International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 6, Issue 3, 147-159, 2025
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V6I3P120

Original Article

A Safety-Constrained Reinforcement Learning Framework for Scheduling with Latency-Tail Guarantees in Industrial URLLC

Paramesh Sethuraman
Verification Project Manager, Nokia America Corporations, Dallas, TX, USA.

Received On: 06/07/2025 Revised On: 20/07/2025 Accepted On: 14/08/2025 Published On: 05/09/2025

Abstract *- In industrial wireless networks, Ultra-Reliable Low-Latency Communications (URLLC) require strict end-to-end latency requirements with the reliability level of higher than 99.999% especially in time-based control, robotics and safety systems. Although current schedulers used in 5G-Advanced networks are based on minimizing the average latency or maximizing the throughput, they tend to lack a guarantee on strict latency-tails in bursty traffic and in dynamically changing channel scenario. Deterministic scheduling methods are not adaptable to stochastic variations in networks, and reinforcement learning (RL) based schedulers usually seek long-term available levels, but do not have enforceable limits on safety and tail-risk creating the risk of deadline breaches in industrial challenges that are mission-critical. To overcome these shortcomings, a Safety-Constrained Reinforcement Learning (SCRL) framework to latency-sensitive scheduling on Industrial URLLC settings is proposed in this paper. The scheduling is formulated based on a Constrained Markov Decision Process (CMDP) that has clear latency-tail and reliability constraints. An embedded Lyapunov-based virtual queue mechanism is utilized to guarantee the presence of queue stability and hard deadline guarantee, whereas a risk-sensitive competitive objective, which is grounded on Conditional Value-at-Risk (CVaR), is minimized to reduce extreme lateny events. The suggested framework also uses the concept of queue-aware executing state representations in helping to be responsive in burst traffic conditions. Realistic industrial 5G-Advanced simulations show that the given method yields 99.999% reduction in the over standardized queue stability and reliability topology as compared to deterministic earliest deadline first and classic RL schedulers. The scheme delivers close deterministic performance at the expense of flexibility to offer a scalable, and safety guaranteed scheduling system to AI-controlled Radio Access Networks (RAN) in industrial wireless systems.*

Keywords *- 5G-Advanced, Industrial URLLC, Safety-Constrained Reinforcement Learning, Latency-Tail Guarantees, Deterministic Scheduling, AI-Driven RAN, Constrained Optimization, Time-Critical Communications, RAN Intelligence, Risk-Sensitive Optimization, Queue-Aware Scheduling, Lyapunov Optimization, Industrial Wireless Networks.*

1. Introduction

1.1. Industrial URLLC and the Latency-Tail Challenge

The wireless communication used in Industry 4.0 applications (closed-loop motion control, collaborative robotics and safety interlocking) requires ultra-low latencies of milliseconds and extremely high reliabilities (more than 99.999). [1] These rigid prerequisites, which have been formalized under URLLC in 5G and amplified in 5G-Advanced by the 3GPP, move the emphasis of performance to average throughputs to extreme reliability guarantees. Rare latency peaks instead of average delay is the most perilous factor in industrial settings that combine bursty traffic and heavy connectivity, metallic objects, and electromagnetic reverberation. Even a single delay violation of high percentile (e.g. 99.999th percentile) may destroy robotic coordination or jeopardize safety. In turn, the essential functional design objective to industrial wireless wireless scheduling is latency-tail control rather than mean delay optimization.

1.2. Limitations of Existing Scheduling and Learning Approaches

Deterministic-based or Time-based scheduling techniques like EDF exhibit limited-delay behavior with simplistic assumptions, but cannot adapt to stochastic meshwire model wireless phenomenon. [2] On the other hand, schedulers based on utilize reinforcement learning (RL)-based schedulers provide adaptive intelligence (learns traffic and channel variations) but most of the formulations maximize average reward in the long term without specifically handing rare extreme events. This can lead to them attaining low mean latency and yet breaking the ultra-reliable conditions in tail conditions. The literature on latency-tail-sensitive industrial URLLC scheduling has explored stability, constraint-handling, or minimizing of tail individually during constrained Markov decision processes, Lyapunov optimization and risk-sensitive RL, but has not explored combining these approaches altogether. The existence of this gap indicates the necessity to have a framework that imposed hard constraints of reliability collectively but still capable of adaptive optimization.

1.3. Proposed Safety-Constrained RL Contribution

In order to overcome such challenges, in this work, a Safety-Constrained Reinforcement Learning (SCRL) framework is proposed that combines Conditional Value-at-Risk (CVaR)-based tail minimization with virtual queue stabilization with positively constrained Lyapunov-equifications in a constrained Markov decision process model. The framework directly introduces probabilistic deadline constraints in policy optimization, with limited queue growth and stability even in infinite characteristics to stochastic arrivals and fading channels. The proposed solution proves vast improvements in the reduction of 99.999% latency-tail violation against deterministic and conventional RL schedulers just over theoretical guarantees and realistic modeling of an industrial environment within the context of 5G-Advanced URLLC. The SCRL framework helps to provide a common base of reliable AI-based scheduling in the next-generation industrial RANs by providing a linkage between deterministic safety assurance and adaptive intelligence.

2. Related Work

2.1. URLLC Scheduling in the 5G and 5G-advanced

Since 5G Release 15, URLLC has been an important service category and is still under development by 5G-Advanced standardization by the 3GPP. [3] In contrast to eMBB services which focus on the maximization of throughput, the URLLC scheduling focuses on the tight latency (1-10 ms) and ultra-high reliability (≥99.999%). Deterministic scheduling schemes, including the earliest-deadline-first (EDF), fixed-priority, and time-sensitive networking (TSN)-like resource reservation have been well studied to ensure guaranteed delays. The fighter capabilities such as preemptive scheduling, semi-static resource partitioning and grant-free access are also improved to decrease the control-plane latency. Nevertheless, they are highly prone to using Mac modeling or simplistic and dead-beat assumptions, and can be ugly in the presence of bursty industrial traffic state or when channel conditions change rapidly. Tradeoffs between collision probability, spectral efficiency, and reliability in traversing dense factory deployment applications hamper their capability to continuously tame extreme events on latency, and thus like with the latency-tail problem their current capability to avoid such events high.

2.2. AI-Driven and Constrained Reinforcement Learning for RAN

Resource management has also been made adaptive due to the introduction of an artificial intelligence into the RAN and reinforcement learning (RL) has been applied extensively in the scheduling, power control and link adaptation. The model is an RL-based scheduler with the network treated as a Markov decision process thus using the long-term cumulative rewards usually as throughput or expected delay. [4] Actor-critic, deep Q-networks, and policy-gradient approaches show good flexibility in changing the conditions of a channel and traffic dynamics. Nevertheless, the majority of AI-based schedulers are also interested in expected performance as opposed to worst-case or tail guarantees, which makes them unsuitable in safety-critical industry URLLC applications. Constrained reinforcement learning surpasses this weakness by introducing scheduling as a constrained Markov decision process (CMDP) with safety constraints modeled through Lagrangian relaxation or Lieberman auLooking glass Lyapun serves, eccv battery, tau on Labur 56 ACSAC mini electronic pot Blood and fall 246. Although Lyapunov optimization gives verifiable stability to queues, and has limited constraint violations, it is usually short-sighted and does not fully enjoy long-term learning benefits. There is limited literature that exists combining both theoretically stable constrained RL tools and explicit latch-time reduction in wireless scheduling with scarce research on latency-optimal wireless scheduling.

