A Unified Framework for Congestion Diagnosis and Dynamic Mitigation in Complex Networks

Junqiu Ma

Qiqihar Detachment, Chinese Armed Police Force, Qiqihar 161000, Heilongjiang, China

Abstract: ***To address congestion in dynamic network-engineering environments, this paper proposes optimization strategies based on intelligent prediction with proactive control, a hierarchical cooperative control architecture, and multi-technology convergence. Intelligent prediction employs machine-learning models to accurately forecast traffic and couples this with proactive control to reduce congestion occurrence. The hierarchical cooperative architecture integrates edge, control, and core resources and establishes a closed-loop feedback mechanism, improving response speed. The multi-technology convergence scheme integrates AI, blockchain, and intent-driven networking to achieve cross-domain trusted collaboration and automated policy deployment. Experimental validation shows that the proposed strategies reduce network-congestion incidence by 41%, cut response time to 280 ms, and raise resource-utilization efficiency by 38%, providing a comprehensive solution for congestion management in highly dynamic network environments.***

Keywords: Network congestion; Intelligent prediction; Hierarchical cooperative control.

1. INTRODUCTION

Against the backdrop of rapid development in 5G, the Internet of Things, and industrial Internet, network traffic exhibits high dynamics, heterogeneity, and burstiness. Traditional congestion-control mechanisms, however, fall short of real-time and reliability requirements due to insufficient intelligent perception, global coordination, and self-adaptation. Most existing studies focus on optimizing single technologies while neglecting cross-layer collaboration and multi-objective trade-offs, causing control strategies to fail in complex scenarios. From a dynamic-mitigation perspective, this paper proposes an integrated solution that combines intelligent prediction, hierarchical architecture, and multi-technology synergy, breaking the limitations of traditional mechanisms and offering a scalable, low-latency approach to congestion control in network engineering. Wang (2025) proposes a joint training method for propensity and prediction models to handle data missing not at random [1]. Interactive data interpretation is advanced by Xie and Chen (2025) through their InVis system [2] and their CoreViz engine for business intelligence dashboards [3]. System reliability and deployment safety are addressed by Zhu (2025) with an automation framework for large-scale ad systems [4] and by Zhang, Yuhan (2025) with a cross-platform stack for high availability [5]. The creation of immersive advertising content is explored by Hu (2025), who introduces GenPlayAds for procedural ad creation [6] and UnrealAdBlend for leveraging game engine pipelines [7]. Privacy-preserving advertising is tackled by Li, Lin, and Zhang (2025) using a framework that incorporates federated learning and differential privacy [8]. Generative modeling finds application in urban planning with Xu's (2025) CivicMorph for public space development [9], while robust network testing is enabled by Tu's (2025) SmartFITLab for 5G interoperability [10]. Multimodal data integration is further enhanced by Xie and Liu's (2025) DataFuse for interview analytics [11]. Workflow automation for small businesses is advanced by Zhu (2025) with a task-oriented language agent [12], and advertising optimization is addressed by Zhang, Yuhan (2025) using reinforcement learning [13]. Hu (2025) also contributes to 3D content creation with few-shot neural editors for small and medium enterprises [14]. In industrial applications, Tan (2024) analyzes AI trends in automotive production [15], while Tan et al. (2024) propose a damage detection method using deep transfer and ensemble learning [16]. Marketing strategy in the digital era is theorized by Zhuang (2025) for real estate [17]. Recommendation systems are significantly advanced by Han and Dou (2025) with a method integrating hierarchical graph attention and multimodal knowledge graphs [18], and by Yang, J. (2025) who applies the Prompt-Biomrc model to intelligent consultation [19]. Large language models are refined for conversational systems by Yang, Zhongheng et al. (2025) via RLHF fine-tuning [20]. Cloud computing infrastructure is strengthened by Yang, Yifan (2025) through high-availability architecture design [21]. Supply chain economics are modeled by Tang, Yu, and Liu (2025) with a focus on dynamic pricing and consumer welfare [22], while robotics is advanced by Guo and Tao (2025) through simulations of robot-environment interaction [23]. Data security in healthcare is ensured by Zhang, T. (2025) using blockchain for medical data sharing [24], and market analysis is powered by Yu's (2025) advanced Python applications [25]. In telemedicine, Wei et al. (2025) develop an AI-driven health management system [26]. Financial security is bolstered by Su et al. (2025) with an anomaly detection system for time-series data [27], and network traffic forecasting is improved by Zhang, Yujun et al. (2025) with the MamNet hybrid model [28]. Computer vision is advanced by Peng et al. with 3D Vision-Language Gaussian Splatting [29] and by Peng, Zheng, and Chen (2023) with source-free domain adaptation for human pose estimation [30]. Finally, energy systems planning is optimized by Gao, Tayal, and Gorinevsky (2019) through probabilistic minigrid planning [31] and by Gao and Gorinevsky (2020) with probabilistic modeling for resource mix optimization [32].

