Mitigating Partial Observability in Adaptive Traffic Signal Control with Transformers

X. Wang[a,*], A. Taitler[b], S. Sanner[a] and B. Abdulhai[a]

[a] University of Toronto, Toronto, Canada
cnxiaoyu.wang@mail.utoronto.ca, baher.abdulhai@utoronto.ca, ssanner@mie.utoronto.ca
[b] Ben-Gurion University of the Negev, Be'er Sheva, Israel
ataitler@gmail.com
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 29, 2024

---

1 INTRODUCTION

Traffic signal control in urban areas plays a pivotal role in managing transportation, reducing congestion, and enhancing safety and sustainability (Grote *et al.*, 2016). While traditional fixed-timing systems struggle to adapt to changing traffic conditions, Adaptive Traffic Signal Control (ATSC) systems use dynamic algorithms to adjust signal timings based on real-time data (El-Tantawy *et al.*, 2013, Wang *et al.*, 2023a), aiming to optimize traffic flow and minimize delays.

Reinforcement Learning (RL) has emerged as a promising approach to enhancing ATSC systems. RL algorithms enable controllers to learn optimal policies by interacting with the environment. One notable RL-based ATSC framework is eMARLIN (Wang *et al.*, 2023b). eMARLIN introduces a lightweight distributive collaboration architecture, allowing a learned communication signal to be transmitted between neighbouring intersections while conserving the communication bandwidth.

However, a significant challenge in ATSC arises from partial observability (PO) in traffic networks. Agents rely on local sensors that can fail to accurately report system states or restricted by their maximum detection range. PO leads to uncertainty and complexity, hindering the control effectiveness. To address this, eMARLIN+ (Wang *et al.*, 2023c) extends eMARLIN by incorporating historical information using a Long Short-Term Memory (LSTM) network. It communicates abstracted local observation histories among agents, aiming to mitigate the effects of partial observability and improve overall system performance. However, LSTM's training throughput poses challenges.

Transformers, known for their success in sequence modelling tasks, including natural language processing (NLP), offer a promising alternative (Vaswani *et al.*, 2017). This paper explores integrating Transformer-based controllers into eMARLIN systems to effectively address partial observability. We propose strategies to enhance training efficiency and effectiveness, presenting results on real-world scenarios and comparing with strong baselines. The Transformer-based model demonstrates improved coordination capability by capturing significant events from historical data for better control policies.

2 TRANSFORMERS PRELIMINARIES

Attention mechanisms in machine learning assign weights to focus on specific parts of inputs, providing powerful feature abstraction capabilities. Transformers (Vaswani *et al.*, 2017) are a significant advancement, demonstrating strong performance across various fields.

The Transformer encoder (Figure 1), a key component, introduces two key innovations: positional encoding and the Transformer block. Positional encoding, typically using sinusoidal functions, provides order awareness to the embedded inputs, allowing Transformers to process sequences in parallel. This method is more scalable and interpretable than traditional recurrent architectures, offering a more efficient solution to long-term memory challenges.

The Transformer block's central innovation is its scaled dot-product attention mechanism, which calculates attention scores between linearly transformed input slices to produce a weighted combination of features. Multiple attention heads enhance understanding by focusing on different aspects of the input simultaneously. This multi-head approach improves scalability and adaptability, making Transformers highly effective for handling complex tasks in various domains.
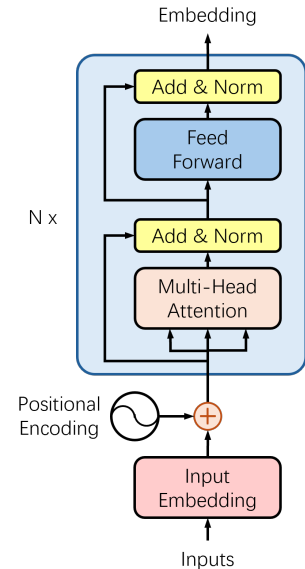

