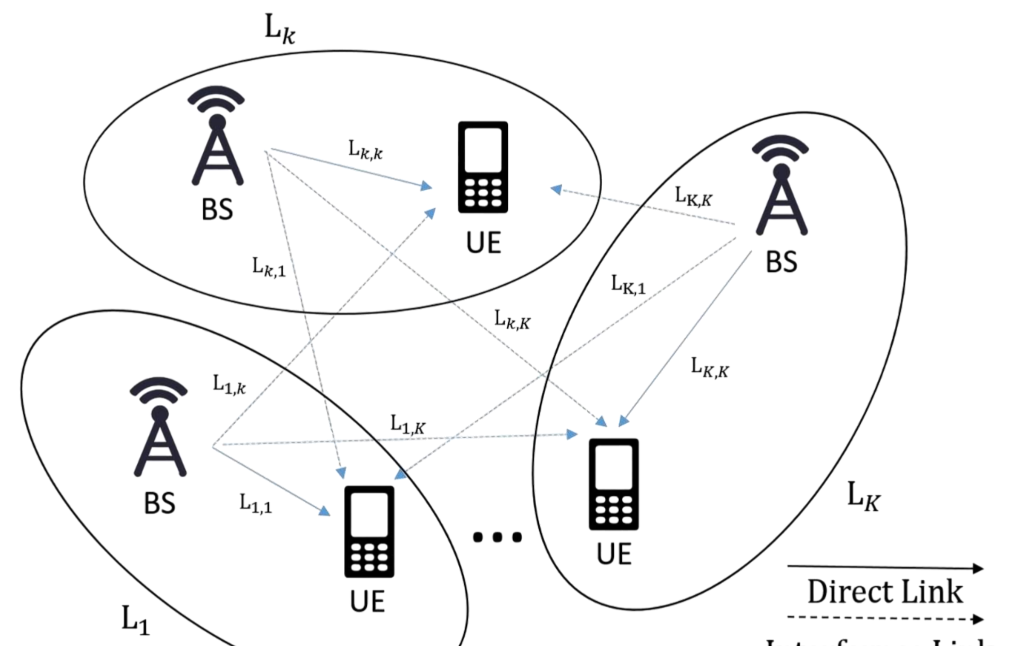
融合GNN與Tarnsformer技術的智能網路優化

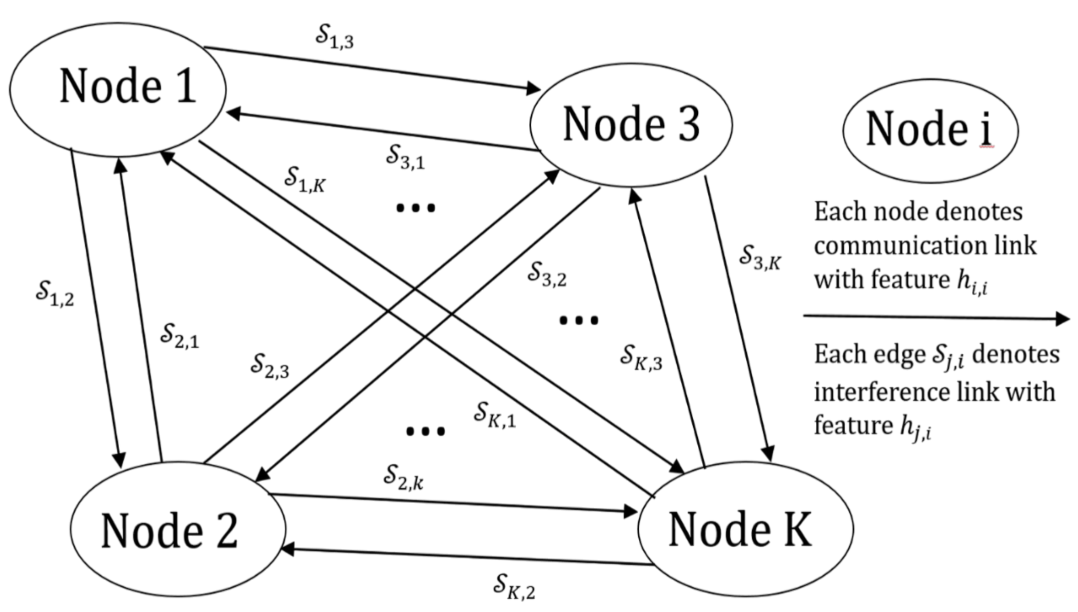
陳世汶; 林庭緯; 劉建興老師

摘要

為滿足大規模超可靠低延遲通訊（mURLLC）對超低延遲、高可靠性與可擴展性的需求，我們提出結合 Transformer 的 G4U 與 PG4U 框架。原有 G4U 透過無線聚合減少訊號與計算開銷，PG4U 則以並行聚合提升穩健性，但兩者皆受限於局部訊息傳遞，難以捕捉長距離依賴關係。為克服此問題，新架構在嵌入更新階段整合 Transformer 編碼器，使模型能全域關注所有節點，實現跨節點依賴建模且無需額外訊號交換。

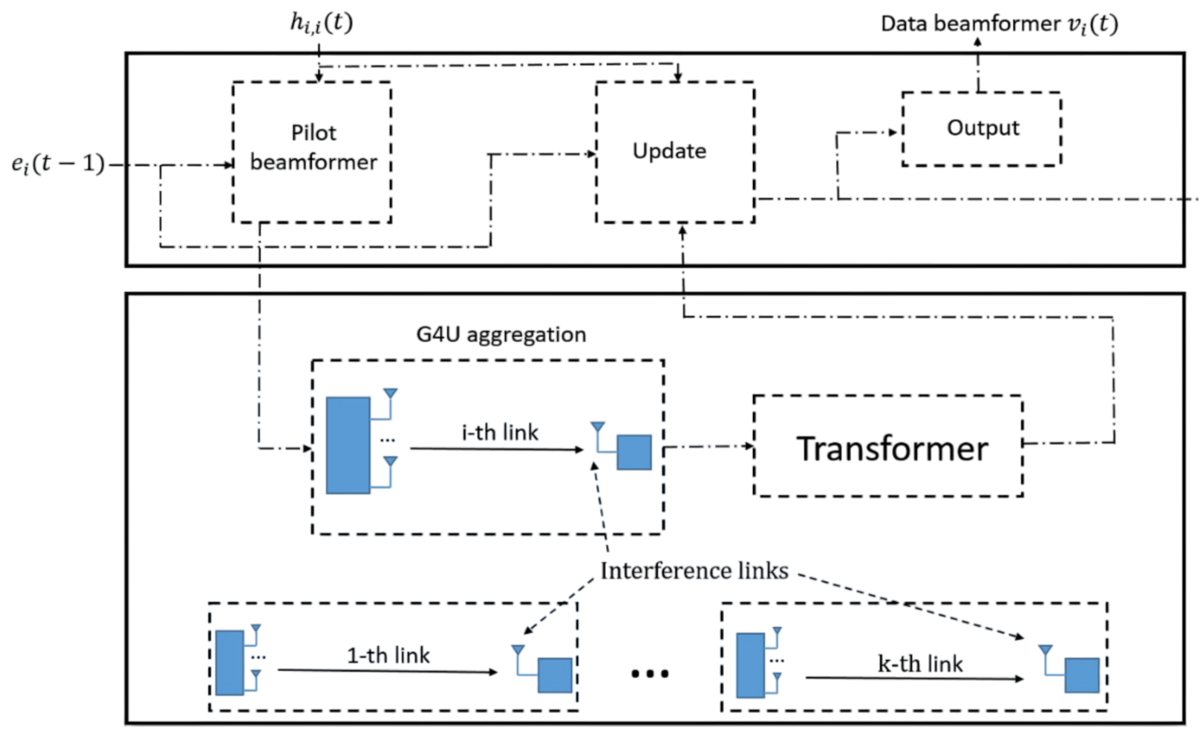
基於 K 連結的所考慮的 mURLLC 網路的系統模型





我們考慮一個由 K 條通訊鏈路 $L_1, L_2, \dots, L_K$ 組成的下行 mURLLC 網路，其中每條鏈路連接一個基地台 (BS) 及其關聯的用戶設備 (UE)，所有鏈路均工作在相同的頻寬上。每個基地台都配備有 N_t 根天線，每個 UE 只有一個天線，所有 BS 同步傳輸。實體網路映射對應圖拓樸結構。

G4U with Transformer 模型圖



訊息的產生、聚合和更新步驟定義如下

$$m_{j\to i}^{(n)}(t) = \Phi\left(h_{j,j}(t), h_{j,i}(t), e_j^{(n)}(t); \theta\right)$$

$$a_i^{(n)}(t) = agg\left(m_{j,i}^{(n)}(t), j \in \mathcal{N}(i)\right)$$

$$e_i^{(n+1)}(t) = U\left(e_i^{(n)}(t), a_i^{(n)}(t), h_{i,i}(t); \omega\right)$$

$$v_i(t) = \Omega\left(e_i^{(N)}(t);\ \varphi\right).$$

連結 L_i 的訊號與干擾加雜訊比 (SINR) 為下

$$\mathrm{SINR}_i(H(t), V(t)) = \frac{\left|h_{i,i}^H(t) v_i(t)\right|^2}{\sum_{j=1, j\neq i}^{K} \left|h_{j,i}^H(t) v_j(t) s_j\right|^2 + \sigma^2}$$

其中 $V(t) = [v_1(t), \dots, v_K(t)]^T \in C^{(K\times N_t)}$ [3]。令 T_o 表示策略推理期間產生的總開銷，包括訊號延遲和計算延遲。

最佳化問題可以表述為:

$$\arg \min_{\mathrm{v_1(t)},\cdots,\mathrm{v_K(t)}} E[U(t)]$$

$$\text{s. t. } \varepsilon_{\max}(t) = \max_{i=1,\cdots,K} \varepsilon_i(V(t), H(t)),$$

$$U(t) \triangleq \log_{10}(\varepsilon_{\max} t + 10^{-\beta}) + \beta,\quad \beta \geq 0,$$

$$0 \leq \|v_i(t)\|^2 \leq P_{\max},\quad \forall i.$$

$$S_q(t) = \sum_{j=1}^{K} \phi(q_j^h(t))\ ;\ S_{kv}(t) = \sum_{j=1}^{K} \phi(k_j^h(t)) \otimes v_j^h(t)$$

