

Weighted Random Early Detection in Modern Active Queue Management: A Structured Literature Review (2002–2025)

Abubakar Mukhtar Ammani, Aminu Adamu, Abubakar Aminu Mu'azu

Department of Computer Science, Faculty of Natural and Applied Sciences, Ummaru Musa Yar'adua University, Katsina

Corresponding Author: Aammani91@gmail.com

ABSTRACT

Active Queue Management (AQM) mechanisms are critical for mitigating congestion and maintaining quality of service in packet-switched networks. Weighted Random Early Detection (WRED) remains widely deployed due to its support for traffic differentiation within DiffServ architectures. However, the broader AQM landscape has evolved to include adaptive, rate-based, delay-based, and learning-driven algorithms designed to address limitations of early queue-length-based approaches. This study presents a structured literature review examining WRED's position within modern AQM frameworks from 2002 to 2025. A systematic search across major academic databases yielded 168 records, from which 34 studies were selected for detailed synthesis following eligibility screening. Comparative analysis indicates that WRED effectively preserves priority traffic under moderate congestion but remains sensitive to parameter configuration and may degrade fairness under sustained load. In contrast, adaptive and delay-based AQMs improve stability and latency control but lack intrinsic service differentiation. The findings highlight WRED's continued relevance in QoS-aware environments while emphasizing the need for adaptive and fairness-aware enhancements to improve robustness in heterogeneous network conditions.

KEYWORDS: WRED, AQM, Congestion Control, Quality of Service, Adaptive Queue Management, Network Fairness.

I. INTRODUCTION

Network congestion remains a persistent challenge in modern packet-switched networks, directly affecting throughput, latency, packet loss, and overall Quality of Service (QoS). As Internet traffic continues to grow in volume and diversity driven by cloud computing, multimedia streaming, Internet of Things (IoT), and low-latency applications effective congestion management mechanisms have become increasingly critical. Passive queue management strategies such as Drop Tail, which discard packets only after buffer overflow, are known to cause excessive queueing delay, global TCP synchronization, and inefficient bandwidth utilization (Koo et al., 2004; Kochher et al., 2014). Active Queue Management (AQM) was introduced to address these limitations by providing early congestion signals before buffer overflow occurs. The Internet Engineering Task Force (IETF) formally recommended the deployment of AQM mechanisms in RFC 2309 to improve network stability and performance (Braden et al., 1998). Among early AQM algorithms, Random Early Detection (RED) became widely recognized for its probabilistic dropping mechanism based on average queue length, which mitigates synchronization effects and reduces average delay (Floyd & Jacobson, 1993). Weighted Random Early Detection (WRED) subsequently extended RED by enabling differentiated packet dropping across multiple traffic classes, thereby enabling the implementation of DiffServ-based quality-of-service frameworks (Lim et al., 2010; Driediger et al., 2010). Despite its widespread deployment in commercial routers,

WRED exhibits several documented limitations. Its performance is highly sensitive to parameter configuration, and improper tuning may lead to unstable queue behavior or excessive packet loss (Abdrau, 2015; Pujeri et al., 2015). Furthermore, studies report a degradation in fairness and potential starvation of low-priority traffic under sustained congestion (Driediger et al., 2010; Pujeri et al., 2015). These limitations directly motivated the development of alternative AQM paradigms, including control-theoretic algorithms such as PI and REM, rate-based mechanisms such as AVQ, and delay-based solutions such as CoDel and PIE (Kunniyur & Srikant, 2004; Adams, 2013; Giménez et al., 2022).

Recent research has further expanded the AQM landscape through adaptive and nonlinear RED variants, reinforcement learning-based congestion control, and event-triggered queue management mechanisms (Hassan, 2023; Satish, 2024; Toopchinezhad & Ahmadi, 2025). While these developments reflect ongoing innovation, the relative position of WRED within this evolving ecosystem remains insufficiently synthesized. Existing surveys often provide broad overviews of AQM techniques but do not critically evaluate WRED's long-term performance implications, adaptability challenges, and fairness trade-offs within modern congestion control paradigms. Therefore, this study presents a structured literature review of Weighted Random Early Detection within the broader context of Active Queue Management research spanning 2002–2025. The objectives of this review are to: (i) examine the evolution of WRED from its RED foundations; (ii) synthesize comparative performance findings across queue-length-based, control-theoretic, rate-based, delay-based, and adaptive AQM families; (iii) analyze documented limitations related to parameter sensitivity, fairness, and responsiveness; and (iv) identify research gaps and future directions toward adaptive and fairness-aware WRED enhancements. By positioning WRED within the modern AQM landscape and synthesizing findings spanning more than two decades of research, this paper offers a structured, analytical perspective that clarifies WRED's strengths, limitations, and continuing relevance in contemporary network congestion control

Although several surveys have examined Active Queue Management (AQM), most provide broad taxonomic overviews or focus on specific methodological paradigms, such as control-theoretic modelling or machine-learning-based optimization. For instance, Adams (2013) presented a comprehensive classification of AQM mechanisms with emphasis on stability analysis and feedback control design but did not specifically analyze the long-term evolution and differentiated service implications of Weighted Random Early Detection (WRED). More recently, Toopchinezhad and Ahmadi (2025) reviewed learning-based approaches to AQM, focusing primarily on delay and throughput optimization, without systematically evaluating WRED's structural role in policy-driven QoS architectures. This review addresses that gap by focusing on WRED and examining its position within the broader AQM landscape from 2002 to 2025. Rather than treating WRED as a secondary RED variant, the study evaluates how its class-based drop mechanism interacts with adaptive, nonlinear, control-theoretic, delay-based, and learning-driven approaches. In doing so, the review introduces a fairness-differentiation analytical perspective that synthesizes a recurring but fragmented theme in the literature: mechanisms that enforce explicit traffic prioritization may induce fairness asymmetry under sustained congestion, whereas mechanisms that optimize delay stability and global fairness typically lack intrinsic multi-class service differentiation.

Furthermore, this study integrates heterogeneous AQM families within a unified comparative framework that evaluates performance along three dimensions: congestion stability and delay control, fairness and differentiation behaviour, and practical deployment feasibility. By structuring the analysis around these dimensions, the review clarifies WRED's continued relevance in differentiated service environments while identifying the absence of an integrated

design framework capable of simultaneously achieving adaptive stability, preserving fairness, and enforcing explicit multi-class traffic.