2.3. Latency-Tail Optimization and Remaining Gaps

The relevance of latency-tail optimization in wireless and cloud systems has been popular because some extreme delay events are very infrequent, but when they do occur, the entire service is greatly affected. [5] Conditional Value-at-Risk (CVaR) has been proposed as a consistent risk measure that has the ability to explicitly specify small percentiles, extreme edges, and therefore, is applicable on URLLC reliability goals. Other recent complementary methods like the extreme value theory and large deviation-based probabilistic bound can give analytical information about tail behavior when much of the analytical details rely on stationary traffic or simplified channel models. Notably, the majority of tail-optimization methods are not dependent on adaptive scheduling or reinforcing learning algorithms. Lyapunov-based stability of queues, and modification reinforcement learning in industrial URLLC scheduling. It is imperative that the gap be bridged in order to have formal reliability assurees as well as dynamic dynamism in the next generation intelligent RAN systems.

3. System Model

3.1. Industrial URLLC Network Architecture

3.1.1. Factory Deployment Scenario

We take into account a smart factory implementation via both on a privately operated 5G-Advanced industrial network standardized by the 3rd Generation Partnership Project. [6] The factory floor is composed of robotic arms, programmable logic controllers (PLCs), autonomous guided vehicles (AGVs), safety controllers and distributed sensors installed in an enclosed indoor setting, which has metallic obstructions, multipath fading, and dynamic interference. Time-slotted uplink and downlink scheduling to all devices on shared spectrum resources is offered by a centralized next-generation Node B (gNB).

Time is cut into discrete transmission time intervals (TTIs) numbered by $\mathrm{t} = 0,1,2, \ldots t$ in every time slot, the gNB sees the condition of the whole network, both the queue backlog and channel state information and decides on a scheduling action. The scheduler assigns resource blocks, the modulation and coding schemes (MCS) used as part of it, assigns transmission priorities to ensure that the latency and reliability requirements of the industrial Ultimate-Reliable Low-Latency Communications (URLLC) are achieved.

Scheduling Model and Device Classes
N= 1,2,....N are the set of industrial devices. The system state at time

$$\mathrm{s(t)} = \{\mathrm{Q1(t)}, \dots, \mathrm{QN(t)}, \mathrm{H1(t)}, \dots, \mathrm{HN(t)}\},$$

and $\mathrm{Qi(t)}$ is the queue backlog, and $\mathrm{Hi(t)}$ is the channel state information (CSI) of device i Note that scheduler applies a policy, $\pi\colon \mathrm{s(t)} \to \mathrm{a(t)}$ the action $\mathrm{a(t)}$ with reference to the measured state.

There are three types of industrial devices: (i) safety-critical control devices with latency requirements below 1-5 ms and a reliability of ≥ 99.999% (ii) mission-critical monitoring devices with more relaxed requirements in the form of a small latency and (iii) best-effort sensors with intermediate latency tolerance. Every class has its identifiable deadlines Di and reliability thresholds ei which directly affect the scheduling priorities and the definition of the constraints.

3.2. Traffic Model

3.2.1. Arrival and Burst Modeling

Let $\mathrm{Ai(t)}$ the arrivals of the packet of the device i at time slot t Traffic entries are represented by stochastic processes. [7] Deterministic or quasi-periodic control traffic, periodic control traffic, and event-driven control traffic are characterized by the periodicity occurrence of specific control traffic patterns, periodic control traffic and the Poisson or Markov-modulated Poisson process (MMPP) with mean arrival rate, respectively.

$$\mathrm{E[Ai(t)]} = \lambda\mathrm{i}.$$

A two-state Markov model is developed with normal and burst states in order to model the bursty nature of an industrial environment. The arrival rate in the burst state B rises such that li(B) >li(N). One is that the burstiness tends to impact strongly on the growth of queues and the tail-latency behavior and therefore requires a larger scheduling based on the average.

3.2.2. Queue Dynamics and Stability

A logical queue exists in every device at the gNB. The evolution of queues is controlled by

$$\mathrm{Qi(t+1)} = \max\{\mathrm{Qi(t)} - \mu\mathrm{i(t)}, 0\} + \mathrm{Ai(t)},$$

Where $\mu\mathrm{i(t)}$ refers to the rate of service given by the person in charge of the schedule. Queue stability requires

$$\limsup_{\mathrm{T}\to\infty} \frac{1}{T} + \sum_{t=0}^{T-1} E[Qi(t)] < \infty.$$

Queue backlog has a direct effect on the delay of packets through the Littles law, which defines a good relationship between the stability guarantees and the latency performance. Hence, a narrow queue is one of the conditions under which the URLLC delay requirements can be fulfilled.

3.3. Channel and Reliability Model.

3.3.1. Wireless Channel Model

Small scale fading and shadowing takes place in the wireless channel between every gNB and the devices. We suppose interstationary block fading, [8] such that the channel will be unchanged on a per-TTI basis but different across the slots based on a stationary distribution. $\mathrm{Hi(t)}$Denote the instantaneous channel gain.

The achievable rate is given by

$$\mathrm{Ri(t)} = \mathrm{W}\, log_2 1 + \frac{\mathrm{Pi(t)Hi(t)}}{No}$$

W is bandwidth allocated, $\mathrm{Pi(t)}$ is transmit power and N_0 is noise power. Service rate The service rate $\mu\mathrm{i(t)}$ also varies depending on resources used and the current channel quality.

3.3.2. Reliability and Deadline Constraints

The likelihood of the fact that the latency of the packet L_i is not greater than its deadline Di: is referred to as reliability:

$$\Pr(\mathrm{Li} \le \mathrm{Di}) \ge 1 - \epsilon\mathrm{i},$$

Where the $\epsilon\mathrm{i} \le 10^{-5}$ to ensure that the traffic is safe. Error in transmission is characterized through block error rate (BLER) and this relies on signal-noise ratio (SNR), and the choice of modulation/coding. The mechanisms of hybrid automatic repeat request (HARQ) do have the potential to enhance reliability, but dial up delay variability, hence influencing the latency-tail behavior.

3.4. Latency-Tail Formulation

3.4.1. 99.999% Latency Bound

Industrial URLLC aims at the highest possible extent of reliability, commonly at the 99.999 th percentile of latency:

$$\Pr L_i > {L_i}^{(99.999\%)} \le 10^{-5}$$

The bound in percentile should meet

$${L_i}^{(99.999\%)} \le D_i$$

This representation is an analysis of rare and critical delay events unrepresented by average latency measures.

3.4.2. Tail Probability and CVaR Constraint

The probability of the tail violation is given as.

$${L_i}^{(99.999\%)} \le D_i$$

With constraint $\delta_i \le \epsilon\mathrm{i}$ Conditionally speaking, however, tail risk can still be measured by Conditional Value-at-Risk (CVaR):

$$\mathrm{CVaR}\alpha(\mathrm{Li}), \alpha = 0.99999$$

Minimisation of CVaR minimises the anticipated latency in worst case situations exceeding the percentile point. The following latency-tail definitions are formal safety constraints in the proposed safety-constrained reinforcement learning model, which guarantees flexibility and ultra-reliable operation in future 5G-Advanced RAN 5G-based industrial settings.