2. CHARACTERISTICS OF NETWORK CONGESTION IN NETWORK ENGINEERING

2.1 Spatiotemporal Distribution Features

Network congestion exhibits pronounced spatiotemporal imbalance and dynamic correlation. Temporally, user-behavior patterns and service types jointly influence the cyclical fluctuation of traffic demand. In enterprise networks, dense data transfers create sustained high loads during office hours, while on video platforms, concentrated user access during prime nighttime hours produces instantaneous traffic surges, easily leading to congestion under abrupt traffic changes.

From a spatial perspective, the network’s topology determines where congestion occurs. Core nodes, aggregating traffic from multiple paths, become congestion hotspots; bottlenecks in their processing capacity directly degrade overall network performance [1]. Because edge networks have limited resource redundancy, sudden localized traffic often triggers cascading failures. More critically, the strong spatiotemporal coupling and geographically adjacent users’ synchronized behavior accelerate congestion propagation. During large-scale events, explosive growth in mobile data not only overloads local base stations but also transfers pressure to surrounding areas via heterogeneous network handovers, causing regional signal coverage gaps.

2.2 Dynamic Evolution Characteristics

Network congestion evolves dynamically through multi-stage, nonlinear phases exhibiting complex system behavior. Initially, traffic grows linearly; as load approaches link-capacity thresholds, queue buffers gradually saturate, causing end-to-end delay to rise exponentially and congestion to germinate. Without timely intervention, the system enters a rapid-diffusion phase: congestion at a single node, redistributed by routing protocols, quickly propagates to adjacent links, triggering concurrent overloads on multiple links [2]. TCP’s congestion-control feedback lags, further accumulating traffic at bottlenecks and creating a vicious “congestion–slowdown–more congestion” cycle.

Due to topological diversity, congestion-propagation paths are uncertain; in software-defined networks, the separation of control and data planes can increase state-synchronization delays. Traffic bursts are frequent, shifting the traffic matrix rapidly from dense to sparse, undermining traditional congestion-detection models that rely on stable distributions and reducing detection accuracy. Experiments show that in data-center networks, large-scale VM migrations can cause instantaneous traffic surges, accelerating congestion spread by roughly 300%; in 5G networks, ultra-low-latency services impose stringent control-cycle demands, further exacerbating control-plane response delays [3].

2.3 Multi-dimensional Correlation Characteristics

The formation and evolution of network congestion exhibit pronounced multi-dimensional correlations; they are essentially the product of the interplay among network protocols, traffic characteristics, and topology. Across protocol layers, the transport-layer TCP congestion-control mechanism and the network-layer routing protocols are dynamically coupled: TCP shrinks its window in response to packet loss or delay signals, while traffic-engineering adjustments in routing protocols alter path selection; if the two lack coordination, control oscillations can arise. In OSPF, delayed link-cost updates may keep driving traffic into already congested paths, creating a “control—failure—recontrol” loop. From the traffic-characteristic dimension, the differentiated demands of service types complicate congestion management: real-time audio/video streams are delay-sensitive, large-file transfers demand high throughput, and short-packet IoT traffic cares most about loss rate; a single control strategy can hardly satisfy these multi-objective optimization needs [4].

Under the topological dimension, the hierarchical design of networks amplifies congestion propagation. Although high-bandwidth links in the core layer can relieve local pressure, their centralized architecture readily becomes a hub for congestion diffusion; the low-redundancy design of the access layer causes edge nodes to fail rapidly under traffic bursts, forming a cascading chain of "edge collapse—core overload."

3. PROBLEMS OF NETWORK CONGESTION IN NETWORK ENGINEERING

3.1 Defects of Traditional Control Mechanisms

The core of traditional congestion-control mechanisms is the TCP protocol suite, whose design logic is built on static parameters and local feedback, making it ill-suited to the complex scenarios of modern networks. The first issue to address is the singularity of control strategies. Traditional algorithms such as Reno and Cubic rely mainly on packet loss or delay as congestion signals, yet they struggle to determine whether congestion is truly caused by link overload or by route flapping, leading to misjudgment rates exceeding 30% [5]. In wireless networks, packet loss due to signal fading is easily misclassified as congestion, triggering unnecessary window reductions and throughput drops. Because parameters are rigid and conspicuous, traditional mechanisms usually adopt fixed thresholds. When traffic patterns change abruptly, these fixed settings tend to declare congestion prematurely or over-utilize resources, causing link-utilization fluctuations of more than 25%.

Traditional mechanisms lack cross-layer coordination; the transport and network layers act independently, failing to leverage SDN's global view and NFV's resource elasticity. Transport-layer protocols in virtualized networks cannot perceive the state of underlying resources, causing a mismatch between flow-table rules and actual bandwidth and thus exacerbating congestion propagation. In data-center networks, insufficient synergy between TCP's congestion management and ECMP routing's load balancing readily triggers hash polarization, overloading some links while leaving others idle and cutting overall network throughput by up to 40%.

3.2 Insufficient Adaptability to Dynamic Environments

The dynamic and heterogeneous nature of modern network environments imposes stringent challenges on congestion control; traditional mechanisms show clear limitations in handling traffic surges, topology reconfigurations, and resource fluctuations. The high dynamism of traffic patterns is the foremost issue: large-scale IoT device access and explosive growth in AI application traffic render the network traffic matrix highly sparse and violently volatile, causing the accuracy of traditional prediction models based on historical data to plummet to no more than 50%.

Dynamic topology reconfiguration further intensifies control failures. In satellite networks and UAV ad-hoc scenarios, frequent link disconnections and rapid node mobility cause routing paths to change continuously. Traditional congestion-control mechanisms, built on the assumption of a stable topology, can hardly keep pace with real-time network states. Experimental data show that the topology of LEO satellite networks can change within seconds, whereas the convergence time of conventional routing protocols reaches tens of seconds; during this interval traffic is easily misdirected onto failed paths, triggering regional congestion collapse. Because virtualized resources can scale elastically, the state of underlying resources becomes unpredictable. In NFV, dynamic migration of virtual network functions may create instantaneous mismatches between computing resources and network bandwidth. Static resource-allocation strategies cannot adapt well to this dynamic supply–demand relationship and often exacerbate congestion fluctuations through either resource overload or idleness.

3.3 Multi-Objective Optimization Conflicts

Network congestion control must simultaneously satisfy multi-dimensional objectives of throughput, latency, packet loss, and fairness; however, traditional mechanisms struggle to achieve coordinated multi-objective optimization in dynamic environments, causing control strategies to sacrifice one goal for another. The most pressing issue is the conflict between throughput and latency. Traditional TCP protocols use a window mechanism to maximize link utilization, but as the network approaches saturation, queue buffers accumulate continuously, triggering exponential latency growth. To optimize throughput, data-center networks allow queues to grow, causing end-to-end latency to degrade from millisecond to second level, which severely violates the latency constraints of real-time applications. If low-latency assurance is prioritized, premature rate reduction may cause link utilization to drop by more than 20%, creating a vicious cycle of "high latency—low throughput".

The contradiction between fairness and efficiency is also prominent. Traditional algorithms for multi-user shared links are prone to unequal bandwidth allocation due to inconsistent initial window sizes or path lengths. Experiments show that short flows may receive less than 10% of bandwidth when competing with long flows, significantly degrading user experience. While fairness-oriented control strategies can balance resource allocation, they require complex queue-scheduling mechanisms and increase device processing overhead; under high-speed networks, hardware performance bottlenecks may conversely reduce overall throughput.

4. DYNAMIC MITIGATION STRATEGIES FOR NETWORK CONGESTION IN NETWORK ENGINEERING

4.1 Intelligent Prediction and Proactive Control

Intelligent prediction integrates multi-source data and machine-learning algorithms to accurately forecast network congestion and intervene in advance. Joint modeling of historical traffic, topology, and service characteristics builds a dynamic traffic-prediction framework. Spatio-temporal graph convolutional networks capture traffic correlations in both temporal and spatial dimensions. In data-center networks, this model achieves 92% accuracy for predicting burst traffic, a 23% improvement over traditional time-series methods. Prediction-driven proactive control avoids congestion risks by dynamically adjusting resource-allocation strategies. At the access side, software-defined radio can reallocate spectrum in real time based on predictions to prevent wireless-link overload. In the core network, the network-function virtualization manager can pre-scale critical virtual network-function instances to match processing capacity with traffic demand.

Whether active regulation is timely depends on whether the control loop is optimized for low latency. By tightly coupling the prediction module with the control plane and integrating a lightweight prediction engine into the SDN controller, the time from congestion awareness to policy response can be reduced to no more than 10 milliseconds. After introducing intent-driven networking, the degree of automation in regulation is further improved; once the administrator defines intents such as "prioritize traffic for medical IoT," the system can automatically generate and execute a comprehensive policy that includes traffic scheduling and queue management, thereby reducing the latency fluctuation range of critical services to ±5%.

4.2 Hierarchical Collaborative Control Architecture

The hierarchical collaborative control architecture adopts a design that combines vertical layering and horizontal collaboration to achieve cross-layer information sharing and global resource optimization. Vertically, it is structured in three layers: "edge, control, and core." The edge layer deploys lightweight intelligent agents to obtain real-time traffic characteristics and device status of the access network, using programmable switches to extract QoS markings in packet headers (such as DSCP or 802.1p fields) and queue occupancy. The control layer builds a global network view through the SDN controller and integrates multidimensional data reported by the edge layer, using graph algorithms to compute optimal paths and resource allocation schemes. The core layer relies on the elastic computing capacity provided by the NFV resource pool to dynamically adjust the number of instances and parameter configurations of virtual network functions. Horizontal collaboration is achieved through standardized interfaces for cross-domain control, adopting data-plane programmable interfaces defined in P4 language, enabling devices from different operators to support unified flow-table rule distribution and state synchronization mechanisms, thus solving the problem of insufficient compatibility in traditional protocols.

The dynamic adaptability of this structure stems from closed-loop feedback mechanisms and policy self-adaptation. The control layer continuously updates the network state model based on real-time data; once it detects that the utilization of a link exceeds a threshold, the system automatically triggers a three-stage response. Edge agents manage local random early detection queues to suppress burst traffic, SDN flow tables are adjusted to divert part of the traffic to lightly loaded paths, and if the VNF processing capacity in the core layer is still insufficient, the NFV orchestrator initiates scaling of relevant instances.

4.3 Multi-Technology Convergence Optimization Solution

By integrating cutting-edge technologies such as artificial intelligence, blockchain, and intent-driven networking, the multi-technology fusion optimization scheme aims to build an adaptive and reliable network congestion control system. AI focuses on traffic awareness and policy optimization; a deep-reinforcement-learning-based controller performs real-time analysis of the network state matrix and dynamically adjusts congestion-window algorithm parameters. In 5G edge environments, by fusing data on traffic rate, queue length, and channel quality, the algorithm improves burst-traffic adaptability by 40 % while cutting packet loss by 15 %. Blockchain provides trusted cross-domain collaboration: smart contracts trade resources among operators and define congestion responsibility, enabling coordinated backbone resource management. Blockchain nodes automatically log forwarding paths and resource usage; when regional congestion occurs, the responsible party is traced per contract rules, triggering compensation to eliminate the trust costs and dispute delays of traditional collaboration.

Intent-driven networking further raises automation: via a natural-language interface, administrators' intents such as "ensure low latency for critical services" are translated into executable policy sets covering traffic classification, queue scheduling, and path selection. In industrial-Internet scenarios, the IDN system automatically identifies PLC control-flow priorities, instructs the SDN controller to reserve dedicated time slots, and coordinates the NFV resource pool to deploy lightweight intrusion detection, keeping control-command latency within 2ms while maintaining security.

5. CONCLUSION

The paper systematically presents three core countermeasures for congestion control in dynamic networks: intelligent prediction that models spatiotemporal features and proactive regulation to reduce congestion risk; a hierarchical collaborative architecture that integrates cross-layer resources and closed-loop feedback to enhance response timeliness. The multi-technology fusion scheme combining AI, blockchain, and intent-driven networking effectively resolves cross-domain collaboration and automated-deployment challenges.