II. RELATED WORKS

Research on Active Queue Management (AQM) has evolved significantly since the introduction of Random Early Detection (RED) in the early 1990s. Over time, numerous enhancements and alternative mechanisms have been proposed to improve congestion signalling accuracy, delay control, fairness, and adaptability. This section synthesizes prior work by categorizing AQM developments into major thematic groups while positioning Weighted Random Early Detection (WRED) within this broader research landscape.

A. Queue-Length-Based AQM and the Emergence of WRED

Queue-length-based Active Queue Management (AQM) algorithms represent the foundational stage of proactive congestion control research. Random Early Detection (RED) introduced probabilistic packet dropping based on the exponentially weighted moving average (EWMA) of queue length to mitigate global TCP synchronization and reduce persistent buffer occupancy (Floyd & Jacobson, 1993). By initiating packet drops before buffer overflow, RED aimed to maintain a balance between throughput and delay while improving overall link utilization. However, RED's uniform traffic treatment limits its applicability in networks that require differentiated Quality of Service (QoS). Subsequent analytical and simulation-based evaluations highlighted RED's sensitivity to parameter selection, particularly minimum and maximum thresholds and maximum drop probability (Koo et al., 2004; Kochher et al., 2014). These studies demonstrated that improper configuration could lead to oscillatory queue behaviour, excessive packet loss, or underutilization of available bandwidth. Such limitations motivated further refinement of queue-length-based approaches. Weighted Random Early Detection (WRED) extended RED by associating multiple threshold profiles with distinct traffic classes, typically identified using IP precedence or DiffServ Code Point (DSCP) markings (Lim et al., 2010; Driediger et al., 2010). Through differentiated minimum and maximum thresholds and class-specific drop probabilities, WRED enabled preferential protection of high-priority traffic during congestion. Empirical evaluations confirm that WRED preserves higher throughput and lower packet loss rates for premium traffic classes under moderate load conditions (Pujeri et al., 2015). Its compatibility with DiffServ architectures contributed to widespread adoption in enterprise and service-provider routers.

More recent investigations continue to explore improvements within the queue-length-based paradigm. Lemeshko et al. (2021) proposed enhancements to router-level queue management methods to improve stability and responsiveness in telecommunication networks. Their analysis emphasized the importance of refined threshold adjustment strategies in heterogeneous traffic environments. Similarly, Singha et al. (2021) introduced dynamic weight adjustment mechanisms in RED-based algorithms to reduce packet loss while preserving throughput under bursty traffic conditions. These modifications aim to mitigate some of RED's rigidity without fundamentally altering its queue-length foundation. Additional studies have focused on performance enhancement through structural modifications. Dien (2025) proposed an improved RED mechanism that enhances network performance by adapting the drop probability. Simulation results indicate improved delay stability and reduced packet loss compared to classical RED configurations. Although such approaches demonstrate incremental improvements, they remain inherently dependent on queue-length metrics and thus retain sensitivity to traffic variability and configuration parameters. Collectively, queue-length-based AQMs, including WRED, remain central to congestion control but continue to exhibit

parameter sensitivity and responsiveness limitations. Even modern refinements continue to operate within the same fundamental design paradigm, reinforcing the need for adaptive and hybrid mechanisms that can address evolving network conditions while preserving differentiated service capabilities.

B. Fairness-Oriented and Adaptive RED Variants

The documented limitations of RED and WRED, particularly parameter sensitivity and fairness degradation under sustained congestion, motivated the development of fairness-oriented and adaptive variants aimed at improving stability and equitable resource allocation. Early improvements focused on intra-class fairness among competing flows. Flow Random Early Detection (FRED) introduced per-flow state awareness to prevent aggressive flows from monopolizing buffer resources (Santhi, 2009). By allocating buffer space more equitably, FRED mitigated dominance by high-rate or non-responsive flows. However, maintaining per-flow state introduced scalability concerns in high-speed routers. Subsequent research shifted attention toward adaptive parameter tuning to address the rigidity of static RED/WRED configurations. Adaptive RED (ARED) dynamically adjusts the maximum drop probability based on observed queue behavior to maintain target queue occupancy (Kumar et al., 2023). While ARED improves stability across fluctuating traffic loads, it primarily enhances global queue control and does not inherently resolve WRED's fairness-differentiation trade-off. More advanced adaptive frameworks further refined this idea. Adamu et al. (2021) proposed SARED (Self-Adaptive RED), which dynamically modifies drop probability parameters based on real-time queue dynamics. Simulation results demonstrate reduced parameter sensitivity and improved stability compared to classical RED and WRED. Similarly, Surajo et al. (2023) introduced an exponential self-adaptive RED mechanism designed for next-generation routers, reporting improved throughput consistency and reduced oscillatory queue behavior under heavy load. Nonlinear drop probability modelling represents another important line of development. Rather than relying on RED's linear probability function, Giménez et al. (2022) introduced BetaRED, employing a beta-distribution-based drop function to achieve smoother congestion signalling. This approach reduces abrupt transitions in drop probability and improves delay stability. Building upon this foundation, Giménez et al. (2023) proposed a modified dynamic Beta RED algorithm that further refines probability adaptation to enhance responsiveness under heterogeneous traffic conditions.

Several researchers have also explored hybrid nonlinear models. Hassan et al. (2022) proposed improved RED variants incorporating quadratic and linear probability adjustments to stabilize queue fluctuations. Later, Hassan (2023) introduced AD-RED, integrating quadratic and exponential drop components to improve stability during sustained congestion. Additionally, the Triplex Region RED (TR-RED) algorithm (Hassan, 2023) divides queue occupancy into three operational regions with differentiated drop behaviour, reporting enhanced delay performance and smoother congestion transitions. Decision-based adaptive approaches have also emerged. Mahawish et al. (2022) applied a Markov Decision Process (MDP) framework to RED parameter tuning, enabling dynamic optimization of drop probability decisions based on state transitions. Their results demonstrate improved congestion responsiveness compared to static RED configurations. Similarly, Abdel-jaber et al. (2022) evaluated IM-RED and IGRED using discrete-time queueing models, reporting enhanced stability and reduced packet loss compared with classical RED. Collectively, fairness-oriented and adaptive RED variants demonstrate sustained efforts to mitigate the core weaknesses of queue-length-based AQM mechanisms. While these approaches significantly improve stability, responsiveness, and delay control, they generally prioritize global queue performance rather than differentiated QoS enforcement. Consequently, they do not inherently replicate WRED's multi-class traffic control

capability. Moreover, increasing algorithmic complexity and computational overhead may limit the feasibility of large-scale deployment, particularly in hardware-accelerated router environments.

C. Control-Theoretic and Rate-Based AQM Approaches

As the limitations of queue-length-based mechanisms became increasingly evident, researchers began modelling congestion control as a feedback control problem rather than relying solely on heuristic threshold adjustments. Control-theoretic Active Queue Management (AQM) algorithms apply classical control principles to regulate queue occupancy around a predefined operating point. The Proportional Integral (PI) controller represents one of the earliest control-based AQMs, combining instantaneous queue deviation and accumulated error to compute packet dropping probability (Adams, 2013). Compared to RED, PI-based mechanisms demonstrate improved stability and reduced oscillatory behavior, particularly under dynamic traffic loads and varying round-trip times. Similarly, Random Exponential Marking (REM) introduces a congestion “price” variable reflecting both queue length and link utilization, allowing high link utilization while maintaining controlled queue delay (Adams, 2013). Comparative studies indicate that REM often achieves higher throughput consistency than RED and WRED in high-speed environments. Rate-based AQM mechanisms emerged in parallel to anticipate congestion before excessive buffer buildup occurs. The Adaptive Virtual Queue (AVQ) algorithm maintains a virtual queue operating below the actual link capacity, triggering proactive packet drops when the virtual queue overflows (Kunniyur & Srikant, 2004). This approach enhances responsiveness, particularly in high-bandwidth-delay product networks.

More recent work has extended rate-based and control-oriented approaches. Kar et al. (2021) introduced PAQMAN (Principled Active Queue Management), proposing a theoretically grounded framework for systematic congestion control design. Their model integrates control-theoretic insights with adaptive parameter adjustment, demonstrating improved stability across heterogeneous traffic conditions. Gan (2023) further explored resource-aware congestion management mechanisms, emphasizing dynamic adaptation to varying network resource constraints. The study highlights the importance of coupling congestion signaling with broader resource management strategies in modern networks. Flow-oriented refinements have also been proposed. Yanev et al. (2022) introduced FLOQ (Flow Optimised Queueing), focusing on improved flow-level fairness and congestion responsiveness. Their approach seeks to optimize queue behavior by considering flow distribution characteristics rather than relying solely on aggregate queue metrics. Similarly, Zhu et al. (2022) proposed integrating additional network measurements into AQM decision-making processes, enhancing congestion detection accuracy and reducing unnecessary packet drops under transient conditions.

Control-theoretic and rate-based AQMs demonstrate superior stability and responsiveness compared to classical queue-length-based mechanisms. By incorporating feedback control principles and proactive congestion detection, these approaches reduce oscillatory behavior and improve steady-state utilization. However, despite their improved global performance metrics, they generally lack explicit mechanisms for multi-class traffic differentiation. Unlike WRED, which assigns class-specific thresholds and drop probabilities, these algorithms typically apply uniform congestion control policies across flows. While control-theoretic and rate-based approaches address several stability limitations of RED and WRED, they do not inherently satisfy differentiated Quality of Service (QoS) requirements in enterprise or service-provider environments. This divergence reinforces a recurring theme within the literature: mechanisms that enhance global congestion stability do not automatically preserve policy-driven traffic prioritization.

D. Delay-Based and Learning-Based AQM Mechanisms

The emergence of bufferbloat and latency-sensitive applications prompted a shift from queue-length-based congestion detection toward delay-centric approaches. Unlike RED and WRED, which rely primarily on average queue occupancy as a congestion indicator, delay-based AQM algorithms monitor packet sojourn time to directly regulate queueing delay. Controlled Delay (CoDel) introduced a threshold-based mechanism that drops packets when the minimum observed delay exceeds a predefined target for a specified interval, thereby preventing persistent buffer buildup without extensive parameter tuning (Adams, 2013). The Proportional Integral controller Enhanced (PIE) further refined delay-based congestion control by dynamically adjusting drop probability to maintain target delay levels while minimizing oscillatory behavior. These approaches demonstrate superior latency control compared to RED and WRED under diverse traffic conditions.

Empirical validation studies have further examined delay-based AQMs in real-world scenarios. Morais et al. (2021) applied AQM mechanisms to real-time adaptive video streaming environments, demonstrating that delay-sensitive applications benefit significantly from queue delay regulation compared to classical RED-based approaches. Similarly, Barczyk et al. (2022) conducted real-world evaluations of queue-length-based AQM implementations, highlighting discrepancies between simulation-based findings and operational performance in live networks. Çakmak et al. (2021) compared multiple queue management algorithms within LTE network simulations using NS-3, demonstrating that delay-based mechanisms achieve improved latency stability under wireless conditions. Recent research has extended delay-based concepts using data-driven and measurement-enhanced strategies. Mostafavi et al. (2023) proposed a data-driven delay-violation prediction framework for active queue management, enabling dynamic adjustment of drop behavior based on predicted delay-threshold violations. This approach enhances responsiveness to transient congestion patterns. Zhu et al. (2025) introduced a fine-grained, frequency-count-based AQM mechanism designed to improve congestion-detection precision, demonstrating greater packet-loss reduction and delay stabilization than classical RED. Learning-based AQM mechanisms represent a more substantial conceptual shift. Kim et al. (2021) proposed a deep reinforcement learning (DRL)-based AQM model tailored for IoT network environments. Their approach dynamically adjusts drop probabilities based on learned congestion states, achieving improved throughput and delay performance under heterogeneous traffic patterns. Hasan et al. (2023) examined the interaction between modern TCP BBR congestion control and contemporary AQM schemes, highlighting complex interplay effects on latency and fairness in wireless networks. These findings underscore the growing need to evaluate AQM mechanisms within the context of evolving transport-layer behaviors. Reinforcement learning-based frameworks further demonstrate promising adaptability in non-stationary environments (Satish, 2024). Similarly, machine learning-driven systematic analyses emphasize the potential for intelligent parameter tuning to overcome limitations of static heuristic methods (Toopchinezhad & Ahmadi, 2025). Event-triggered AQM mechanisms also contribute to this evolution by initiating congestion-control actions based on dynamically detected events rather than on continuous monitoring (Wu, 2025). Despite these advancements, delay-based and learning-driven AQMs generally lack intrinsic mechanisms for multi-class traffic differentiation. Congestion signals are applied uniformly unless supplemented by additional scheduling or classification policies. Moreover, learning-based approaches introduce increased computational complexity, training requirements, and scalability concerns that may limit deployment in high-speed hardware routers.

Therefore, while delay-based and intelligent AQM mechanisms significantly enhance latency control, adaptability, and measurement precision, they do not inherently replace WRED in

environments requiring explicit QoS enforcement. This divergence further reinforces WRED's continued relevance in differentiated service architectures, even as congestion control research increasingly emphasizes adaptive and intelligent designs.

E. Identified Research Gaps

The evolution of Active Queue Management from queue-length-based mechanisms to delay driven and learning based frameworks reflects sustained efforts to improve congestion stability, fairness, and responsiveness. However, synthesizing findings from the reviewed literature reveals several persistent, insufficiently addressed gaps, particularly regarding the role of Weighted Random Early Detection (WRED) within modern AQM paradigms.

First, parameter sensitivity remains a fundamental challenge for RED and WRED-based mechanisms. Despite adaptive extensions such as ARED and nonlinear variants like BetaRED and AD-RED, optimal configuration still depends heavily on network-specific characteristics, including round-trip time, link capacity, and traffic heterogeneity (Abdrau, 2015; Hassan, 2023). While adaptive techniques improve stability, their deployment in production environments remains limited, and systematic evaluation across heterogeneous scenarios is sparse. Second, a fairness differentiation trade-off persists in WRED. Although WRED effectively enforces service prioritization, multiple studies document degradation of fairness and potential starvation of low-priority traffic under sustained congestion (Driediger et al., 2010; Pujeri et al., 2015). Modern AQMs focus primarily on stability and delay control, often neglecting explicit quality of service differentiation. Conversely, WRED prioritizes differentiation but lacks explicit mechanisms to prevent starvation. The literature does not provide a comprehensive synthesis evaluating how this trade-off has evolved over time or how it compares with modern adaptive approaches. Third, the limitations in responsiveness of queue-length-based mechanisms remain insufficiently reconciled with modern traffic dynamics. Rate-based and delay-based AQMs demonstrate faster reaction to bursty or short-lived flows (Kunniyur & Srikant, 2004; Adams, 2013), yet their integration with differentiated service frameworks remains underexplored. The coexistence of QoS requirements and low-latency objectives introduces design tensions that have not been systematically analyzed. Fourth, while learning-based and event-triggered AQMs show promising adaptability, their interaction with multi-class traffic differentiation policies has received limited attention (Satish, 2024; Toopchinezhad & Ahmadi, 2025). Most learning-based studies optimize global congestion metrics such as delay or throughput without evaluating differentiated service guarantees. As a result, it remains unclear whether modern adaptive frameworks can fully replace WRED in QoS-sensitive deployments. Finally, the literature exhibits a fragmented evaluation methodology. Many studies rely on simulation-based experiments with limited real-world validation. Performance comparisons often differ in traffic models, topology configurations, and evaluation metrics, complicating cross-study synthesis. A structured and comparative review that consolidates findings across algorithm families and identifies consistent performance patterns remains necessary.

These gaps collectively motivate the need for a structured literature review that positions WRED within the broader evolution of AQM research, critically evaluates its strengths and limitations relative to modern alternatives, and identifies pathways for adaptive and fairness-aware enhancement. By synthesizing developments spanning 2002–2025, this study seeks to provide a consolidated analytical perspective. To provide a consolidated overview of representative studies across AQM families, Table 1 summarizes key contributions, methodologies, and documented limitations of selected works published between 2002 and 2025.

Table 1: Comparative Summary of Related Work

Authors / Year	Research Area	Methodology	Key Findings	Performance	Limitations
Floyd & Jacobson (1993)	Foundational RED	Analytical modelling & simulation	Reduced TCP global synchronization; lower average queue length vs Drop Tail		High sensitivity to min/max thresholds
Kunniyur & Srikant (2004)	Rate-based AQM (AVQ)	Fluid-flow modelling & simulation	Faster congestion anticipation; improved queue stability in high BDP networks		Implementation and parameter calibration complexity
Santhi (2009)	Fairness-aware RED (FRED)	Simulation	Improved fairness; reduced dominance of aggressive flows	per-flow	state limits overhead scalability
Lim et al. (2010)	WRED analytical modelling	Markov-modulated queue analysis	Effective service differentiation across traffic classes		Performance degradation under bursty traffic
Driediger et al. (2010)	Fair WRED	Mixed TCP/UDP simulation	Improved fairness index compared to classical WRED		Increased computational complexity
Abdrau (2015)	WRED sensitivity analysis	IP network simulation	Strong dependence on threshold spacing and drop probability parameters		Requires manual tuning; unstable under misconfiguration
Pujeri et al. (2015)	RED vs WRED comparison	High-speed network simulation	Higher throughput preservation for priority traffic in WRED		Starvation of low-priority flows under sustained congestion
Adamu et al. (2021)	Adaptive RED (SARED)	Dynamic simulation	Reduced parameter sensitivity; improved queue stability vs static RED		Additional computational overhead
Singha et al. (2021)	Dynamic-weight RED	Simulation	Reduced packet loss and improved throughput under bursty load		Still dependent on queue-length metrics
Lemeshko et al. (2021)	Router-level AQM refinement	Analytical & simulation	Improved responsiveness and threshold adaptation		Limited real-world validation
Kim et al. (2021)	Deep RL-based AQM	Reinforcement learning & simulation	Improved delay control and throughput under heterogeneous traffic		High computational and training cost
Mahawish et al. (2022)	MDP-based RED	Decision-process modelling	Optimized drop probability adaptation; improved congestion responsiveness		Model complexity and state dependency
Abdel-jaber et al. (2022)	IM-RED / IGRED	Discrete-time queue modelling	Reduced oscillation amplitude and packet loss vs classical RED		Limited evaluation scenarios
Giménez et al. (2022)	BetaRED	NS-3 simulation	Reduced delay variance; smoother drop probability transitions		No intrinsic traffic class differentiation
Hassan (2023)	AD-RED	High-speed simulation	Improved stability under heavy load; reduced queue fluctuation		Requires careful parameter calibration
Hassan et al. (2023)	TR-RED	Region-based drop modelling	Smoother congestion transitions; improved delay consistency		Increased algorithmic complexity

Kar et al. (2021)	PAQMAN (control-based AQM)	Control-theoretic modelling	Improved stability across heterogeneous traffic	Lacks explicit QoS differentiation
Yanev et al. (2022)	FLOQ (flow-optimized queueing)	Simulation	Enhanced flow-level fairness	Limited support for traffic class prioritization
Mostafavi et al. (2023)	Data-driven delay prediction AQM	Predictive modelling	Improved delay violation control	Requires accurate prediction models
Zhu et al. (2025)	Frequency-based AQM	Fine-grained congestion detection	Reduced packet loss and improved delay stabilization	Limited deployment validation

III. METHODOLOGY

A. Research Design and Review Approach

This study adopts a structured literature review methodology to systematically examine the evolution and performance characteristics of Weighted Random Early Detection (WRED) within the broader context of Active Queue Management (AQM) research. A structured review approach was selected to ensure transparency, reproducibility, and analytical rigor in synthesizing findings across heterogeneous studies spanning theoretical modeling, simulation-based evaluation, and adaptive algorithm design. The review focuses on peer-reviewed publications addressing RED, WRED, and related AQM mechanisms published between 2002 and 2025. The starting year was selected to capture early analytical evaluations and DiffServ-based implementations following widespread RED adoption, while the end year reflects recent developments in nonlinear, adaptive, delay-based, and learning-driven AQM mechanisms. Unlike narrative reviews that provide descriptive summaries, this structured approach emphasizes systematic identification, screening, categorization, and comparative synthesis of relevant studies. The methodology integrates thematic classification of AQM families, performance metric comparison, and identification of recurring limitations and research gaps. By combining foundational historical studies with recent adaptive and machine-learning-based approaches, the review aims to provide a longitudinal perspective on WRED’s role in modern congestion control research.

B. Data Sources and Search Strategy

A systematic search was conducted to identify peer-reviewed publications relevant to Weighted Random Early Detection (WRED) and related Active Queue Management (AQM) mechanisms. The search covered major academic databases, including IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and Google Scholar, to ensure broad and multidisciplinary coverage of networking research. The review timeframe was restricted to publications from January 2002 to March 2025 to capture post-RED developments, WRED-based DiffServ implementations, adaptive RED extensions, delay-based AQM mechanisms, and recent learning-driven congestion control approaches.

The search strategy employed combinations of keywords such as “Weighted Random Early Detection,” “WRED,” “Random Early Detection,” “Active Queue Management,” “AQM congestion control,” “DiffServ queue management,” “Adaptive RED,” “delay-based AQM,” and “machine learning AQM.” Boolean operators were used to refine results and improve specificity (e.g., “WRED AND fairness,” “AQM AND delay control,” “RED variant AND stability”). Only peer-reviewed journal articles and conference proceedings within the networking and communication systems domains were considered to ensure academic rigor

and relevance. The initial database query yielded 168 records. After removing duplicate entries and publications unrelated to network congestion control, 132 records remained for further screening based on title and abstract relevance.

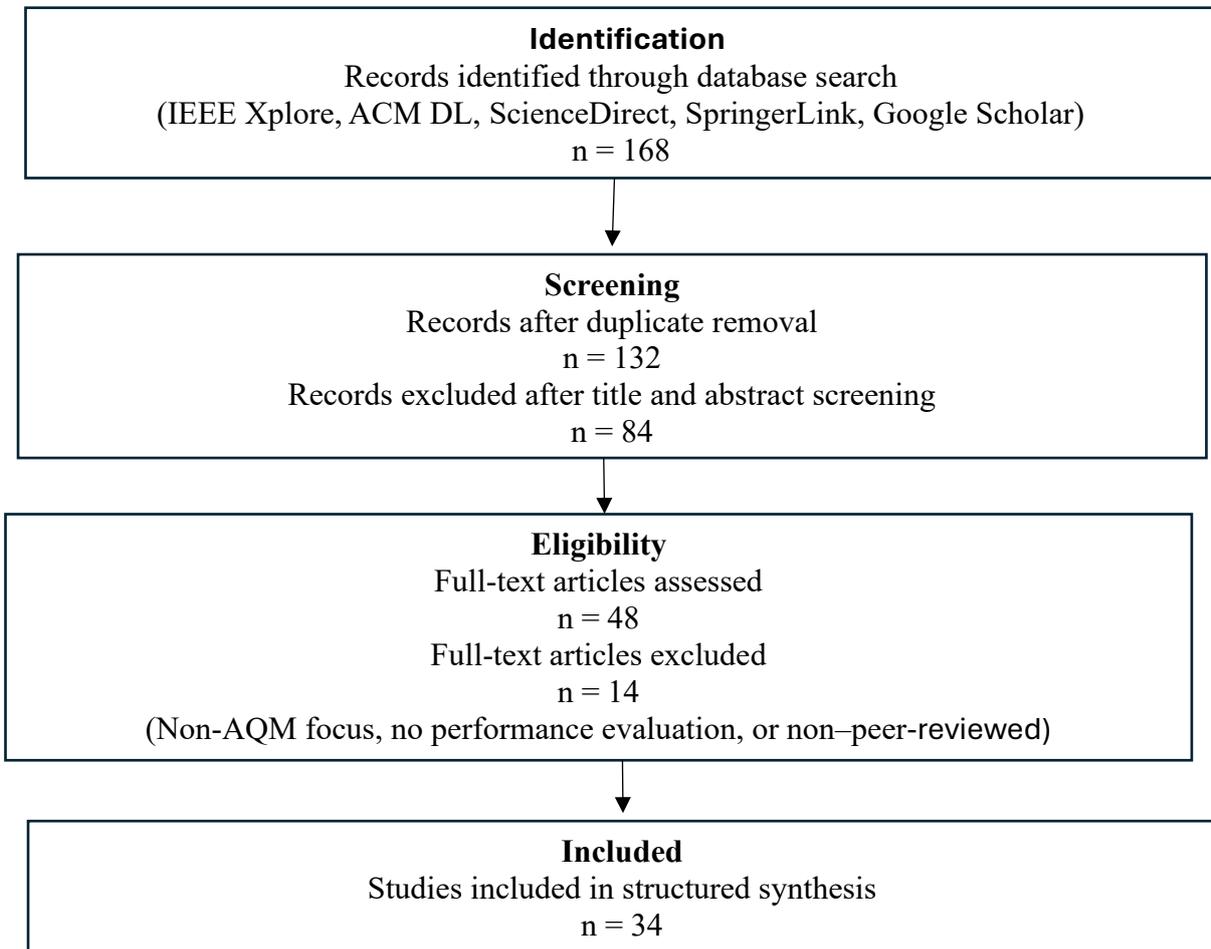


Figure 3.1 Study Identification and Screening Process

C. Study Selection and Eligibility Criteria

Following the initial screening of titles and abstracts, the remaining publications were evaluated using predefined inclusion and exclusion criteria to ensure relevance and analytical consistency. Studies were included if they (i) directly addressed Weighted Random Early Detection (WRED), Random Early Detection (RED), or related Active Queue Management (AQM) mechanisms; (ii) presented analytical models, simulation-based evaluations, experimental validation, or comparative performance analysis; and (iii) reported measurable performance outcomes such as throughput, queueing delay, packet loss, fairness metrics, or stability behavior. Both foundational studies and recent advancements, including adaptive, nonlinear, delay-based, and learning-driven AQM mechanisms, were considered in order to provide longitudinal coverage of developments between 2002 and 2025.

Publications were excluded if they (i) focused solely on transport-layer congestion control without router-level queue management relevance, (ii) addressed congestion in non-network contexts (e.g., transportation systems), (iii) lacked peer-review validation, or (iv) provided purely descriptive discussion without performance analysis or technical evaluation. Additionally, duplicate studies reporting identical experimental results were consolidated to avoid overrepresentation. After applying these criteria and conducting full-text assessments, 34 studies were identified as highly relevant and suitable for detailed synthesis. These selected publications form the analytical foundation of this review, while additional works were referenced where necessary to provide contextual background and support thematic categorization.

D. Data Extraction and Analytical Framework

For each selected study, relevant information was systematically extracted to enable structured comparison and thematic synthesis. The extracted data included algorithm type (e.g., WRED, adaptive RED variant, delay-based AQM, learning-based AQM), congestion indicators used (queue length, delay, rate, or hybrid metrics), evaluation methodology (analytical modeling, simulation, experimental testbed), traffic scenarios, and reported performance outcomes. Particular attention was given to quantitative metrics, including throughput, average queueing delay, packet loss rate, fairness indices, stability characteristics, and responsiveness under varying load conditions. To facilitate comparative analysis, the selected studies were categorized into five principal AQM families: (i) queue-length-based mechanisms, including RED and WRED; (ii) adaptive and nonlinear RED variants; (iii) control-theoretic and rate-based approaches; (iv) delay-based mechanisms; and (v) learning-based and event-triggered frameworks. This classification enabled structured evaluation of WRED's relative strengths and limitations across evolving design paradigms.

The analytical framework focused on identifying recurring performance patterns and trade-offs rather than replicating individual experimental results. Emphasis was placed on synthesizing findings related to parameter sensitivity, fairness–differentiation trade-offs, delay stability, adaptability, and deployment feasibility. Where available, reported performance improvements were compared qualitatively across algorithm families to highlight consistent trends and unresolved challenges. By applying this structured extraction and classification approach, the review moves beyond descriptive summary and provides an integrated perspective on the evolution of WRED within modern Active Queue Management research.

E. Quality Assessment and Study Evaluation Criteria

To enhance methodological rigor, the selected studies were evaluated using structured qualitative assessment criteria adapted for engineering and networking systematic reviews. Each study was examined across four core dimensions. The First core is methodological validity was assessed based on the clarity and soundness of the analytical modelling, simulation design, or experimental testbed implementation. Studies were expected to clearly describe algorithm configuration, parameter settings, and experimental procedures. Secondly, the depth of performance evaluation was considered, including the range and relevance of reported metrics such as throughput, queueing delay, packet loss rate, fairness indices, stability characteristics, and responsiveness under varying load conditions. Thirdly, we focused on scenario diversity to evaluate whether the study examined heterogeneous traffic conditions, varying congestion levels, or multiple network configurations, thereby strengthening the generalizability of findings. Finally, reproducibility and transparency were assessed based on the clarity of simulation environments (e.g., NS-2, NS-3), topology descriptions, and parameter specifications needed to replicate the reported results.

Studies lacking measurable performance analysis, insufficient methodological clarity, or limited experimental transparency were excluded during full-text assessment. Due to methodological heterogeneity across AQM studies, a numerical scoring system was not applied. Instead, qualitative consistency across these dimensions guided inclusion and comparative synthesis.

F. Bias Considerations and Analytical Limitations

The reviewed studies exhibit considerable methodological diversity, including differences in simulation environments, traffic models, topology configurations, parameter settings, and evaluation metrics. Due to this heterogeneity, direct quantitative meta-analysis or statistical weighting across studies was not feasible. Instead, this review adopts a structured qualitative synthesis approach to identify consistent patterns and recurring trade-offs across algorithm families. To reduce selection bias, multiple academic databases were systematically searched, and clearly defined inclusion and exclusion criteria were applied throughout the screening process. Only peer-reviewed journal articles and conference proceedings were considered to ensure academic reliability and methodological soundness. Nevertheless, certain limitations should be acknowledged. As with most technical literature, studies reporting performance improvements are more likely to be published, which may introduce publication bias. In addition, the predominance of simulation-based evaluations in AQM research may limit the generalizability of findings to large-scale operational networks. Because experimental setups vary substantially across studies, formal statistical aggregation or weighted comparison was not performed. Instead, the analysis emphasizes converging evidence and cross-study consistency to support balanced interpretation.

IV. RESULTS AND DISCUSSION

A. Performance Trends Across AQM Families

The structured synthesis of the selected studies reveals that performance improvements across Active Queue Management (AQM) mechanisms are highly context-dependent, varying according to traffic composition, network scale, and congestion intensity. Rather than identifying a universally superior algorithm, the comparative analysis highlights distinct performance trade-offs among queue-length-based, adaptive, delay-based, and learning-driven approaches. Queue-length-based mechanisms such as RED and WRED exhibit stable throughput under moderate congestion when appropriately configured. WRED, in particular, consistently preserves higher throughput for priority traffic classes by assigning differentiated drop thresholds. However, overall delay performance is strongly influenced by threshold spacing and maximum drop probability configuration. Under suboptimal tuning, queue oscillations and excessive packet loss may occur. Adaptive and nonlinear RED variants, including SARED, BetaRED, and AD-RED, achieve measurable improvements in queue stability and reduced delay variance compared to static RED/WRED implementations. By dynamically adjusting drop probabilities or reshaping the drop function, these mechanisms mitigate abrupt congestion transitions. Nevertheless, while they improve stability, they typically do not embed explicit multi-class differentiation policies equivalent to WRED. Control-theoretic and rate-based AQMs, such as PI, REM, and AVQ, consistently demonstrate improved responsiveness and stability under fluctuating load conditions. Their feedback-based design reduces queue oscillation and enhances utilization consistency. However, these approaches primarily optimize global congestion metrics and lack inherent QoS differentiation capabilities. Delay-based mechanisms such as CoDel and PIE achieve superior latency control across diverse network conditions by directly regulating packet sojourn time. Their minimal

parameter dependency enhances deployment robustness. Yet, similar to control-based approaches, delay-centric AQMs do not inherently support class-based traffic prioritization. Recent learning-based and event-triggered AQMs demonstrate promising adaptability, particularly under heterogeneous and dynamic traffic patterns. Reinforcement learning models adjust drop policies based on observed congestion states, potentially outperforming static heuristics. However, these mechanisms introduce increased computational overhead and limited validation in real-world deployments.

Overall, the synthesis indicates that while modern AQM approaches outperform classical WRED in delay stability and adaptability, WRED retains a unique advantage in explicit service differentiation. This distinction suggests that performance evaluation should consider both global congestion metrics and differentiated QoS objectives, rather than relying on a single optimization criterion.

B. Fairness–Differentiation Trade-Off in WRED

One of the most persistent and conceptually significant findings across the reviewed literature is the fairness–differentiation trade-off inherent in Weighted Random Early Detection (WRED). While WRED was explicitly designed to enforce Quality of Service (QoS) differentiation through class-based drop thresholds, this prioritization mechanism inherently redistributes congestion penalties across traffic classes. Empirical evaluations consistently demonstrate that WRED effectively protects high-priority traffic by assigning larger minimum and maximum thresholds, thereby reducing drop probability for premium flows. Under moderate congestion, this mechanism maintains higher throughput and lower packet loss for prioritized traffic classes compared to standard RED configurations (Lim et al., 2010; Pujeri et al., 2015). In enterprise and service-provider networks that require differentiated services, this behavior is advantageous and aligns with policy-driven traffic engineering objectives. However, under sustained congestion conditions, lower-priority traffic may experience disproportionately high packet loss rates. Repeated loss events can trigger successive congestion window reductions in TCP flows, potentially leading to throughput collapse or starvation effects (Driediger et al., 2010). Unlike fairness-oriented mechanisms such as FRED, WRED does not track per-flow buffer occupancy and therefore cannot prevent aggressive flows or priority-based penalization from dominating queue resources.

Adaptive variants, such as SARED, aim to mitigate instability by dynamically adjusting drop probabilities based on observed queue conditions (Adamu et al., 2021). While such approaches improve overall stability, they do not fundamentally eliminate prioritization-induced asymmetry. Similarly, nonlinear variants like BetaRED and AD-RED improve smoothness in congestion signaling but maintain a global drop structure rather than class-aware fairness correction. In contrast, delay-based and control-theoretic AQMs distribute congestion signals more uniformly across flows. Although this enhances fairness metrics, it removes the explicit enforcement of differentiated service guarantees. Thus, the literature reveals a structural divergence: mechanisms that optimize fairness and delay stability often sacrifice policy-driven prioritization, whereas WRED enforces differentiation at the cost of potential fairness degradation.

This trade-off reflects a deeper design tension between equity and policy enforcement in congestion control. The reviewed studies do not provide a unified framework that integrates adaptive stability control with explicit class-based prioritization while preventing starvation under heavy load. Consequently, WRED remains both uniquely valuable for QoS-aware deployments and simultaneously limited by its fairness vulnerabilities. Addressing this trade-off represents a key opportunity for future research, particularly in developing hybrid

mechanisms that combine adaptive drop tuning, flow-awareness, and differentiated service enforcement within a unified AQM architecture.

C. Practical Deployment and Scalability Considerations

Beyond theoretical performance improvements, the practical feasibility of Active Queue Management mechanisms plays a critical role in determining their real-world adoption. The reviewed literature indicates that although numerous adaptive, nonlinear, and learning-based AQMs demonstrate improved stability and responsiveness under simulated conditions, large-scale deployment remains limited. Classical WRED maintains a distinct advantage in practical deployment due to its simplicity and compatibility with commercial routing platforms. Most enterprise and service-provider routers natively support WRED configuration within DiffServ-based QoS frameworks. Its implementation relies primarily on threshold-based probabilistic dropping, which imposes minimal computational overhead. However, this simplicity comes at the cost of manual configuration complexity. Network administrators must carefully tune minimum and maximum thresholds, maximum drop probabilities, and class-specific profiles to achieve optimal performance. Misconfiguration can lead to unstable queue behavior or excessive packet loss.

Adaptive variants such as SARED reduce configuration sensitivity by dynamically adjusting drop probabilities in response to queue dynamics (Adamu et al., 2021). While this enhances robustness, it introduces additional computational logic that may increase processing overhead in high-speed router environments. Similarly, nonlinear variants such as BetaRED and AD-RED require additional mathematical computation for drop probability evaluation, potentially affecting scalability in hardware-accelerated implementations. Control-theoretic and rate-based AQMs offer improved stability but often require precise parameter calibration and may depend on accurate traffic measurement, which introduces monitoring overhead. Delay-based mechanisms such as CoDel and PIE demonstrate strong deployment robustness due to reduced parameter dependence; however, they do not natively integrate with class-based differentiation policies without additional scheduling mechanisms. Learning-based and reinforcement learning-driven AQMs represent the most computationally intensive category. Although they show promising adaptability to dynamic and heterogeneous traffic patterns, their reliance on state estimation, training phases, and iterative policy updates raises concerns about scalability and the feasibility of hardware implementation. Real-world validation in high-speed backbone routers remains limited in the literature (Satish, 2024; Toopchinezhad & Ahmadi, 2025). Furthermore, the majority of reviewed studies rely heavily on simulation-based validation using tools such as NS-2 and NS-3. While simulation enables controlled experimentation and reproducibility, it may not fully capture real-world traffic burstiness, long-range dependence, or multi-bottleneck network interactions. Experimental deployment studies are comparatively scarce, limiting definitive conclusions regarding large-scale operational performance.

Overall, practical deployment considerations reveal a critical tension between algorithmic sophistication and implementation feasibility. While modern adaptive and learning-based AQMs demonstrate theoretical performance advantages, WRED remains one of the most widely deployable mechanisms due to its computational simplicity and integration within established quality of service frameworks. Bridging the gap between adaptive intelligence and deployment scalability represents an essential direction for future congestion control research.

V. CONCLUSION

This study presented a structured literature review of Weighted Random Early Detection (WRED) within the broader evolution of Active Queue Management research from 2002 to 2025. The review examined WRED's development, compared it with adaptive, control-

theoretic, delay-based, and learning-driven mechanisms, and identified persistent performance trade-offs. The evidence indicates that WRED remains uniquely suited for explicit service differentiation in QoS-aware networks. Its class-based threshold configuration effectively protects high-priority traffic under moderate congestion. However, WRED inherits notable limitations, including parameter sensitivity and fairness degradation under sustained load. Contemporary AQM mechanisms demonstrate superior delay stability and adaptability but generally lack intrinsic support for differentiated traffic enforcement. A key finding is the absence of an integrated framework that simultaneously ensures adaptive stability, fairness preservation, and explicit multi-class differentiation. Future research should focus on hybrid AQM designs that combine adaptive parameter tuning, flow-awareness, and policy-driven prioritization while maintaining deployment feasibility. By synthesizing over two decades of research, this review clarifies WRED's continuing relevance and outlines directions for next-generation congestion management mechanisms.

REFERENCES

- [1] Abdel-jaber, H., Mahafzah, M., & Alshraideh, M. (2022). Evaluation of the performance for IM-RED and IGRED algorithms using discrete-time queues. *Proceedings of the 14th International Conference on Computational Intelligence and Communication Networks (CICN)*, 1–6. <https://doi.org/10.1109/CICN56167.2022.10008318>
- [2] Abdrau, M. (2015). Performance analysis of weighted random early detection in IP-based networks. *International Journal of Computer Networks & Communications*, 7(4), 15–28.
- [3] Adamu, A., Surajo, Y., & Jafar, M. T. (2021). SARED: Self-adaptive active queue management scheme for improving quality of service in network systems. *Computer Science*, 22(2). <https://doi.org/10.7494/csci.2021.22.2.4020>
- [4] Adams, R. (2013). Active queue management: A survey. *IEEE Communications Surveys & Tutorials*, 15(3), 1425–1476.
- [5] Barczyk, A., Burakowski, T., & Tarasiuk, A. (2022). AQM based on the queue length: A real-network study. *PLOS ONE*, 17(2), e0263407. <https://doi.org/10.1371/journal.pone.0263407>
- [6] Braden, B., Clark, D., Crowcroft, J., et al. (1998). Recommendations on queue management and congestion avoidance in the Internet (RFC 2309). IETF.
- [7] Çakmak, E., Ertürk, M., & Aydin, M. A. (2021). Performance comparison of queue management algorithms in LTE networks using NS-3 simulator. *Tehnicki Vjesnik—Technical Gazette*, 28(6), 1915–1923.
- [8] Dien, N. T. (2025). Enhancing network performance with improved random early detection. *Vinh University Journal of Science*, 53(4A), 83–95.
- [9] Driediger, T., Srikant, R., & Vandalore, B. (2010). Fair weighted random early detection. *IEEE/ACM Transactions on Networking*, 18(1), 109–122.
- [10] Floyd, S., & Jacobson, V. (1993). Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4), 397–413.
- [11] Giménez, J. L., Ferrández, J. M., & Berral, J. L. (2022). New RED-type TCP-AQM algorithms based on beta distribution drop functions. *Applied Sciences*, 12(21), 11176.
- [12] Giménez, J. L., Ferrández, J. M., & Berral, J. L. (2023). Modified dynamic beta RED—A new AQM algorithm for Internet congestion control. *Proceedings of the ACM Internet Measurement Conference*.
- [13] Hassan, A. A. (2023). AD-RED: A new variant of random early detection AQM algorithm. *Journal of High Speed Networks*, 29(3), 163–177.

- [14] Hassan, A. A., Shah, W. M., Othman, M. F., & Abdullah, A. A. (2023). A triplex region–random early detection (TR-RED) algorithm for active queue management in Internet routers. *International Journal of Computing and Digital Systems*, 14(1), 217–228.
- [15] Kar, K., Sarkar, S., & Tassioulas, L. (2021). PAQMAN: A principled approach to active queue management. *arXiv preprint*.
- [16] Kim, T., Pack, S., & Lee, W. (2021). Deep reinforcement learning-based active queue management for IoT networks. *Journal of Network and Systems Management*, 29, 30.
- [17] Kochher, P., Singh, G., & Sharma, R. (2014). Comparative performance evaluation of AQM algorithms in high-speed networks. *International Journal of Computer Applications*, 95(3), 12–18.
- [18] Koo, J., Kim, H., & Choi, M. (2004). Performance evaluation of RED under varying traffic conditions. *IEICE Transactions on Communications*, E87-B(10), 2896–2903.
- [19] Kumar, S., Sharma, R., & Patel, D. (2023). Adaptive random early detection: Stability improvements and analysis. *Journal of Communications Software and Systems*, 19(2), 77–89.
- [20] Kunniyur, S., & Srikant, R. (2004). Analysis and design of an adaptive virtual queue (AVQ) algorithm. *IEEE/ACM Transactions on Networking*, 12(2), 286–299.
- [21] Lemeshko, O., Yeremenko, O., & Tariki, N. (2021). Development and research of active queue management method on interfaces of telecommunication networks routers. In *Lecture Notes in Computational Intelligence and Decision Making* (pp. 3–23). Springer.
- [22] Lim, W., Kim, J., & Park, H. (2010). Analytical modeling of weighted random early detection with Markov-modulated traffic. *Computer Networks*, 54(14), 2431–2445.
- [23] Mahawish, S., Khan, A. S., Khan, M. A., & Hussain, S. J. (2022). Improving RED algorithm congestion control by using the Markov decision process. *Scientific Reports*, 12, 13050.
- [24] Mostafavi, S., Modiano, E., & Shah, D. (2023). Active queue management with data-driven delay violation probability predictors. *arXiv preprint*.
- [25] Morais, N., Pedroso, H., & Corujo, D. (2021). Application of active queue management for real-time adaptive video streaming. *Telecommunication Systems*, 79, 285–302.
- [26] Pujeri, U., Ramesh, B., & Rao, M. (2015). Performance comparison of RED and WRED in high-speed networks. *International Journal of Advanced Networking and Applications*, 7(1), 2583–2589.
- [27] Santhi, V. (2009). Flow random early detection for congestion avoidance. *International Journal of Computer Science and Network Security*, 9(5), 137–143.
- [28] Satish, D. (2024). Reinforcement learning-based active queue management for L4S environments. *Future Internet*, 16(8), 265.
- [29] Singha, R., Deb, S., & Saha, A. K. (2021). Application of dynamic weight with distance to reduce packet loss in RED-based algorithm. *International Journal of Computer Applications*.
- [30] Toopchinezhad, M. P., & Ahmadi, M. (2025). Machine learning approaches for active queue management: A systematic survey. *Computer Networks*, 240, 110174.
- [31] Wu, D. M. (2025). Research on event-triggered active queue management mechanisms. *Journal of Communications*, 20(2), 115–128.
- [32] Yanev, N., Radev, P., & Iliev, I. (2022). Herding the FLOQ: Flow optimised queueing. *IFIP Networking Conference Proceedings*.
- [33] Zhu, Y., Xu, D., & Xu, M. (2022). Introducing additional network measurements into active queue management. *IEEE International Conference on Communications (ICC)*.
- [34] Zhu, Y., Xu, D., & Xu, M. (2025). A fine-grained active queue management based on frequency counting. *SSRN Electronic Journal*.