4. Problem Formulation

The problem of the Industrial URLLC scheduling is formulated as a time-varying, stochastic constrained optimization problem involving wireless channel and traffic conditions. [9] Since arrivals, fading, and decoding errors change randomly with time, the system is by default represented as a Constrained Markov Decision Process (CMDP) featuring clearly defined safety and reliability constraints. In contrast to traditional delay minimizations issues, such formulation jointly optimizes the expected latency alongside managing extreme delay events and achieving stability of queues.

4.1. Scheduling Objective

4.1.1. Expected Latency Minimization

Let Li(t) be the latency of packets of device iii at time slot t. The mean latency on long-run average basis among devices is denoted as

$$\bar{L} = \limsup_{\mathrm{T}\to\infty} \frac{1}{T} + \sum_{t=0}^{T-1}\sum_{i=1}^{N} \mathrm{E}[\mathrm{Li(t)}]$$

Any reduction in expected queue length is the same thing as reducing the average packet delay since the queue backlog is directly proportional to that delay (as shown in Little's Law).

$$\underset{\pi}{min} = \limsup_{\mathrm{T}\to\infty} \frac{1}{T} + \sum_{t=0}^{T-1}\sum_{i=1}^{N} \mathrm{E}[\mathrm{Qi(t)}]$$

Nevertheless, ensuring the optimal average delay does not avoid very low probability but critical latency spikes, which are inadmissible in safety-critical systems of industry.

4.1.2. Tail Probability Control

In order to clearly control the extreme delay events, determine the latency violation indicator:

$$\mathrm{i(t)} = \begin{matrix} 1, \text{if } \mathrm{Li(t)} > \mathrm{D} \\ 0 \; otherwise. \end{matrix}$$

The probability of long term violation is

$$\mathrm{i(t)} = \begin{matrix} 1, \text{if } \mathrm{Li(t)} > \mathrm{D} \\ 0 \; otherwise. \end{matrix}$$

The URLLC reliability constraint entails.

$$\delta \mathrm{i} \le \epsilon \mathrm{i},$$

where $\epsilon \mathrm{i} \le 10^{-5}$ is to ensure safety-critical devices. Thus, the scheduling goal should collaboratively reduce the anticipated latency as well as impose probabilistic tail guarantees.

4.2. Safety Constraints

The URLLC systems applied in industries introduce numerous [10] stringent safety requirements, in addition to the average delay optimization.

4.2.1. Hard Deadline Constraint

Every packet has to meet a probabilistic deadline constraint:

$$\Pr(\mathrm{Li} > \mathrm{Di}) \le \epsilon \mathrm{i}.$$

This achieves the adherence to the highest reliability requirements (e.g., 99.999 percent), required in motion control automation and automation of safety.

4.2.2. Reliability Constraint

Let $pi^{fail}\,(t)$ represent the probability of packet decoding failure. The long term reliability requirement is:

$$\limsup_{\mathrm{T}\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathrm{E}[pi^{fail}\,(t)] \le \epsilon \mathrm{i}.$$

This will ensure that there are no transmission errors at the physical-layer that are beyond acceptable limits in industrial usage.

4.2.3. Queue Stability Constraint

Stability of queues is necessary to avoid uncontrollable growth of delays. Strong stability requires:

$$\limsup_{\mathrm{T}\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathrm{E}[Qi\;(t)] \le \infty.$$

In the same words, the long term service rate should be more than the rate of arrival:

$$\lambda \mathrm{i} < \bar{\mu}\mathrm{i}.$$

Queue stability is also a safety requirement as unstable queues are sure to cause latency violations.

4.3. Constrained Markov Decision Process (CMDP)

Through the scheduling problem is defined [11] as a CMDP characterized by the following:

$$(\mathrm{S}, \mathrm{A}, \mathrm{P}, \mathrm{r}, \{\mathrm{ck}\}).$$

4.3.1. State Space

The system state at time T is

$$\mathrm{s(t)} = \{\mathrm{Qi(t)}, \mathrm{Hi(t)}, \mathrm{Ai(t)}\}_{i}^{n}{=}1$$

Where the congestion is represented by queue backlogs, variability of channels is modeled by wireless and variability in arrival is modeled by traffic. State space combines network and physical layer knowledge.

4.3.2. Constraint Functions

There are three constraint cost functions which are:

Latency violation cost:

$$\mathrm{c1(t)} = \sum_{i=1}^{n} I_i\,(t)$$

Reliability cost:

$$\mathrm{c2(t)} = \sum_{i=1}^{n} p_i\, fail\;(t)$$

Stability-related cost:

$$\mathrm{c3(t)} = \sum_{i=1}^{n} Q_i\;(t)$$

The CMDP objective is:

$$\max_{\pi} E\pi \sum_{t=0}^{\infty} \gamma \mathrm{tr}(\mathrm{s(t), a(t)})$$

Subject to

$$\mathrm{E\pi[ck] \leq dk, \forall k.}$$

4.4. Risk-Sensitive Formulation

In order to explicitly penalize the extreme delay [12] events we introduce the risk-sensitive objectives.

4.4.1. CVaR-Based Objective

Let L denote the latency random variable. The Conditional Value-at-Risk (CVaR) at confidence level α is defined as:

$$\mathrm{CVaR\alpha(L) = E[L \mid L \geq VaR\alpha(L)].}$$

In the case of industrial URLLC, α =0.99999 alpha = 0.99999 α =0.99999. The optimization problem turns to be:

$$\min_{\pi} \mathrm{CVaR\alpha(L).}$$

CVaR crime-handicaps far latency values that is higher than either the 99.999th percentile, so that there is tail-conscious scheduling.

4.4.2. Exponential Utility Development.

Another method applies an exponential utility functional:

$$\min_{\pi} \mathrm{CVaR\alpha(L).}$$

Where $\theta > 0$ controls risk sensitivity. Larger θ values impose stronger penalties on extreme delays. The objective becomes:

$$\max_{\pi} \mathrm{E[U(L)].}$$

4.4.3. Worst-Case Robust Formulation

A strong formulation works to ensure that worst-case expected latency is minimized on top of an uncertainty set *P* to take into account uncertainty in both traffic and channel statistics.

$$\min_{\pi} \; \max_{\pi} \mathrm{EP[L].}$$

4.4.4. United Optimization Problem

The final scheduling problem is formulated as:

$$\min_{\pi} \mathrm{CVaR\alpha(L)}$$

subject to:

- $\mathrm{Pr(Li > Di) \leq \epsilon i}$
- Reliability constraints
- The limitation of queue stability

This risk-aware CMDP is where the Safety-Constrained Reinforcement Learning framework can be found and its mathematical foundation where adaptive time amounts can be achieved coupled with formal industrial-grade reliability and latency guarantees.

5. Proposed Safety-Constrained Reinforcement Learning Framework

This section presents the suggested Safety-Constrained Reinforcement Learning (SCRL) framework, which is based on the optimization of risk-sensitive problems combined with Lypov-constrained enforcement that can be used to guarantee latency-tail guarantees in Industrial URLLC networks. [13] The architecture is programmed to be implemented in real time at the gNB scheduler and at the same time performance in terms of delay is optimized and the safety and reliability requirements are enforced.

5.1. Framework Overview

5.1.1. Architecture Description

The SCRL model has four fundamental modules that run within the gNB scheduler:

- State Observer Module - Gathers queue backlog, channel state information (CSI), arrivals at the traffic, and deadline status of all the industrial devices.
- Risk-Sensitive Policy Network - The model is a deep actor-critic that compiles the scheduling actions based on the queue-aware state and channel-aware state.
- Lyapunov Safety Layer - Enforcement of stability by ensuring drift by maintaining queues of virtual measuring the constraint violation.
- Dual Update Module - Reliability and deadline multipliers lagrange multipliers are updated.

The working process is as follows.

5.1.2. Network State → SCRL Agent → Action → Environment Update

The SCRL agent will combine three complementary mechanisms:

- Adaptive scheduling: Actor-Critic learning.
- Objective shaping in latency control using CVaR.
- Reduction of Lyapunov drifts of constraint satisfaction.

Such integrated design allows the performance and safety assurances to be combined.

5.1.3. Online Learning Process

At each time slot t

- The state of the system $\mathrm{s(t)}$ is monitored by the agent.
- Scheduling decision $\mathrm{a(t)}$is the output of the actor network.
- The environment care-takes backlogs of queues and states of channels.
- The reward costs and constraint costs are calculated.
- Dual variables and virtual queues are modified.
- The update of these policy parameters is through the use of gradient descent.

All learning process is online and it makes it adaptable to bursty traffic, channel fading as well as other needs that are heterogeneous to various devices. In contrast to offline optimization, SCRL operates constantly to optimize the scheduling policy as real-time dynamics in the industry.

5.1.4. RAN Integration

SCRL scheduler is intended to be used in AI-based RAN architectures like O-RAN-defined near-real-time RAN Intelligent Controller. Scheduler is at the MAC layer of the gNB and is connected to:

- CSI feedback modules
- Buffer management systems
- link adaptation controllers and HARQ

Inference latency is capped to URLLC time-slot constraints (≤1 ms), which makes the use of inference real time in industrial 5G-Advanced networks.

5.2. Lyapunov-Based Constraint Enforcement

In a bid to ensure safety in the process of learning, [14] SCRL integrates the principles of Lyapunov optimization that implement probabilistic constraints through virtual queues.

5.2.1. Virtual Queue Construction

To each constraint k, a virtual queue is defined Zk(t)

$$\mathrm{Zk(t+1) = \max\{Zk(t) + ck(t) - dk, 0\}.}$$

For latency violation probability of device i

$$\mathrm{Zlat, i(t+1) = \max\{Zlat, i(t) + Ii(t) - \epsilon i, 0\}.}$$

When the violation rate is above the permitted limit, the virtual queue is extended, and it adds a punishment in subsequent scheduling decision making.

5.2.2. Lyapunov Function and Drift Minimization

Define the quadratic Lyapunov function:

$$\mathrm{L(t)} = \frac{1}{2}\sum_{k} z_{k}^{2} t$$

The Lyapunov drift is:

$$\mathrm{\Delta(t) = E[L(t+1) - L(t) \mid s(t)].}$$

The SCRL reduces the expression of drift-penalty:

$$\mathrm{\Delta(t) + VE[CVaR\alpha(L(t))],}$$

Where V balances performance and stability. Larger V emphasizes performance optimization, while smaller V prioritizes constraint tightness.

5.2.3. Stability Guarantee

If the drift satisfies:

$$\Delta(\mathrm{t}) \leq \mathrm{B} - \sum_{k} z\,k(\mathrm{t})\epsilon$$

Assuming that B is some finite constant, then:

- The virtual queues are highly stable.
- There are constraint violations that are limited.
- Violation probability in the long-run is satisfied.

$$\limsup_{\mathrm{T}\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathrm{E}[ck\ (t)] \leq \mathrm{dk}.$$

This is to give formal guarantees on deadline compliance, reliability control and queue stability.

5.3. Queue-Aware State Representation

State representation is therefore [15] vital in scheduling decisions that are tail aware.

5.3.1. Backlog-Aware Encoding

Backlog of the queues is normalized:

$$\overline{Q\imath}(t) = \frac{\mathrm{Qi(t)}}{Q\ max}$$

The urgency of the deadline is coded as remaining slack time:

$$\mathrm{D}_{i}^{rem}\mathrm{(t) = Di - Li(t).}$$

Such a representation enables the policy to give priority to near deadline packets based on their occurrence.

5.3.2. Channel–Traffic Feature Fusion

The state vector combines:

$$\mathrm{s(t)} = \left[\mathrm{\overline{Q}i(t), Hi(t), Ai(t), D}_{i}^{rem}\mathrm{(t)}\right]\mathrm{i} = 1_{i}^{n} = 1$$

Deep neural layers combine congestion and channel information and hence the prioritization can be adapted to change based on the fading and burst traffic. The result of this combination is much more robust than a queue-based or channel-only scheduler.

5.4. Policy Optimization Algorithm

The SCRL uses the risk-sensitive actor-critic algorithm without Lagrangean dual update.

5.4.1. Lagrangian Objective

The Lagrangian function is:

$$\mathrm{L(\theta,\lambda) = E\pi[r(t)]} - \sum_{k} \mathrm{\lambda k(E[ck(t)] - dk)}$$

The parameters of the actors are updated through policy gradient:

$$\mathrm{\nabla\theta J = E\pi[\nabla\theta log\pi\theta(a \mid s)A\pi(s,a)].}$$

To minimize the variance to enhance convergence, the critic approximates the value function Vp(s).

5.4.2. Dual Variable Updates

Projection gradient ascent is used to update Lagrande multipliers:

$$\mathrm{\lambda k(t+1) = [\lambda k(t) + \eta(ck(t) - dk^{+}]}$$

Where η is the dual learning rate. Violations increase penalty weight, dynamically reinforcing constraint satisfaction.

5.4.3. Safety-Layer Projection

The feasibility of actions is determined prior to the actual execution of the action of choice by use of a safety projection step:

$$a_{safe}\mathrm{(t) = arg} \underset{a\varepsilon Asafe}{\mathrm{min}} \parallel \mathrm{a - a(t)} \parallel^{2}$$

The projection eliminates risky schedule allocations that may lead to immediate deadline violation or overflow.

6. Theoretical Analysis

This part forms the theoretical basis of the proposed Safety-Constrained Reinforcement Learning (SCRL) framework. We establish stability of queues, [16] deduce the latencies hydraulic changes of tails, as well as show convergence under constrained optimization of Lyapunov. The analysis shows that the proposed scheduler is reliable to the objective of the industrial URLLC factors and ensures stability of learning.

6.1. Queue Stability Proof

6.1.1. Queueing System Model

Suppose a wireless system at discrete times with each device i with a queue which changes as:

$$\mathrm{Qi}(t+1) = \max[\mathrm{Qi}(t) - \mu\mathrm{i}(t), 0] + \mathrm{Ai}(t),$$

$\mathrm{Qi}(t)$ represents backlog, $\mathrm{Ai}(t)$ represents arrivals, and $\mu\mathrm{i}(t)$, is the service rate chosen by the RL scheduler. Mean $\lambda\mathrm{i} = \mathrm{E}[\mathrm{Ai}(t)]$, and the service decisions are based on perceived system state.