Figure 1 – *The architecture of a Transformer encoder model.*

3 METHODOLOGY

3.1 The Underlying Model

eMARLIN+ (Wang *et al.*, 2023c) is a distributed ATSC coordination algorithm based on the Deep Q-Network (DQNs) (Mnih *et al.*, 2013). It formulates the generic ATSC problem as a Partially Observable MDP (POMDP), where state variables define network dynamics, and observations capture sensor data. To reduce the model complexity, eMARLIN+ factorizes the POMDP by introducing a Dominance Relaxation assumption, prioritizing immediate neighbourhood influences over distant ones under second-level operation. This relaxation factorizes the joint conditional distribution, that represents the global dynamics, and streamlines the model. This relaxation allows eMARLIN+ agents to learn only from neighbouring information, rather than the complex global data. Finally, eMARLIN+ decouples the global reward down to the sum over individual intersection delays, decoupling the joint problem into separate learnable and manageable sub-problems.

In this setting, each of the agents observes locally, receives a local reward, and maximizes this surrogate target directly. The model is designed to be:

- *States:* The state of the system $s \in \mathcal{S}$ contains all the information related to the network, including vehicle positions, velocities, traffic lights status, etc. This information in general is hidden and defines the underlying dynamics of the problem.

- *Observations:* The local observation $o_i \in \mathcal{O}_i$ directly sensed by an agent consists of the count of vehicles and the count of queued vehicles (driving below a threshold speed, 2 m/s used in this work) in each incoming lane within the detection range, as well as the current phase index and its elapsed duration.

- *Actions:* The action $a_i \in \mathcal{A}_i$ represents the *desired phase* in the next time-step. The traffic light EXTENDs the current phase if it is the same as the action, or CHANGEs to the selected phase by going through the corresponding yellow change and red clearance

intervals. A policy constraint permits only those valid transitions under the standard NEMA two-ring-two-barrier scheme (Wang *et al.*, 2023b). Minimum and maximum green time constraints are also applied.

- *Rewards:* The local reward $r_i \in \mathcal{R}_i$ is the agent sensed total queue length at the intersection.

3.2   eMARLIN-Transformer

To address partial observability caused by sensor limitations, we propose eMARLIN-Transformer, an integration of Transformers into the eMARLIN architecture. This extension replaces the fully connected network in the eMARLIN Encoder with a Transformer-based design and augments input with observation history. Unlike eMARLIN+, which uses LSTM and relies on recurrent hidden states, eMARLIN-Transformer leverages attention mechanisms for parallelized computations, improving efficiency and capturing temporal information more effectively. Note that for isolated intersections, eMARLIN-Transformer operates as DQN-Transformer.
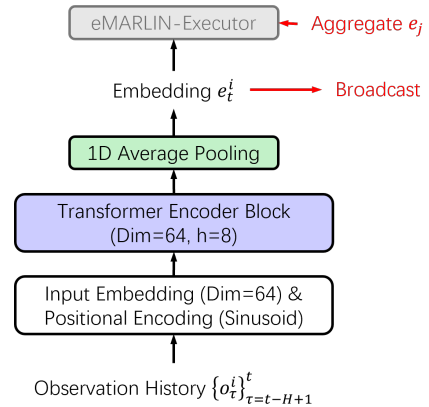

Figure 2 – *The architecture of Encoder in eMARLIN-Transformer.*

3.3   Implementation details

RL and Transformers offer potential in traffic management tasks but face challenges like high memory usage and latency. To stabilize Transformer learning for ATSC, selecting a proper model structure is crucial (Figure 2).

Handling historical information in Transformers requires careful data order and alignment. Our approach involves randomly selecting a recorded episode and a time slice from the replay buffer, retrieving the history up to a certain length. To maintain semantic alignment, we reverse sampled sequences before padding and positional encoding, emphasizing the current time slice.

Padding with zeros can confuse learning agents, as zeros might represent either padding or a zero-valued feature. To address this, we use padding masking, which ensures that attention mechanisms do not misinterpret padded sequences. This masking technique is also applied at the pooling layer after the attention block to ensure clarity throughout the learning process.

4   RESULTS

We modelled a neighbourhood surrounding the intersection of Yonge Street and Steeles Avenue in Toronto, Canada using the Aimsun simulator. City signal timing plans, obtained from Toronto, follow the standard NEMA scheme with semi-actuated control (FHWA, 2007). RL agents run on five intersections along Yonge Street, while the remaining follow city signal timing plans.