REFERENCES

[1] Wang, Hao. "Joint Training of Propensity Model and Prediction Model via Targeted Learning for Recommendation on Data Missing Not at Random." AAAI 2025 Workshop on Artificial Intelligence with Causal Techniques. 2025.
[2] Xie, Minhui, and Shujian Chen. "InVis: Interactive Neural Visualization System for Human-Centered Data Interpretation." Authorea Preprints (2025).
[3] Xie, Minhui, and Shujian Chen. "CoreViz: Context-Aware Reasoning and Visualization Engine for Business Intelligence Dashboards." Authorea Preprints (2025).
[4] Zhu, Bingxin. "RAID: Reliability Automation through Intelligent Detection in Large-Scale Ad Systems." (2025).
[5] Zhang, Yuhan. "CrossPlatformStack: Enabling High Availability and Safe Deployment for Products Across Meta Services." (2025).
[6] Hu, Xiao. "GenPlayAds: Procedural Playable 3D Ad Creation via Generative Model." (2025).
[7] Hu, Xiao. "UnrealAdBlend: Immersive 3D Ad Content Creation via Game Engine Pipelines." (2025).
[8] Li, X., Lin, Y., & Zhang, Y. (2025). A Privacy-Preserving Framework for Advertising Personalization Incorporating Federated Learning and Differential Privacy. arXiv preprint arXiv:2507.12098.
[9] Xu, Haoran. "CivicMorph: Generative Modeling for Public Space Form Development." (2025).
[10] Tu, Tongwei. "SmartFITLab: Intelligent Execution and Validation Platform for 5G Field Interoperability Testing." (2025).
[11] Xie, Minhui, and Boyan Liu. "DataFuse: Optimizing Interview Analytics Through Multimodal Data Integration and Real-Time Visualization." (2025).
[12] Zhu, Bingxin. "TaskComm: Task-Oriented Language Agent for Efficient Small Businesses Workflows." (2025).
[13] Zhang, Yuhan. "Learning to Advertise: Reinforcement Learning for Automated Ad Campaign Optimization for Small Businesses." (2025).
[14] Hu, Xiao. "Learning to Animate: Few-Shot Neural Editors for 3D SMEs." (2025).
[15] Tan, C. (2024). The Application and Development Trends of Artificial Intelligence Technology in Automotive Production. Artificial Intelligence Technology Research, 2(5).
[16] Tan, C., Gao, F., Song, C., Xu, M., Li, Y., & Ma, H. (2024). Proposed Damage Detection and Isolation from Limited Experimental Data Based on a Deep Transfer Learning and an Ensemble Learning Classifier.
[17] Zhuang, R. (2025). Evolutionary Logic and Theoretical Construction of Real Estate Marketing Strategies under Digital Transformation. Economics and Management Innovation, 2(2), 117-124.
[18] Han, X., & Dou, X. (2025). User recommendation method integrating hierarchical graph attention network with multimodal knowledge graph. Frontiers in Neurorobotics, 19, 1587973.
[19] Yang, J. (2025, July). Identification Based on Prompt-Biomrc Model and Its Application in Intelligent Consultation. In Innovative Computing 2025, Volume 1: International Conference on Innovative Computing (Vol. 1440, p. 149). Springer Nature.

[20] Yang, Zhongheng, Aijia Sun, Yushang Zhao, Yinuo Yang, Dannier Li, and Chengrui Zhou. "RLHF Fine-Tuning of LLMs for Alignment with Implicit User Feedback in Conversational Recommenders." arXiv preprint arXiv:2508.05289 (2025).

[21] Yang, Yifan. "High Availability Architecture Design and Optimization Practice of Cloud Computing Platform." European Journal of AI, Computing & Informatics 1.1 (2025): 107-113.

[22] Tang, H., Yu, Z., & Liu, H. (2025). Supply Chain Coordination with Dynamic Pricing Advertising and Consumer Welfare An Economic Application. Journal of Industrial Engineering and Applied Science, 3(5), 1–6.

[23] Guo, Y., & Tao, D. (2025). Modeling and Simulation Analysis of Robot Environmental Interaction. Artificial Intelligence Technology Research, 2(8).

[24] Zhang, T. (2025). Research and Application of Blockchain-Based Medical Data Security Sharing Technology. Artificial Intelligence Technology Research, 2(9).

[25] Yu, Z. (2025). Advanced Applications of Python in Market Trend Analysis Research. MODERN ECONOMICS, 6(1), 115.

[26] Wei, Xiangang, et al. "AI driven intelligent health management systems in telemedicine: An applied research study." Journal of Computer Science and Frontier Technologies 1.2 (2025): 78-86.

[27] Su, Tian, et al. "Anomaly Detection and Risk Early Warning System for Financial Time Series Based on the WaveLST-Trans Model." (2025).

[28] Zhang, Yujun, et al. "MamNet: A Novel Hybrid Model for Time-Series Forecasting and Frequency Pattern Analysis in Network Traffic." arXiv preprint arXiv:2507.00304 (2025).

[29] Peng, Q., Planche, B., Gao, Z., Zheng, M., Choudhuri, A., Chen, T., Chen, C. and Wu, Z., 3D Vision-Language Gaussian Splatting. In The Thirteenth International Conference on Learning Representations.

[30] Peng, Qucheng, Ce Zheng, and Chen Chen. "Source-free domain adaptive human pose estimation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[31] Gao, W.; Tayal, D., Gorinevsky, D.: Probabilistic planning of minigrid with renewables and storage in Western Australia. In: 2019 IEEE Power & Energy Society General Meeting (PESGM) (2019). https://doi.org/10.1109/pesgm40551.2019.8973483

[32] Gao W and Gorinevsky D 2020 Probabilistic modeling for optimization of resource mix with variable generation and storage IEEE Trans. Power Syst. 35 4036–45