6.1.2. Lyapunov Function and Drift.

Define a quadratic Lyapunov function:

$$\mathrm{L}\big(\mathrm{Q}(t)\big) = \frac{1}{2}\sum_{i} \mathrm{Q}_{2}^{i}(t)$$

The conditional Lyapunov drift is:

$$\Delta\mathrm{L}(t) = \mathrm{E}[\mathrm{L}(\mathrm{Q}(t+1)) - \mathrm{L}(\mathrm{Q}(t)) \mid \mathrm{Q}(t)].$$

Using standard drift expansion techniques introduced by Michael J. Neely, the drift can be upper bounded as

$$\Delta\mathrm{L}(t) \leq \mathrm{B} + \sum_{i} \mathrm{Qi}(t)\mathrm{E}[\mathrm{Ai}(t) - \mu\mathrm{i}(t) \mid \mathrm{Q}(t)],$$

B is a finite constant, which relies on limited arrival and service rates.

6.1.3. Stability Condition

If the scheduling policy ensures:

$$\sum_{i} \mathrm{Qi}(t)\mathrm{E}[\mu\mathrm{i}(t)] \geq \sum_{i} \mathrm{Qi}(t)\lambda\mathrm{i} + \epsilon\mathrm{i}\sum_{i} \mathrm{Qi}(t),$$

construe that $\epsilon > 0$; then the drift is negative outside some finite range. Consequently:

$$\limsup_{\mathrm{T}\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{i} \mathrm{E}[\mathrm{Qi}(t)] < \infty$$

Therefore, all queues that are highly stable.

6.1.4. Stability Implications

The Lyapunov-based design guarantees:

- Mean-rate stability
- Limited queue length of long-term average.
- Optimality in the network capacity region throughput.

Thus, SCRL planner stabilizes the network when subjected to admissible traffic loads.

6.2. Latency-Tail Bound Derivation.

The URLLC systems need probabilistic guarantee of delay and not just averageness. [17] Bound limits on violation of delays are now obtained.

6.2.1. Delay-Queue Relationship

By Little's Law:

$$Di = \frac{\mathrm{Qi}}{\lambda\mathrm{i}}$$

Therefore, the limits of queue length directly limit packet delay.

6.2.2. Tail Bound via Markov Inequality

For any threshold $\mathrm{d} > 0$:

$$\Pr(\mathrm{Di} > \mathrm{d}) = \Pr(\mathrm{Qi} > \lambda\mathrm{id}).$$

Using Markov inequality:

$$\Pr(\mathrm{Qi} > \mathrm{x}) \leq \frac{\mathrm{E}[\mathrm{Qi}]}{x}$$

Therefore:

$$\mathrm{r}(\mathrm{Di} > \mathrm{d}) \leq \frac{\mathrm{E}[\mathrm{Qi}]}{\lambda\mathrm{id}}$$

This indicates that the probability of violation decreases with the decrease of expected queue backlog.

6.2.3. Exponential Tail Bound

In light-tailed arrivals and restricted service rates, the large deviation theory provides:

$$\Pr(\mathrm{Qi} > \mathrm{x}) \leq \mathrm{C}e^{-\theta\mathrm{x}}$$

Where $\theta > 0$ is a question of system slack ϵ.
Thus:

$$\Pr(\mathrm{Di} > \mathrm{d}) \leq \mathrm{C}e^{-\theta\lambda\mathrm{id}}$$

6.2.4. Interpretation

The SCRL framework ensures:

- Delay distribution that decays sub-exponentially.
- Very low probability of violation.
- Tunable reliability through Lyapunov control V.

This enables the URLLC reliability requirements (e.g., 10^{-5} probability of violation) of 5G-Advanced industrial networks.

6.3. Convergence Analysis

We study convergence of the policy [18] and the ultimate constraint satisfaction.

6.3.1. Primal-Dual Formulation

The constrained MDP is expressed as follows:

$$\max_{\pi} \mathrm{J}(\pi) \text{ subject to } \mathrm{Ck}(\pi) \leq \mathrm{ck}$$

With Lagrangian relaxation:

$$\mathrm{L}(\pi,\lambda) = \mathrm{J}(\pi) - \sum_{k} \lambda\mathrm{k}(\mathrm{Ck}(\pi) - \mathrm{ck})$$

The SCRL algorithm performs stochastic gradient ascent on π and gradient ascent on λ.

6.3.2. Dual Updates
Dual variables evolve as:

$$\lambda\mathrm{k}(\mathrm{t}+1) = [\lambda\mathrm{k}(\mathrm{t}) + \eta\mathrm{t}(\mathrm{Ck}(\mathrm{t}) - \mathrm{ck})]^{+},$$

With diminishing step size $\eta\mathrm{t} = 1/\mathrm{t}$.

6.3.3. Policy Convergence
If:

- There is smooth parameterization of policy.
- Gradient estimations are not biased.
- Robbins-Monro conditions are met by step sizes.

Then:

$$\nabla\theta\mathrm{L}(\theta,\lambda) \to 0.$$

In this way the algorithm will reach a local saddle point of the Lagrangian.

6.3.4. Constraint Satisfaction
Under Slater's condition:

- Strong duality holds
- The instances of constraint violation are eliminated asymptotically.

$$\limsup_{\mathrm{T}\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathrm{Ck}(\mathrm{t}) \leq \mathrm{ck}.$$

6.3.5. Safety-Layer Projection Stability
The safety projection operator:

$$a_t' = \Pi_{\mathrm{Asafe}}(\mathrm{at})$$

Is non-expansive:

$$a_t' - \mathrm{a}*\| \leq \| \mathrm{at} - \mathrm{a}*\|$$

Therefore, it does not destabilize gradient updates and preserves convergence guarantees.

6.4. Theoretical Guarantees Generally.
The SCRL model offers the below provable properties:

Table1: Summary of Theoretical Guarantees of the Proposed SCRL Framework

Property	Guarantee
Queue Stability	Strong stability via negative Lyapunov drift
Delay Bound	Exponential latency-tail decay
Constraint Satisfaction	Asymptotic feasibility
Policy Convergence	Convergence to local saddle point
Safety	Hard constraint enforcement through projection

6.4.1. Theoretical Significance
By integrating:

- Optimization of the Lyapunov drift.
- Objective shaping risk-sensitively.
- Primal-dual reinforcement learning.
- Safety-layer projection

The presented SCRL scheme attains a rigorously stable, latency-conscious, and reliability ensured scheduling scheme to be applied to next-generation 5G-Advanced and future 6G intelligent RAN systems.

7. Simulation Setup

In this section, the simulation environment, the deployment assumptions, the comparisons of baselines, and evaluation statistics applied to confirm the proposed Safety-Constrained Reinforcement Learning (SCRL) framework to the URLLC conscious scheduling in industrial Radio Access Network (RAN) settings will be described. [19] The installation complies with the requirements of URLLC established by 3rd set of the International Telecommunication Union 3 rd generation Partnership Project and a standard of indoor propagation of industrial propagation. This is aimed at strictly testing stability, latency-tail reliability and learning performance in the realistic 5G-Advanced industrial conditions.

7.1. Industrial Deployment Scenario

7.1.1. Factory Layout

The simulation model is a smart plant that is modelled based on 5G industrial automation use cases that were described in 3GPP URLLC specifications. The factory floor is 120m x 80m and it is a medium-scale production facility. One centralized next generation NodeB (gNB) is attached at the centre of the ceiling to imitate the indoor macro-cell coverage. Several industrial devices over the floor (30-80) are robotic arms, autonomous guided vehicles (AGVs), and wireless sensors. This environment consists of both the mobile and fixed nodes in order to capture the heterogeneous mobility patterns. It is mainly line-of-sight (LoS)-based propagation, and the stochastic blockage process is by the presence of machinery and moving vehicles.