The results in Table 1 highlights the effectiveness of eMARLIN-Transformer in coordination performance compared to other methods. Specifically, eMARLIN-Transformer reduces delay time by 65.8% compared to a semi-actuated city plan, indicating a substantial improvement in traffic flow.

When compared to the naive eMARLIN, which does not attempt to address the PO issue, eMARLIN-Transformer achieves a 30.8% improvement in delay time, demonstrating the importance of accounting for sensing limitations and utilizing historical data. Additionally, eMARLIN-Transformer outperforms eMARLIN-LSTM, the previous state-of-the-art approach that uses LSTM to handle sequential data, with 16.6% greater delay time savings, underscoring the advantages of Transformer-based models in addressing PO issues.

Table 1 – *eMARLIN-Transformer outperforms all baselines on the Toronto test-bed. Numbers reported in the table are the episodic delay (s) of each intersection.*

	North2	North1	Yonge-Steeles	South1	South2	Sum
City plan	2,120	77,222	465,500	6,091	49,965	600,898
PressLight	23,590	101,091	276,824	60,180	115,070	576,755
DQN	7,169	61,029	197,544	40,999	50,288	357,029
eMARLIN	**1,027**	60,800	174,464	8,174	48,170	292,635
DQN-LSTM	1,134	65,033	155,048	4,499	38,406	264,120
eMARLIN-LSTM	1,407	51,228	156,408	**3,996**	29,706	242,745
eMARLIN-Transformer	1,416	**40,491**	**136,731**	4,126	**22,747**	**205,511**

These results underscore the Transformer's capacity to capture sequential information from historical observations, enhancing multi-agent collaboration in ATSC, leading to more efficient traffic management and reduced congestion.

5 CONCLUSION AND DISCUSSION

In this work, we presented Transformers as an alternative encoder for capturing sequential information in traffic scenarios, providing advantages over LSTM networks, particularly in handling PO. We integrated Transformers into the eMARLIN architecture to create eMARLIN-Transformer. The implementation details were discussed, highlighting the Transformer's improved performance compared to other RL baselines. Experiments in the North York test-bed demonstrated that the Transformer-based encoder effectively addresses PO issues, enhances agent coordination, and leads to better control performance.

References

El-Tantawy, Samah, Abdulhai, Baher, & Abdelgawad, Hossam. 2013. Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (MARLIN-ATSC): methodology and large-scale application on downtown Toronto. *IEEE Transactions on Intelligent Transportation Systems*, **14**(3), 1140–1150.

FHWA. 2007. *Traffic Analysis Toolbox Volume IV: Guidelines for Applying CORSIM Microsimulation Modeling Software.* U.S. Department of Transportation Federal Highway Administration. Chap. Appendix F: Actuated Signal Control.

Grote, Matt, Williams, Ian, Preston, John, & Kemp, Simon. 2016. Including congestion effects in urban road traffic CO2 emissions modelling: Do Local Government Authorities have the right options? *Transportation Research Part D: Transport and Environment*, **43**, 95–106.

Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, Graves, Alex, Antonoglou, Ioannis, Wierstra, Daan, & Riedmiller, Martin. 2013. Playing Atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz, & Polosukhin, Illia. 2017. Attention is all you need. *Advances in neural information processing systems*, **30**.

Wang, Xiaoyu, Abdulhai, Baher, & Sanner, Scott. 2023a. A critical review of traffic signal control and a novel unified view of reinforcement learning and model predictive control approaches for adaptive traffic signal control. *Handbook on Artificial Intelligence and Transport*, 482–532.

Wang, Xiaoyu, Taitler, Ayal, Smirnov, Ilia, Sanner, Scott, & Abdulhai, Baher. 2023b. eMARLIN: Distributed Coordinated Adaptive Traffic Signal Control with Topology-Embedding Propagation. *Transportation Research Record*, 1–13.

Wang, Xiaoyu, Taitler, Ayal, Smirnov, Ilia, Sanner, Scott, & Abdulhai, Baher. 2023c. eMARLIN+: Overcoming Partial-Observability Caused by Sensor Limitations and Short Detection Ranges in Traffic Signal Control. *In: 2023 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE.