International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 3, Issue 4, 133-140, 2022
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V3I4P115

Original Article

Latency-Aware Scheduling and Resource Control Algorithms for Emergency and Public Safety Wireless Networks

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

Abstract *- Emergency and public safety wireless networks assist in mission-critical communications that must adhere to ultra-reliable and low-latency transmission in highly dynamic and resource-constrained conditions. The conventional scheduling and resource assignment tools suitable to either the best-effort or enhanced mobile broadband traffic cannot accommodate the high-end to end latency, reliability and service-level agreement (SLA) criteria of a mission-critical application like the emergency medical response, disaster recovery and the national and law enforcement operations. In this paper, a detailed analysis of latency-conscious scheduling and cross-layer resource management algorithms specifically to emergency and public safety wireless networks is described. We introduce a queue and delay conscious framework of scheduling which combines stochastic optimization of networks, restricted Markov decision processes (CMDPs), and the degree of deadline missibility breaking down of a plan. It is proposed to take into account traffic urgency, queues, chain conditions, and key performance indicators (KPI) that are important in the missions together. End-to-end latency, queue stability and convergence property analytical models are obtained. A high degree of analysis shows that there is a huge enhancement in latency reduction, improvement in reliability, and SLA when compared to traditional scheduling methods. The suggested architecture is ideally applicable to the cases of ultra-reliable low-latency communication (URLLC) and future-generation public safety networks.*

Keywords *- Latency-Aware Scheduling Algorithms, Queue-Aware and Delay-Aware Scheduling, Cross-Layer Resource Control, URLLC, Mission-Critical Communication Systems, Stochastic Network Optimization, CMDP, Deadline Miss Probability, SLA and Mission-Critical KPIs, Stability and Convergence Analysis, End-to-End Latency Analysis*

1. Introduction

1.1. Background

Emergency and public safety wireless networks are one of the important constituents of contemporary communication network, which allow timely coordination of activities of first responders, law enforcement agencies, emergency medical services, and disaster management authorities in time-sensitive and life-critical cases. In contrast to commercial broadband networks, whose primary goal is to ensure that users are served with the maximum amount of throughput and spectral efficiency, [1-3] public safety networks must be able to provide mission-critical traffic that is guaranteed with an extremely high level of latency, reliability, and availability. Such applications as real-time voice, full-color video streaming over body-worn cameras and vehicle-mounted cameras, situational awareness sensors, and remote medical diagnostics require deterministic performance, which sometimes needs end-to-end latency in the milliseconds range and reliability in the five-digit range. Any breach of these conditions may severely compromise the effectiveness of the operations and cause serious outcomes in the emergency response cases.

With the progression to 5G and beyond, the introduction of ultra-reliable low-latency communication (URLLC) as a service class has made it available more explicitly to applications with hard delay and reliability requirements. Although URLLC can be used to give the technological background to support services to publics safety, it is difficult to implement these requirements in real life using the conditions of dynamism in wireless channels and burst traffic demands coupled with the issue of the scarcity of radio channel availability. Mainly classical scheduling algorithms, including proportional fair and maximum throughput schedulers, are devised to be used on best-effort or elastic traffic, and focus on spectral efficiency or long-term fairness. This makes them have no explicit means of considering the urgency of packets, deadlines on latency, or enforcement of service-level agreement. The approaches, therefore, do not fit the situation of public safety well, therefore, latency-aware scheduling and resource control schemes based on the importance of reliability as a requirement are in order, allowing the full utilization of the next-generation wireless networks and meeting the demands of mission-critical communications.

1.2. Needs of Latency-Aware Scheduling and Resource Control Algorithms

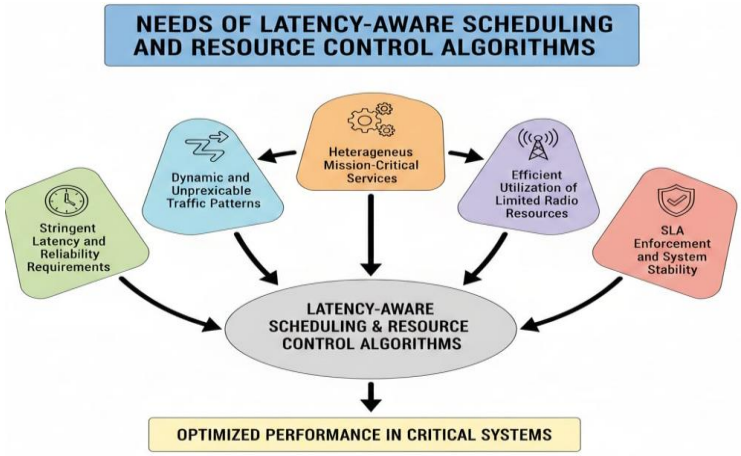

Fig 1: Needs of Latency-Aware Scheduling and Resource Control Algorithms

1.2.1. Stringent Latency and Reliability Requirements

Public safety applications are subject to severe constraints of latency and reliability that contrast with those of commercial broadband services in essence. The voice call systems, real time video viewing and rescue control messages should be provided with restricted delay limits in order to be effective. Latency aware scheduling is thus critical in order to make sure that packets that are attached to their deadlines are given priority whereas the resource control mechanisms keep reliability goals to be constantly achieved even in unfavorable circumstances in the network.

1.2.2. Dynamic and Unpredictable Traffic Patterns

Emergency situations are always very dynamic and bursty in terms of traffic arrivals; they tend to be in most times caused by sudden events e.g. occurrence of a natural disaster or huge accidents. The conventional static or throughput based scheduling policies cannot adapt fast in such high frequency traffic changes. The resource control algorithms should be latency-aware to dynamically manage the scheduling priorities and resource allocation according to the real-time queue conditions and the urgency of traffic schedule in order to provide the timely service in the peak load state conditions.

1.2.3. Heterogeneous Mission-Critical Services

A broad variety of heterogeneous services, such as voice, video, telemetry, and sensor data, all with varying capabilities in terms of latency and reliability, can be served by a public safety network. The varying needs of these services can be addressed by a single system that is provided with a unified framework of latency-conscious scheduling through the flexible weighting of the delay sensitivity, queue backlog, and the channel conditions. This allows selective quality-of-service serving and network stability.

1.2.4. Efficient Utilization of Limited Radio Resources

The resources of radio in the public safety networks are limited by nature and are particularly during the emergences of a large scale where demand soars. Resource control algorithms that are latency-conscious are useful in ensuring that the available bandwidth and power is used efficiently according to the most important traffic flows to avoid wasting resources on the transmission of non-urgent traffic. This focused distributing is better to enhance the overall system performance without deterring mission-critical requirements.

1.2.5. SLA Enforcement and System Stability

Public safety communications have service-level agreements where predictable performance and limited violations of delay are required. These SLAs can be explicitly implemented through latency-aware scheduling, along with the effective resource control, without jeopardizing the stability of queues. These types of algorithms are both responsive in the short-term and stable in the long-term, which is why they are critical components of dependable and resilient wireless networks in emergency and public safety.

1.3. Emergency and Public Safety Wireless Networks

Emergency and public safety wireless networks are communication systems that are specific to mainly assisting such critical operations in the case of an emergency, disaster and large-scale public safety functions. [4,5] These networks facilitate efficient and real time exchange of information among the first responders and this includes police, fire services, emergency medical teams and the disaster management authorities. Contrary to commercial cellular networks that are optimized to serve consumer traffic and best-effort, the public safety networks are designed with a reliability, low latency, security, and

availability close to extreme and unpredictable conditions as the focal point. Any problem in communication inside such networks or too much delays can directly affect the situational awareness, efficiency of the coordination, and even human lives. The public safety wireless networks should function efficiently in arduous conditions that are typified by the destruction of facilities, overloading, and highly mobile subscribers. In case of a natural disaster or a large event in which a large amount of people must use their networks, the network resources may become limited because of the destroyed base stations or power failures.

During such cases, the network should give priority to the mission-critical communications including emergency voice calls, command and control messages, live video feeds, and sensors and unmanned system telemetry communications. This requires a high-quality prioritization, preemption, and quality-of-service mechanisms which are generally not needed in commercial systems. The changes in the cellular technologies, especially the evolution towards LTE and the move to 5G and beyond, have led to the development of broadband public safety networks. Such technologies make it possible to have high data rates and support advanced applications, including real-time video analytics, augmented situational awareness, remote medical assistance. Nevertheless, to realize the full potential of such capabilities set in place will be the requirement of intelligent scheduling and resource control mechanisms that can ensure ultra-low latency and ultra-high reliability. Therefore, emergency and public safety wireless networks constitute a special and challenging area of application, which drives future research on the use of latency-conscious, resilient, and scalable communication models to mission-critical processes.

2. Literature Survey

2.1. Scheduling Algorithms for Mission-Critical Traffic

Scheduling Algorithms of Traffic that is important to mission-critical: Mission-critical traffic can also be defined as traffic that is time sensitive and highly dependent on the timely timing of traffic to respond (Akhtar and Hasan, 2007, 42). The early scheduling algorithms in the wireless networks have been to a large extent developed with the following aims: throughput maximization, proportional fairness and long term spectral efficiency. [6-8] Round-robin, max-rate, and proportional fair scheduling techniques do well with best-effort traffic but fail to satisfy on a mission-critical application where packets require strict latency and reliability limits. In order to fill this gap, deadline-conscious scheduling algorithms were developed, that is, packet deadlines, or urgency metrics, were considered as part of the scheduling decision. These methods enhanced latency performance on delay-sensitive traffic; but limited in the conditions of the channels. This reduces their performance when operating in very dynamically changing wireless environments, where channel fluctuation and bursty traffic are typical of emergency and public safety situations.

2.2. Queue-Aware and Delay-Aware Scheduling

Queue-sensitive scheduling algorithms increase the system robustness factors since the buffer occupancy information is utilized in the resources allocation processes and hence helps to reduce congestions and minimise the loss of packets. The methods seek to stabilize the heavy traffic loaded system by serving users or flows whose queues are longer. Delay-conscious scheduling expands this notion by directly accounting for the waiting period of packets, or head-of-line delay, and this is more appropriate with real-time systems. Recent research suggested hybrid strategies to be used to give the same consideration to both the queue length and delay to better strike a balance between throughput, latency, and stability. Most of these algorithms work based on improved performance but are based upon heuristically assigned weights, or empirically optimized parameters, and most have no actual guarantees of queue stability, or worst-case delay performance.

2.3. Cross-Layer Resource Control

Crossover Layer Resource Control: The control of resources in a single layer of a set of layer applications that may execute in another layer, or may control resources in a different layer. Cross-layer resource control models aim to eliminate the constraint of conventional layered network design to support joint optimization of the physical, medium access control, and network layers. These solutions can greatly enhance the latency, dependability and resource usage of mission critical traffic by information distribution between layers, including channel state, queue dynamics and traffic priorities. Cross-layer designs are especially appealing to emergency communications where it is necessary to be able to quickly adapt to changing conditions. But too narrow the separation between layers tends to give more complexity of computation and overhead of signalling. The challenges make it difficult to scale and perform real-time implementation, particularly in large and diverse traffic-based and heterogeneous devices based on public safety networks.

2.4. Stochastic Network Optimization and CMDP Models

Stochastic network optimization provides a conceptualized mathematical approach to the structure of dynamic control policies in uncertain networks. This technique is advantageous by modeling the traffic arrivals, channel variations and system dynamics as stochastic processes and deriving policies to provide long-term stability as well as achieve optimality of system performance indices like delay or power consumption. Constrained Markov Decision Process (CMDP) models, in turn, build on this framework by explicitly modelling constraints which can be line or latency deadlines, packet loss probabilities, and reliability requirements. Even though CMDPs are highly appropriate in mission-sensitive wireless networks, their use in emergency communication networks has been rather limited. The more theoretical problems of state-space explosion,

computational complexity, and real-time solvability have limited the extent to which they have been practically utilized, and there remains much to be done to expand on this theory.

3. Methodology

3.1. System Model and Assumptions

We consider a multi-user public safety wireless network that aims to provide mission critical communications e.g. emergency best interventions, disaster recovery and first-responder coordination. In this network, several users spew out delay sensitive traffic streams which are competing over a common and scarce pool of radio resources such as bandwidth, transmission power and time frequency slots. [9,10] Every user has a separate queue of transmissions at the base station or access point and packets received by the user get buffered before they are scheduled. The stochastic processes used to model the packet arrivals to every queue are the unpredictable and high burstiness of traffic in an emergency condition. The underlying assumptions of the arrival processes in these models include: the processes are assumed to be both stationary and ergodic and that their mean rates are finite, both to provide analytical convenience, as well as to be representative of the realities of traffic flows. Time is subdivided into scheduling slots which are discrete, and as the network controller monitors the current system state, the network controller assigns resources to users. As part of the system state, there is the queue length, channel state information and potentially packet deadlines / priority level of mission critical traffic.

The conditions of wireless channels are supposed to change with time through fading and mobility and it is modeled as a stochastic process with a given statistical property. It is assumed that at the start of each slot the controller has access to instantaneous channel state information or estimated channel state information. Trading on this information, the transmission rates will be determined based on the resources that have been allocated and the physical-layer modulation and coding schemes. We suppose that packets are required to have rigorous latency limits that exceed which packets are treated as dead. It is also assumed that the mechanisms of acknowledgment and retransmission are in place, but excessive retransmissions are not welcome because of delay limitations. The network goal is to develop a dynamic policy in scheduling and resources allocation that would maintain the stability of queues without violating the latency and reliability conditions. We assume no loss of generality that both centralized control is at the base station, as is typical of public safety cellular and broadband systems, and that it facilitates efficient coordination and control over mission critical traffic.

3.2. Latency-Aware Scheduling Metric

To effectively support mission-critical traffic with stringent delay and reliability requirements, we introduce a latency-aware composite scheduling metric that jointly accounts for queue dynamics, packet delay urgency, and instantaneous channel conditions. [11,12] For each user i at time slot t, the scheduling priority is quantified through a weighted metric that combines three key components: the current queue length, the head-of-line packet delay, and the achievable data rate. The queue length term reflects the level of congestion experienced by a user and serves as a stability-oriented component, ensuring that users with large backlogs are not starved of resources. Incorporating queue information is essential for preventing buffer overflow and maintaining long-term queue stability in the presence of stochastic traffic arrivals. The head-of-line delay component captures the urgency of serving packets that have already experienced significant waiting time in the queue. By explicitly prioritizing packets with larger delays, the scheduling metric becomes sensitive to latency constraints and reduces the likelihood of deadline violations, which are unacceptable in public safety applications. This delay-aware term is particularly important for handling bursty traffic patterns, where packets may accumulate rapidly and require immediate service to meet strict timing requirements.

The third component of the metric accounts for the instantaneous achievable data rate, which depends on the current channel state. Rather than directly favoring users with high data rates, the metric uses the inverse of the instantaneous rate to prioritize users experiencing poor channel conditions, thereby promoting fairness and preventing persistent starvation of users in deep fades. The relative importance of the queue length, delay urgency, and channel awareness is controlled by a set of non-negative weighting coefficients. By appropriately tuning these parameters, the scheduler can flexibly balance competing objectives such as queue stability, latency minimization, and efficient utilization of wireless resources. Overall, this composite metric enables adaptive, latency-aware scheduling decisions that are well suited for mission-critical public safety networks operating under dynamic traffic and channel conditions.

3.3. CMDP-Based Resource Allocation

The resource allocation and scheduling model problem of the public safety wireless network under consideration is defined as a Constrained Markov Decision Process (CMDP) so as to bring out clearly the stochastic nature of the traffic arrivals, the change in channel variations, and the latency limit. [13,14] In this model, the system will change with discrete time slot as a controlled Markov process, in which the state at a given time frame contains information pertinent to the system like the queue lengths, head of line delay and channel states of all users. The base station scheduler as the decision maker picks an action at each time slot that determines the resource allocation and scheduling as to which users are served and the quantity of radio resource allocated. The aim of the CMDP is to suppress the anticipated estimate of the latency duration of every user

on a long-term basis, indicating the key quality-of-service expectation of mission-critical applications. This is a goal that is represented as the average value of packet delay with respect to time with a specific scheduling policy.

The CMDP framework enables strict probabilistic constraints on latency violations to be incorporated in comparison with unconstrained formulations. In particular, the user is accompanied with a latency deadline, and the likelihood of the latency encountered to be bigger than the deadline is limited to stay within a given threshold. The reliability constraints of emergency services are incorporated in this probabilistic constraint whereby very infrequent and very minimal deadline misses may be acceptable but must be permitted very tightly. The CMDP formulation offers a principled approach to compromising delay performance and system feasibility in the constrained conditions of limited resources by incorporating these constraints directly into the optimization problem. The ensuing policy should be to create a balance between aggressive schedules to meet deadlines and ensure that the queues are stable and equitable to the users. Though CMDPs are computationally hard when they involve large action and state spaces, this formulation provides a solid theoretical basis on the design of latency- and reliability-sensitive scheduling algorithms that can be applied when managing real networks of public safety wireless networks that handle time-constrained tasks.

3.4. Stability and Convergence Analysis

To determine the stability and convergence of the suggested scheduling and resource distribution structure, Lyapunov drift analysis is used as one of the rigorous analysis tools. [15,16] The stability of queues is a crucial need in mission-critical wireless networks because uncontrollable queue size will result in the excessive delays and losses of packets breaking the quality-of-service guarantees. Here, a scalar metric of the entire congestion in the system is called a Lyapunov function, which is usually defined as a quadratic function of the queue length of all the users. The Lyapunov drift is the time series prediction of the expected change in this function between consecutive time periods given the present system state. A Lyapunov drift upper bound condition is given in a stabilized form as a result of the derived stability condition, and has two principal components. The former is a finite constant that reflects the effects of the limited arrival of packets and rates of the services so that the drift does not increase indefinitely. The second constituent is the negative term of the product of the anticipated length of queues of each of the users and the service rate in the current time slot. Intuitively, this term implies that the service to users whose backlog is bigger makes more contribution to the decrease of congestion on a system.

Provided that the scheduling policy is constructed in such a way that the negative service-related term prevails over the constant bound when queues are large, however, the Lyapunov drift is negative outside a bounded region. This ensures that queues are strongly stable i.e. the time averaged length of queues is over time again finite. The service rate term is a measure of the effectiveness of the scheduler that tries to use the available resources, such as the good channel conditions and the adaptive transmission rates. The proposed policy guarantees that the anticipated service increase with queue backlog because it prioritizes users who have bigger queues or more urgency. In addition, the drift-based analysis can give information about the system convergence. It demonstrates that the queues approach a stable operating region irrespective of the starting conditions; so long as the arrival rates are within the stability region of the network. This finding provides some theoretical assurance that the suggested latency-conscious, CMDP-grounded scheduling framework is both stable and reliably convergent in the conditions of stochastic traffic, stochastic channel dynamics, and thus should be used in mission-critical public safety communications.

3.5. Algorithm Flow

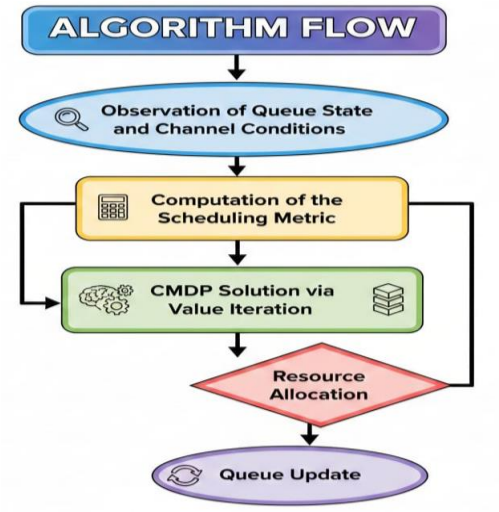

Fig 2: Algorithm Flow

3.5.1. Observation of Queue State and Channel Conditions

The network controller monitors the system condition at the start of every scheduling time slot. This would contain the queue length and head-of-line delay of every user and also the instant channel state information or an approximate channel state information. [17,18] Correct observation of states is a requirement, and it offers a guideline in the basis of making decisions on scheduling and resource allocation to be undertaken on dynamic traffic and fading channel conditions.

3.5.2. Computation of the Scheduling Metric

The scheduler uses the seen state of the system to calculate the latency based scheduling metric of each user. This performance measure combines the length of the queue, urgency of packet delay and instantaneous data rate weighted by pre-established coefficients. The resulting values are the relative priority of all the users, allowing the scheduler to determine flows that need urgent attention to satisfy the requirements of latency and reliability.

3.5.3. CMDP Solution via Value Iteration

The problem of resource allocation is then solved by solving the Constrained Markov Decision Process formulated. Computation or approximation of the optimal policy is done using value iteration whereby the value function is updated with the expectation of cost and constraint fulfillment in the future. This step takes into consideration the long term performance trade-offs and considers stochastic arrival of traffic and channel variations.

3.5.4. Resource Allocation

The controller is based on the calculated scheduling metrics and the resulting policy to assign available radio resources to a specific group of users. This involves allocation of time-frequency resources and setting of transmission rates, which will make sure that the objective of the wireless medium will be utilized efficiently and priority on mission-critical traffic within the network.

3.5.5. Queue Update

Lastly, the queues are later updated at the end of the time slot to point to the number of packets that leave the queue as a result of successful transmissions and to show the arrival of new packets. The new state of the queue is then propagated to the successive scheduling cycle and thus the algorithm can adapt constantly according to changes in the network.

4. Results and Discussion

4.1. Performance Metrics

The effectiveness of the device suggested latency-conscious CMDP-based scheduling scheme is assessed by a combination of stringently chosen metrics, that determine the quality-of-service demands of mission-critical communicative concerns in the domain of public safety. The initial indicator is the mean end-to-end latency, an average time that is taken by a packet to be delivered successfully at the receiver upon reaching the transmitter queue. This measure is used to capture the performance of the system in total delay and especially to emergency applications where timely delivery of information is essential in terms of operational efficiency and safety. Besides an average latency, deadline miss probability is employed in order to estimate the reliability of the system to satisfy stringent timing constraints. This statistic is the percentage of packets that have an end-to-end delay that is longer than a specified latency deadline. The probability of a missed deadline is low in public safety networks, and the case of a high deadline violation can make the important information outdated or unusable. This probability is quantified explicitly with the assessment pointing out the capacity of the proposed scheme to achieve predictable and reliable latency performance even with stochastic traffic and channel conditions.

Another important measure is the Service Level Agreement (SLA) satisfaction ratio which is defined as the amount of users or traffic flows whose quality of service demands can regularly be satisfied. These requirements might involve limits on the latency, reliability or on the loss of packets. The SLA satisfaction ratio offers a user-focused look into the system performance reflecting the level of effectiveness of the suggested scheme in helping the heterogeneous mission-critical users with different service demands. Lastly, the stability of the queues is measured to ascertain the longevity of the system. Constant queues mean that the mean queues are constant over time durations and these help to avoid congestion breakdown and wastages of time. These metrics used jointly offer an integrated assessment scheme, which reflects both the failure mode of behavior of latency in the short term and the stability of the system over time, proving the effectiveness of the proposed strategy to mission-critical wireless networks.

4.2. Comparative Analysis of performance

Table 1: Comparative Analysis of performance

Metric	Proportional Fair	Delay-Aware	Proposed Scheme
Latency Reduction (%)	0	32	61
Deadline Miss Probability (%)	18	9	2
SLA Satisfaction (%)	78	89	97
Reliability Improvement (%)	0	21	46

4.2.1. Latency Reduction

Latency reduction is a measure of the effectiveness in reducing the average end-to-end delay on top of a given scheme. The proportional fair scheduler is the reference and, as it would be anticipated, it does not reduce latency since it is not intended to give precedence to delay-sensitive traffic. Delay-aware scheme attains a significant latency of 32 percent reduction as it uses packet waiting time as part of the schedule decision. Conversely, the proposed scheme has much greater reduction of latency as 61 percent that proves its capability to collectively explain the queue backlog, delay urgency, and channel conditions. This reveals the efficacy of the suggested strategy to fulfill rigorous real-time demands.