The parameters of deploying are as follows:

- Carrier frequency: 3.5 GHz
- System bandwidth: 40 MHz
- Transmission Time Interval (TTI): 1ms.
- Granularity of scheduling: Slot scheduling.
- MIMO configuration: 4x4

The channel model applies the industrial indoor path-loss formulae used with the ITU recommendations, which assumes the shadowing and small-scale fading effects. This guarantees the existence of realistic signal-to-interference-plus-noise ratios (SINR) dynamics which are well suited to the evaluation of URLLC.

7.1.2. Traffic Parameters

The simulated RAN allows the use of heterogeneous classes of traffic to [20] correspond with the realistic patterns of industrial communication:

- URLLC Control Traffic
- Periodic Sensor Traffic
- Background Best-Effort Traffic

URLLC Traffic:

URLLC packets vary in size between 32-128 bytes which corresponds to control-command payloads of appetite common in robotic coordination. Arrivals are stress-tested by taking either a Poisson or a bursty Markov-modulated process. Its strict target latency is ≤ 1 ms and the reliability of the strict target is 99.999% (five-nines reliability).

Sensor Traffic:

Sensor messages are periodically produced with improvements between 5- 20 ms intervals, which are telemetry streams with moderate reliability requirements.

Best-Effort Traffic:

This class is represented as the FTP-like flows, throughput-driven, and has no hard latency constraint. It guarantees resource competing and mixed-load realism.

7.1.3. Queue Structure

Each of the devices has two logical queues:

- A physical queue (store holding the packets in waiting) of packets pending dispatch.
- An enforcing probabilistic constraints using Lyapunov optimization: A virtual latency-violation queue.

The state of the global system is monitored by the SCRL agent, which includes:

- One-second queue backlogs.
- Channel State Information (CSI)
- Latency violations in the course of history.
- Interference measurements

The state representation allows the agent to acquire scheduling policies that trade off maximum throughput and rigid reliability.

7.2. Baseline Methods

To authenticate the benefits of SCRL the framework is contrasted to three standard scheduling bases that encapsulate deterministic, [21] learning-based, and classical stochastic optimization strategies.

7.2.1. Deterministic EDF Scheduler

Earliest Deadline First (EDF) scheduler allocate resource to packets the resources with least remaining deadline. It is a greedy, deadline priority policy that is channel unaware and unlearnable. EDF is not complex and deterministic hence it is used a great deal in real-time systems. But in the case of bursty arrivals or heavy load it experiences deadline clustering and high probability of violation. It fails to guarantee probabilistically and it is not adaptive.

7.2.2. Traditional RL Scheduler

In this baseline, the policy-gradient scheduler is unconstrained with a Proximal Policy Optimization (PPO). The algorithm optimizes a weighted reward which is a combination of throughput and delay not necessarily with any reliability constraints enforced. Although learning enhances channel-conscious semitimed scheduling choices, safety constraints disappear resulting in erratic tail-latency characteristics with heavy traffic density. Here, regardless of whether learning is enforced or not, implementation is restrained is the question here.

7.2.3. Lyapunov-Only Scheduler

This is the scheduler that works by Max-Weight that are at stochastic network optimization theory by Michael J. Neely. The rule of scheduling picks user.

$$\mathrm{Qi(t)\mu i(t)}$$

This approach ensures that the network has stability in terms of queue and throughput optimality in the network capacity region. Nonetheless, it does myopic per-slot optimization and does not do long-term learning or predictive adaptation to channel statistics.

Table 2: Comparative Analysis of Scheduling Methods

Scheduler	Learning	Constraint Enforcement	Channel Awareness
EDF	No	No	No
Traditional RL	Yes	No	Yes
Lyapunov-only	No	Implicit	Yes
Proposed SCRL	Yes	Explicit (Lyapunov + Dual)	Yes

7.3. Evaluation Metrics

Performance analysis revolves around all ultra-reliable, [22] low-latency measures that are consistent with URLLC standards.

7.3.1. Five-Nines Latency (99.999% Delay)

Having the value of $D99.999\%$ this measure is used to indicate the delay limit under which 99.999% of packets have been sent. This measure not only captures extreme tail behavior unlike the average delay, but also like the extreme delay is essential to industrial automation.

7.3.2. Tail Violation Probability

$$P\mathrm{viol} = \mathrm{Pr(D > Dmax)},$$

Where $D\mathrm{max} = 1\ \mathrm{ms}$

The target constraint is:

$$P\mathrm{viol} \leq 10^{5}$$

This appraises itself directly in terms of compliance with URLLC requirements of reliability.

7.3.3. Reliability

$$\text{Reliability} = 1 - P_{\text{packet\_loss}}$$

There are packet losses such as deadline expiry, buffer overflow and channel decoding errors. This is a comprehensive assessment of end to end integrity of communication.

7.3.4. Throughput

Throughput is calculated as those bits that are delivered successfully divided by the total time. Per-class and aggregate throughput are reported, thereby making it fair in the heterogeneous traffic categories.

7.3.5. Queue Stability

The aspect of queue stability is obtained by time-average backlog and maximum observed backlog as the load increases. Stability is confirmed if:

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathrm{Qi(t)} < \infty$$

This confirms theoretical assurances obtained.

7.3.6. Learning Convergence Metrics

Other measures in reference to RL-based schedulers are:

- Reward convergence curves
- Dual variable stability
- Ensuring the constraint breaking is not obsolete.

These indicators are efficient in the optimality of performance and satisfactory of the constraints at training.

7.3.7. Simulation Duration and Reproducibility

The duration of each experiment is 106 time slots to enable steady-state behavior as well as such a latent occurrence. Data are averaged over 10 random seeds that are independent so that 95% confidence intervals are also given. Tuning of hyper parameters is done through grid search to make a fair comparison of baselines.

8. Results and Discussion

In this section, the overall performance analysis of the suggested Safety-Constrained Reinforcement Learning (SCRL) scheduler is provided in URLLC conditions in the industry. Averages of all the 10 independent simulation runs are made in results and 95% confidence intervals are calculated so as to give the results statistical robustness. The assessment is concerning the extreme latency behavior, violation probability, queue stability, real-time deployment because of reliability-latency tradeoffs, and computational feasibility.

8.1. Latency Distribution Analysis

8.1.1. Empirical CDF Evaluation

The cumulative distribution function (CDF) of the final receipt of packets is studied and works as:

$$\mathrm{FD(d)} = \Pr(\mathrm{D} \le \mathrm{d}).$$

There is a huge disparity in the behavior of tails among schedulers as shown in the CDF curves. The SCRL framework proposed has relative steeper slope towards the 1 ms URLLC threshold which shows better tail compression. Contrarily, the deterministic Earliest Deadline First (EDF) scheduler (when under heavy load) exhibits abruptly tail-inflating behaviour, whereas traditional reinforcement learning (RL) exhibits heavy-tail behaviour owing to the absence of explicit constraint enforcement. Lyapunov-only scheduler offers better stability but has no ability to do tail shaping by learning.

