}(t)$$

where $\rho_l^{(c)} = |p_l^{br(c)}|/F_l^U$. Setting $\kappa \to \infty$ recovers the hard N-1 constraint.

PM1-SC (individual fuzzy security-constrained model):

$$\max \lambda$$
$$\lambda \leq z_g^F(t) \qquad \forall g, t \in \Upsilon^u \qquad \text{(a)}$$
$$\lambda + z_g^F(t) \leq 1 \qquad \forall g, t \in \Upsilon^d \qquad \text{(b)}$$
$$\lambda \leq K \cdot \mu^C \qquad \text{(c)}$$
$$\lambda \leq \mu^{SC} \qquad \text{(d} \star\text{)}$$
$$G^{(0)}(x^F) \leq H^{(0)}, \quad 0 \leq \lambda \leq 1 \qquad \text{(e)}$$

PM2-SC uses hourly-aggregate ON/OFF memberships: only $2TC+2$ fuzzy constraints vs. $GTC+2$. K-factor $K \in (0, 1]$ tightens the cost satisfaction region in (c);

Feasibility Guarantees

Proposition 1. *If $G^{(0)}(x) \leq H^{(0)}$ is feasible, PM1-SC and PM2-SC are feasible for any $\tilde{z}^{(c)}$ and any $\kappa > 0$.*

Proof. Set $\lambda = 0$: (a)–(b) hold by integrality; (c) by $C(x^F) \leq \bar{C}$; (d) by $\mu^{SC} \geq 0$ always; (e) by the base-case hypothesis. □

New Evaluation Metrics

$\mathrm{NFR} = \frac{1}{|\Omega_L^{(c)}|S}\sum_{c,s}\mathbf{1}[G^{(c)}(x_s^F) \leq H^{(c)}]$ *(fraction of N-1 scenarios satisfied)*

$\text{C-SOI} = \frac{1}{|\Omega_L^{(c)}|}\sum_c \frac{C(x^{F(c)}) - C(x^{*(c)})}{C(x^{*(c)})}$ *(avg. cost gap vs. optimal)*