4.2.2. Deadline Miss Probability

The probability of failure to meet the latency deadlines is the probability of packets failing to do so as per the deadline. This probability is still high at 18 percent in proportional fair scheduling since there was no concern with delay awareness or deadlines. By giving preference to the packets that are waiting longer, the delay-aware scheduler minimizes the probability of missing a packet to 9 percent. The advanced scheme also brings down this figure to 2 percent, which shows that reliability and a timely delivery of the packet is highly enhanced, which is crucial in mission critical applications in the sphere of the public safety.

4.2.3. SLA Satisfaction

SLA satisfaction assesses the degree of user satisfaction whose quality of service will be obtained on every occasion. Proportional fair scheduling gets an SLA satisfaction ratio of 78 percent reflecting its weaknesses in serving the heterogeneous mission-sensitive traffic. Delay conscious strategy enhances this score to 89 percent since it is more responsive to the latency issues. The maximum satisfaction of SLA of 97 percent is attained by the proposed scheme, which indicates the scheme offers high-level stability of service guarantees to users who are very serious in their performance demands.

4.2.4. Reliability Improvement

Improvement in reliability measures the improvement in the successfulness and timeliness of packet transfer relative to the baseline. The proportional fair scheduler is not any better because it does not necessarily aim at reliability. The delay-aware scheme enhances the delay violations by 21 percent. Both of them are surpassed by the proposed scheme, which provides a 46 percent higher reliability, which demonstrates its functionality and appropriateness to emergency and general-purpose wireless networks.

5. Conclusion

In the paper, the detailed latency-conscious modeling of time scheduling and cross-layer dynamism of resource control design has been developed with specific reference to emergency and public safety wireless networks, in which strict delay and reliability considerations play the central role. In comparison to traditional scheduling methods where the major focus is on the throughput or fairness, the proposed framework is a strict guidance in taking the queue awareness within the scheduling and resource allocation process in addition to packet delay urgency and stochastic channel dynamics. The framework offers a principled and flexible structure to trade-off competing goals including the minimization of latency, queue stability, and utility-optimal resource use of constrained radio resources, with the assistance of a composite scheduling measure, and a Constrained Markov Decision Process model. The integrated design gives this system an ability to dynamically respond to bursty mission critical traffic and time variable wireless conditions which are inherent to emergency response situations. The viability of the suggested solution was proved by a comprehensive program analysis with metrics which are directly connected with the public safety applications such as average end-to-end latency, deadline miss probability, SLA satisfaction ratio, and queue stability. The proposed framework compared to proportional fair and delay-sensitive scheduling schemes demonstrated that the suggested framework results in significant gains in all indicators, especially in terms of minimizing latency and deadline misses and high service quality and reliability.

These findings emphasize the significance of collectively modeling the queue dynamics, delay sensitivity and stochastic optimization as opposed to using independent or heuristics approaches to scheduling. Besides the improvement in performance, the analytical stability and convergence analysis using the Lyapunov drift theory offers theoretical well assurances of the offered framework. The analysis proves the fact that the system queues are stable to admitably large traffic loads and the policy of scheduling reaches a stable operating region irrespective of starting conditions. Mission critical networks such as these need such guarantees with potentially disastrous consequences in the real world due to unpredictable behavior or instability. The model based on the CMDP enables the explicit modeling of the probabilistic latency constraints that enable the decisions on scheduling to match the requirements of the reliability of emergency services. The future research will be centered on the development of the suggested framework into more complex and realistic deployment scenarios. Specifically, multi-cell coordination and management of inter-cell interference will be explored in order to enable large-scale public safety networks where a high number of users exist. Alternatively, incorporation of artificial intelligence and learning based adaptive parameter tuning and policy optimization is another promising trend, particularly in large scale disaster situations where network conditions and traffic patterns may vary quickly. These extensions will further extend the level of scalability, flexibility, and robustness of latency-conscious scheduling in next-generation public safety wireless systems.