Table 3: Five-Nines (99.999%) Latency Comparison of Scheduling Algorithms

Scheduler	99.999% Latency (ms)
EDF	2.31
Traditional RL	1.47
Lyapunov-only	1.12
Proposed SCRL	0.94

SCRL decreases the extreme tail latency by:

- 59% compared to EDF
- 36% compared to traditional RL
- Compared to Lyapunov-only scheduling 16%.

These findings justify that integrating learning with safety constraints can increase the ultra-reliable delay performance greatly.

8.1.2. Tail Behavior Interpretation

The outcomes of the CDF show that EDF is affected by deadline clustering in the event that traffic is bursty, and therefore a synchronized deadline execution is achieved. The classical RL produces the greatest average reward but does not explain the latency spikes associated with the rare events. Lyapunov-only scheduling guarantees that the backlog is stable but does not take predictions into account when it optimizes myopically on a slot-by-slot basis.

SCRL attains compression during tailing as a result of 3 coupled mechanisms:

- Probabilistic deadline enforcement Virtual queue based constraint tracking.
- Safety-layer projection in order to avoid unsafe scheduling actions.
- Dynamic constraint penalties by means of adaptive dual-variable penalties.

This synergy allows successful management of the mean and the peak indicators of latency.

8.2. Tail Violation Probability

In the URLLC requirement, it is mandatory that:

$$\Pr(\mathrm{D} > 1\ \mathrm{ms}) \le 10^{-5}$$

8.2.1. Measured Violation Probability

Table 4: Tail Violation Probability Comparison Under 1 ms URLLC Constraint

Scheduler	Violation Probability
EDF	4.8×10^{-3}
Traditional RL	1.2×10^{-4}
Lyapunov-only	2.4×10^{-5}
SCRL	7.6×10^{-6}

It is only SCRL that meets the five-nines of reliability criterion.

8.2.2. Statistical Validation

The estimation of the probability of rare events should be strictly validated. The methodology that was followed was as follows:

- Wilson confidence intervals for binomial rare-event estimation
- Horizon of simulation 10^6 time slots.
- Hypothesis testing:

$$\mathrm{H0: Pviol} \geq 10^{-5}$$

SCRL does not accept $H0$ with the 95-percent confidence level and this validates statistically high obedience to URLLC reliability limits.

8.2.3. Rare-Event Stability

This probability of violation is constant at:

- Enhanced burstiness of traffic.
- 15% load surge
- High channel fading variance.

Such resilience explains that SCRL is not limited to nominal traffic assumptions and it becomes reliable even in stress conditions.

8.3. Queue Stability

8.3.1. Backlog Evolution

The dynamics of queue length show additional aspects of stability of schedulers. EDF has exponentially increasing backlog with congestion. The classical RL is oscillately unstable, as there is no hard constraint on the prioritization of rewards. Lyapunov-only scheduling has more stable queues but achieves bigger steady-state backlog.

SCRL has a constrained and smoother evolution of queues.

Table 5. Average Queue Backlog Comparison Across Scheduling Schemes

Scheduler	Average Backlog (packets)
EDF	84.2
Traditional RL	46.7
Lyapunov-only	32.5
SCRL	28.9

SCRL has the smallest average backlog all over a stable scheduler.

8.3.2. Stability near Capacity

At the point of system loads of 90 percent capacity:

- EDF fails because of congestion of deadlines.
- In the traditional RL, the reliability constraint is broken.
- Only Lyapounov remains, but is excessively conservative.
- SCRL maintains stability at the same time maintaining throughput efficiency.

This validates earlier existing theory of Lyapunov stability.

8.3.3. Variance Reduction

SCRL improves queue variance by 22 percent over Lyapunov-only scheduling, which means that control decisions are smoother and jitter is less important (which is essential in industrial automation systems).

8.4. Reliability-Latency Tradeoff

The two are the tradeoff between reliability and latency, using Lyapunov weight V and dual learning rate as the independent variables.

8.4.1. Observed Tradeoff Curve

With increasing limitations of reliability:

- Latency increases slightly.
- The throughput is reduced slightly.
- The probability of violation goes down exponentially.

Because SCRL offers a controlled Pareto frontier, operators are able to adjust reliability without using a high level of latency penalty.

8.4.2. Comparative Tradeoff Characteristics

Table 6. Qualitative Comparison of Reliability Control, Latency Efficiency, and Adaptability

Method	Reliability Control	Latency Efficiency	Adaptability
EDF	None	Poor	None
Traditional RL	Weak	Moderate	High
Lyapunov-only	Strong	Moderate	Low
SCRL	Strong	High	High

SCRL has good reliability control but is also latency efficient and adaptable.

8.4.3. Key Insight

In contrast to deterministic or completely being reactive types of schedulers, SCRL:

- Expects the accumulation of congestion.
- Punishes the stock of violations.
- Acquires channel knowledge of priority.

This results in efficient tail shaping and lack of over conservativeness.

9. Limitations and Future Work

Although theoretically assured and showing five-nines URLLC performance, the proposed SCRL framework can only be applied to a single-cell industrial deployment having centralized scheduling at present. In pragmatic industrial campuses, there are several gNBs running concurrently, creating inter-cell, mobility based instability and inter-cell resources coupled. The current formulation of the coordinated multi-point transmission (CoMP), the crossing-cell latency assurances and the interference-mediated queue dynamics are not explicit aspects in the current formulation by Lyapunov. Future studies need to expand SCRL to include multi-agent constrained reinforcement learning, in which its individual cells perform local agents along with coordinating and updating via consensus. This extension would be consistent with intelligent RAN architectures advocated by the O-RAN Alliance, and would be able to utilize hierarchical control loops such as those used in near-real-time and non-real-time RIC models. Nevertheless, extending Lyapunov stability ensures to the distributed, interference-sensitive environments is an open theoretical problem.

The other restriction is on the centralized training assumptions which might not be viable in the various industrial locations because of privacy, proprietary traffic patterns, and isolation of vendors restrictions. Future research may also be inspired by federated optimization technologies (put forward by H. Brendan McMahan) that federate local SCRL agents that are trained at separate facilities and regularly combined without raw data exchanged. Its major challenges are the management of the heterogeneous (non-IID) traffic distributions, coordination of dual variables in the global reliability, model updates in a communication-ileffective way, and convergence guarantees in non-stationary conditions. This type of federated SCRL architecture would allow privacy preserving cross-factory generalization and scalable AI-native RAN deployments.

Lastly, although the latest inference latency meets 1ms TTI requirements, a future with dense deployments, graph encoder based state encoders, or cell to cell coordination could increase computation requirements. Hardware-algorithm co-design will then be a requirement including that accelerates edges, FPGA-based deterministic inference, or an ASIC-level neural have been confused engine. Based on a look into the future of 6G networks encouraged by the International Telecommunication Union, future networks might need deterministic (hard) delay guarantees over probabilistic bounds, time-sensitive networking (TSN) integration, and AI-native control loops. The application of SCRL to bounded-delay reinforcement learning, network calculus integration, semantic-aware scheduling, and 6G-ready digital twins assisted optimization is a potential direction towards self-certifying, explainable, and completely autonomous 6G-ready intelligent RAN systems.