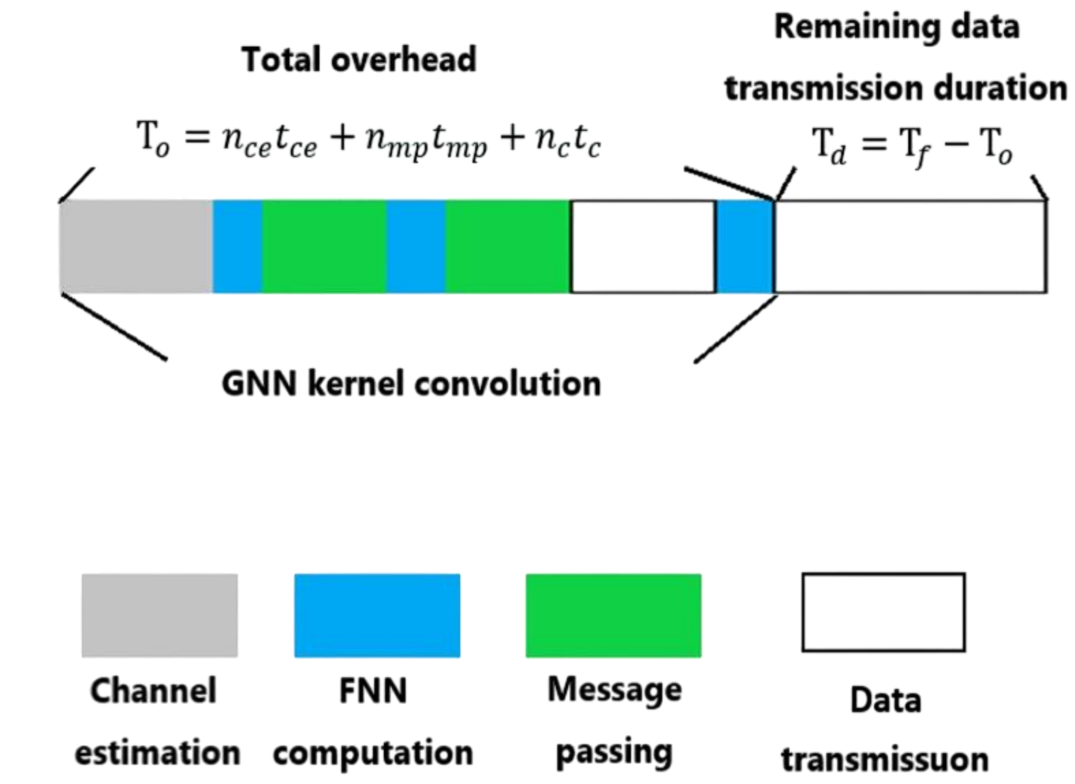
其中 $\phi(\cdot)$ 是 kernel 特徵圖，它將 softmax 注意力線性化，從而允許透過單次 OTA 廣播計算求和。然後每個節點在本地重建其注意力輸出

$$z_i^h(t) = \frac{\phi\left(q_i^h(t)\right)^T S_{kv}(t)}{\phi\left(q_i^h(t)\right)^T S_q(t)}$$

$$\tilde{e_i}(t) = LN\left(x_i(t) + \mathrm{concat}_h\left(z_i^h(t)\right) W_o\right)$$

其中LN(·)表示層歸一化，W_o是輸出投影矩陣，然後應用帶有殘差歸一化的逐位置前饋細化。

在 5G NR 系統中基於 GNN 的策略幀結構設計



$$\varepsilon_i(V(t), H(t)) = Q\left(\frac{-b\ln 2 + B(T_f - T_o)\ln[1 + \mathrm{SINR}_i(t)]}{\sqrt{B(T_f - T_o)V_i(t)}}\right)$$

鏈路 i 的誤塊率(BLER)由 [4]、[5] 給出其中$V_i(t) = 1 - 1/[1 + \xi_i(t)]^{(-2)}$是通道色散[4]，且為高斯 Q 函數。