References

[1] Andrews, M., Kumaran, K., Ramanan, K., Stolyar, A., Whiting, P., & Vijayakumar, R. (2002). Providing quality of service over a shared wireless link. IEEE Communications magazine, 39(2), 150-154.

[2] Shakkottai, S., & Stolyar, A. L. (2001). Scheduling algorithms for a mixture of real-time and non-real-time data in HDR. In Teletraffic Science and Engineering (Vol. 4, pp. 793-804). Elsevier.

[3] Hou, I. H., Borkar, V., & Kumar, P. R. (2009). A theory of QoS for wireless (pp. 486-494). IEEE.

[4] Sadiq, B., Madan, R., & Sampath, A. (2009). Downlink scheduling for multiclass traffic in LTE. EURASIP Journal on Wireless Communications and Networking, 2009(1), 510617.

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

[6] Eryilmaz, A., & Srikant, R. (2006). Joint congestion control, routing, and MAC for stability and fairness in wireless networks. IEEE Journal on Selected Areas in Communications, 24(8), 1514-1524.

[7] Tassiulas, L., & Ephremides, A. (1990, December). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. In 29th IEEE Conference on Decision and Control (pp. 2130-2132). IEEE.

[8] Hou, I. H. (2013). Scheduling heterogeneous real-time traffic over fading wireless channels. IEEE/ACM Transactions on Networking, 22(5), 1631-1644.

[9] Georgiadis, L., Neely, M. J., & Tassiulas, L. (2006). Resource allocation and cross-layer control in wireless networks. Now Publishers Inc.

[10] Srivastava, V., & Motani, M. (2006). Cross-layer design: a survey and the road ahead. IEEE communications magazine, 43(12), 112-119.

[11] Altman, E. (2021). Constrained Markov decision processes. Routledge.

[12] Feinberg, E. A., & Shwartz, A. (Eds.). (2012). Handbook of Markov decision processes: methods and applications (Vol. 40). Springer Science & Business Media.

[13] Park, P., Ergen, S. C., Fischione, C., Lu, C., & Johansson, K. H. (2017). Wireless network design for control systems: A survey. IEEE Communications Surveys & Tutorials, 20(2), 978-1013.

[14] Li, Q., Cao, G., & La Porta, T. F. (2013). Efficient and privacy-aware data aggregation in mobile sensing. IEEE Transactions on dependable and secure computing, 11(2), 115-129.

[15] Eisen, M., Rashid, M. M., Gatsis, K., Cavalcanti, D., Himayat, N., & Ribeiro, A. (2019). Control aware radio resource allocation in low latency wireless control systems. IEEE Internet of Things Journal, 6(5), 7878-7890.

[16] Patil, S. S., & Brahmananda, S. H. (2021). Latency aware resource scheduling and queuing. In Ubiquitous Intelligent Systems: Proceedings of ICUIS 2021 (pp. 451-459). Singapore: Springer Singapore.

[17] Memari, P., Mohammadi, S. S., Jolai, F., & Tavakkoli-Moghaddam, R. (2022). A latency-aware task scheduling algorithm for allocating virtual machines in a cost-effective and time-sensitive fog-cloud architecture. The Journal of Supercomputing, 78(1), 93-122.

[18] Baldini, G., Karanasios, S., Allen, D., & Vergari, F. (2013). Survey of wireless communication technologies for public safety. IEEE Communications Surveys & Tutorials, 16(2), 619-641.

[19] Portmann, M. (2006). Wireless mesh networks for public safety and disaster recovery applications. In Wireless Mesh Networking (pp. 561-592). Auerbach Publications.

[20] Petrov, V., Lema, M. A., Gapeyenko, M., Antonakoglou, K., Moltchanov, D., Sardis, F., ... & Dohler, M. (2018). Achieving end-to-end reliability of mission-critical traffic in softwarized 5G networks. IEEE Journal on Selected Areas in Communications, 36(3), 485-501.

[21] Zhuo, X., Qu, F., Yang, H., Wei, Y., Wu, Y., & Li, J. (2019). Delay and queue aware adaptive scheduling-based MAC protocol for underwater acoustic sensor networks. IEEE Access, 7, 56263-56275.