10. Conclusion

The paper provided a Safety-Constrained Reinforcement Learning (SCRL) framework to URLLC-aware scheduling in industrial radio access networks and has combined Lyapunov drift optimization, primal-dual constrained policy learning, actor-critic adaptation, and safety-layer projection into one architecture. SCRL applies stochastic network optimization concepts to deep reinforcement learning that results in proving the queue stasis and bluffing tight tail-latency proverses. Five-nines reliability (99.999 percent latency less than 1000 micro seconds) and violation probability less than 10^{-5} is proven by extensive simulations, bound backlog at high load, less queue variation, and competition at below milliseconds inference latency. SCRL is the first bridging theory and learning: in contrast to deterministic schedulers like EDF and unconstrained RL strategies, SCRL supports reliable, channel-adaptive, and real-time deployable scheduling, which is appropriate in industrial automation settings.

Industrially/standardization The structure is consistent with intelligent RAN evolution being advanced by the 3rd Generation Partnership Project and the O-RAN Alliance to support AI-enabled scheduling is based on the Near-RT RIC deployment models. SCRL meets fundamental Industry 4.0 demands of ultra-reliable, low latency wireless controls by providing predictive avoidance of congestion, statistical reliability assurances and safe online adaptation. This paper contributes to progress in designing AI-native RAN where it can be shown that stability aware and reliability aware reinforcement learning can stabilize to meet mission-critical constraints, and provides guidance on how in the future it can be extended to work with multi-cell coordination and federated constrained learning, hardware acceleration, and deterministic 6G networking.

Reference

[1] Zhang, W., Derakhshani, M., Zheng, G., & Lambotharan, S. (2024). Constrained risk-sensitive deep reinforcement learning for eMBB-URLLC joint scheduling. IEEE Transactions on Wireless Communications, 23(9), 10608-10624.

[2] Li, J., & Zhang, X. (2020). Deep reinforcement learning-based joint scheduling of eMBB and URLLC in 5G networks. IEEE Wireless Communications Letters, 9(9), 1543-1546.

[3] Li, Q., Chen, J., Cheffena, M., & Shen, X. (2023). Channel-aware latency tail taming in industrial IoT. IEEE Transactions on Wireless Communications, 22(9), 6107-6123.

[4] Khalifa, N. B., Assaad, M., & Debbah, M. (2019, April). Risk-sensitive reinforcement learning for URLLC traffic in wireless networks. In 2019 IEEE Wireless Communications and Networking Conference (WCNC) (pp. 1-7). IEEE.

[5] Wang, X., Yao, H., Mai, T., Guo, S., & Liu, Y. (2023). Reinforcement learning-based particle swarm optimization for end-to-end traffic scheduling in TSN-5G networks. IEEE/ACM Transactions on Networking, 31(6), 3254-3268.

[6] Destounis, A., Paschos, G. S., Arnau, J., & Kountouris, M. (2018, May). Scheduling URLLC users with reliable latency guarantees. In 2018 16th International

Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt) (pp. 1-8). IEEE.

[7] Upadhyay, D., Soni, M., Gupta, S., Sharma, R., & Venu, N. (2025, March). Latency-Aware Network Slicing for 5G URLLC Applications: Design and Optimization Strategies. In 2025 3rd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) (pp. 113-118). IEEE.

[8] Praveen, S., Khan, J., & Jacob, L. (2021, July). Reinforcement learning based link adaptation in 5G URLLC. In 2021 8th International Conference on Smart Computing and Communications (ICSCC) (pp. 159-163). IEEE.

[9] Alsenwi, M., Tran, N. H., Bennis, M., Pandey, S. R., Bairagi, A. K., & Hong, C. S. (2021). Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: A deep reinforcement learning based approach. IEEE Transactions on Wireless Communications, 20(7), 4585-4600.

[10] Shi, W., Ganjalizadeh, M., Ghadikolaei, H. S., & Petrova, M. (2023, September). Communication-efficient orchestrations for urllc service via hierarchical reinforcement learning. In 2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) (pp. 1-6). IEEE.

[11] Azari, A., Ozger, M., & Cavdar, C. (2019). Risk-aware resource allocation for URLLC: Challenges and strategies with machine learning. IEEE Communications Magazine, 57(3), 42-48.

[12] Haque, M. E., Tariq, F., Khandaker, M. R., Wong, K. K., & Zhang, Y. (2023). A survey of scheduling in 5G URLLC and outlook for emerging 6G systems. IEEE access, 11, 34372-34396.

[13] Sohaib, R. M., Onireti, O., Sambo, Y., Swash, R., Ansari, S., & Imran, M. A. (2023). Intelligent resource management for eMBB and URLLC in 5G and beyond wireless networks. IEEE access, 11, 65205-65221.

[14] Liang, F., Yu, W., Liu, X., Griffith, D., & Golmie, N. (2021). Toward deep Q-network-based resource allocation in industrial Internet of Things. IEEE internet of things journal, 9(12), 9138-9150.

[15] Shaik, R. B., Nagaradjane, P., Ioannou, I., Sittakul, V., Vasiliou, V., & Pitsillides, A. (2024). AI/ML-aided capacity maximization strategies for URLLC in 5G/6G wireless systems: A survey. Computer Networks, 249, 110506.

[16] Salh, A., Ngah, R., Hussain, G. A., Alhartomi, M., Boubkar, S., Shah, N. S. M., ... & Alzahrani, S. (2024). Bandwidth allocation of URLLC for real-time packet traffic in B5G: A Deep-RL framework. ICT Express, 10(2), 270-276.

[17] Wang, J., Zheng, Y., Wang, J., Shen, Z., Tong, L., Jing, Y., ... & Liao, Y. (2023). Ultra-reliable deep-reinforcement-learning-based intelligent downlink scheduling for 5G new radio-vehicle to infrastructure scenarios. Sensors, 23(20), 8454.

[18] Neely, M. (2010). Stochastic network optimization with application to communication and queueing systems. Morgan & Claypool Publishers.

[19] Al-Saadeh, O., Wikstrom, G., Sachs, J., Thibault, I., & Lister, D. (2018, December). End-to-end latency and reliability performance of 5G in London. In 2018 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6). IEEE.

[20] Dahlman, E., Parkvall, S., & Skold, J. (2023). 5G/5G-advanced: the new generation wireless access technology. Elsevier.

[21] Khoshnevisan, M., Joseph, V., Gupta, P., Meshkati, F., Prakash, R., & Tinnakornsrisuphap, P. (2019). 5G industrial networks with CoMP for URLLC and time sensitive network architecture. IEEE Journal on Selected Areas in Communications, 37(4), 947-959.

[22] Hamidi-Sepehr, F., Sajadieh, M., Panteleev, S., Islam, T., Karls, I., Chatterjee, D., & Ansari, J. (2021). 5G URLLC: Evolution of high-performance wireless networking for industrial automation. IEEE Communications Standards Magazine, 5(2), 132-140.

[23] Khan, B. S., Jangsher, S., Ahmed, A., & Al-Dweik, A. (2022). URLLC and eMBB in 5G industrial IoT: A survey. IEEE Open Journal of the Communications Society, 3, 1134-1163.

[24] Anand, A., De Veciana, G., & Shakkottai, S. (2020). Joint scheduling of URLLC and eMBB traffic in 5G wireless networks. IEEE/ACM Transactions On Networking, 28(2), 477-490.

[25] Yoshizawa, T., Baskaran, S. B. M., & Kunz, A. (2019, October). Overview of 5G URLLC system and security aspects in 3GPP. In 2019 IEEE Conference on Standards for Communications and Networking (CSCN) (pp. 1-5). IEEE.