QoS Outage Probablitiy 計算

在本研究中，實驗評估指標選定為 QoS Outage Probability，其計算方式定義如下

$$\Pr\{\varepsilon_{\max}(t) > 10^{-5}\}$$

參數表

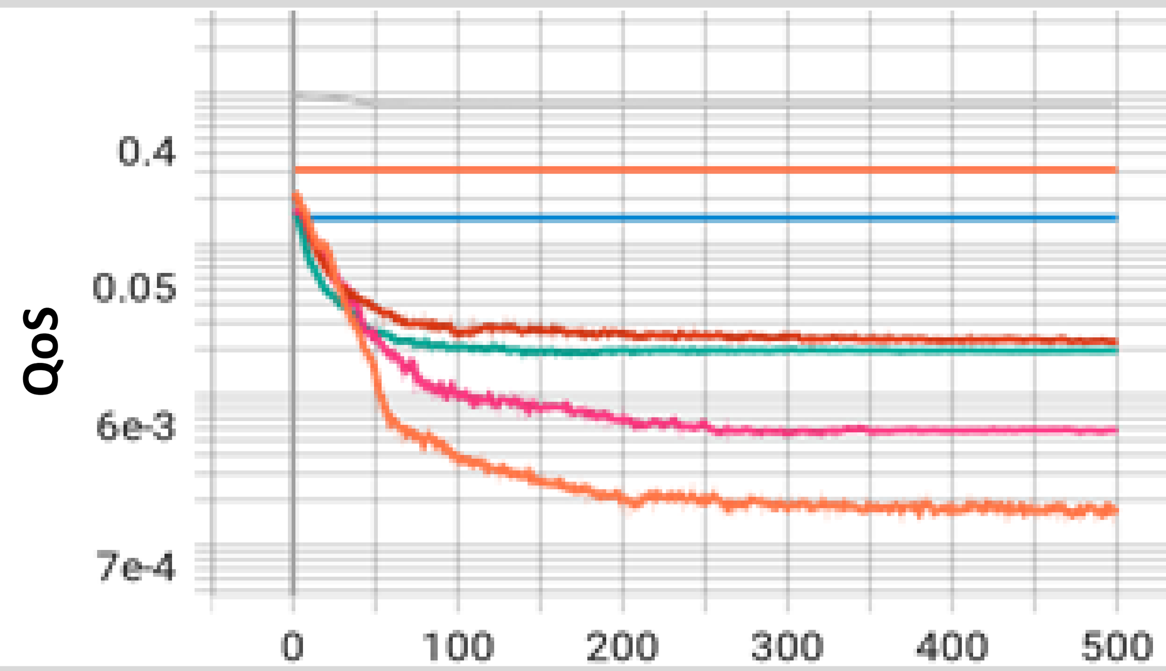
Parameters	Values
Number of antennas N_t	4
Number of frames	10
Number of links	20
Noise power	-174 dBm/Hz
Antenna height	2 m
Maximum transmit power	40 dBm
Bandwidth	5 MHz
Message passing overhead t_{mp}	0.2 µs
Channel correlation coefficient	0.99
Carrier frequency	2.4 GHz
Number of training layouts	2×10^4
Number of testing layouts	5×10^4
Channel estimation overhead t_{ce}	0.2µs
Computation delay t_c	100µs
Number of bits	128
Log-normal shadowing	3 dB

參考論文

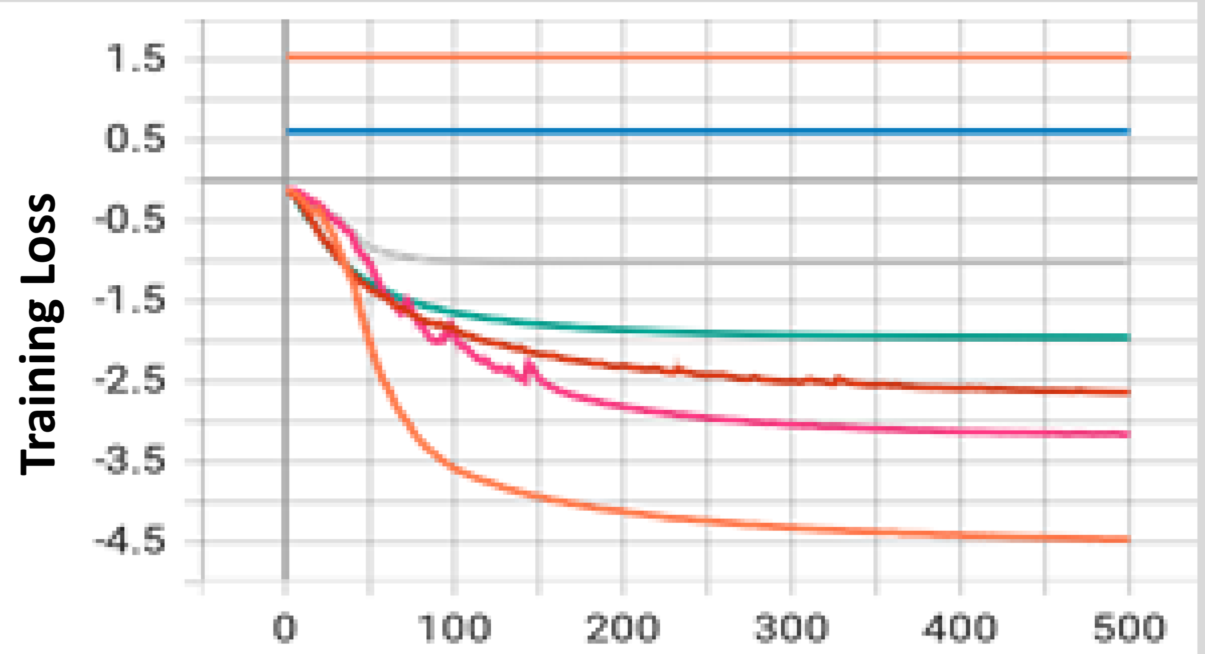
1. Y. Gu, C. She, S. Bi, Z. Quan and B. Vucetic, "Graph Neural Network for Distributed Beamforming and Power Control in Massive URLLC Networks," in IEEE Transactions on Wireless Communications, vol. 23, no. 8, pp. 9099-9112, Aug. 2024.
2. Doshi, S. Yerramalli, L. Ferrari, T. Yoo, and J. G. Andrews, "A deep reinforcement learning framework for contention-based spectrum sharing," IEEE J. Sel. Areas Commun., vol. 39, no. 8, pp. 2526–2540, Aug. 2021.
3. Y. Shen, Y. Shi, J. Zhang, and K. B. Letaief, "Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis," IEEE J. Sel. Areas Commun., vol. 39, no. 1, pp. 101–115, Jan. 2021.
4. W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multipleantenna fading channels at finite blocklength," IEEE Trans. Inf.Theory, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.
5. Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the f inite blocklength regime," IEEE Trans. Inf. Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.
6. Z. Gu et al., "Knowledge-assisted deep reinforcement learning in 5G scheduler design: From theoretical framework to implementation," IEEE J. Sel. Areas Commun., vol. 39, no. 7, pp. 2014–2028, Jul. 2021.
7. X. Liu, C. She, Y. Li, and B. Vucetic, "Edge-wise gated graph neural network for user association in massive URLLC," in Proc. IEEE Globecom Workshops (GC Wkshps), Dec. 2021, pp. 1–6.
8. Y. Gu, C. She, Z. Quan, C. Qiu, and X. Xu, "Graph neural networks for distributed power allocation in wireless networks: Aggregation over-theair," IEEE Trans. Wireless Commun., vol. 22, no. 11, pp. 7551–7564, Nov. 2023.
9. H. Shi, A. Aijaz, and N. Jiang, "Evaluating the performance of over-theair time synchronization for 5G and TSN integration," in Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom), May 2021, pp. 1–6.
10. Propagation Data and Prediction Methods for the Planning of Short Range Outdoor Radiocommunication Systems and Radio Local Area Networks in the Frequency Range 300 MHz to 100 GHz, document I. R. P.1411-11, 2021.

結果

G4U 加 Transformer 能將中斷率降低超過 84%，PG4U 加 Transformer 則降低約 52%。其中，G4U with Transformer 的中斷率最低，甚至優於集中式的 WMMSE，顯示Transformer 的全域注意力能有效提升分散式干擾協調能力。EPA 與 GNN 的中斷率最高，反映其缺乏有效的功率配置機制。



G4U 加入 Transformer 後收斂速度明顯加快，比基線 G4U 快超過 50%，且最終 loss 低約 94%。PG4U 的 Transformer 版本也優於原始 PG4U，但相對改善幅度較小，因其基線本身較低。WMMSE 雖然最終 loss 與 G4U with Transformer 相近，但因為集中式迭代運算，需要更多 epoch 才能收斂；EPA 則維持不變，而分散式 GNN 降幅慢且最終停在較高的 loss。



數線對照表