IPv6 Network Dynamic Label Distribution and Fast Path Reconfiguration Mechanism Based on Deep Reinforcement Learning

Feifei Hu[1*], Xubin Lin[1], Liu Wu[1]
[1]Power Dispatching and Control Center, CSG, Guangzhou Guangdong, 510663, China
Corresponding author: Feifei Hu (e-mail: huff_1024@163.com).

ABSTRACT Against the background of large-scale commercial deployment of IPv6, SRv6 (Segment Routing over IPv6) has become the mainstream forwarding technology. Traditional static label distribution suffers from low resource utilization, while passive fault path reconstruction has high delay and cannot adapt to the graded QoS requirements of services. To realize intensive label management and millisecond-level network self-healing, this paper proposes a joint management and control mechanism based on the time-series weighted improved PPO (RW-PPO) algorithm. Relying on the IPv6 extension header to construct a multi-dimensional network decision space, the algorithm optimizes the weight of training samples and solves the convergence oscillation problem of native PPO under dynamic traffic. A hierarchical distributed label distribution architecture is built to realize on-demand allocation, recycling and aggregation of labels. A three-level reconstruction architecture consisting of fault prediction, path pre-storage and local switching is constructed to reduce the overhead of network-wide routing recalculation. Comparative experiments are carried out based on the Mininet simulation platform. The results show that compared with the traditional OSPF-SR scheme and the native PPO scheme, the proposed mechanism reduces the label occupancy rate by 18.72%, achieves a single-link fault reconstruction delay as low as 9.2 ms, and significantly reduces the interruption duration of high-priority services, which is suitable for the operation and maintenance scenarios of operators' IPv6 heterogeneous service networks.

INDEX TERMS IPv6 network; SRv6 segment routing; deep reinforcement learning; dynamic label distribution; fault self-healing

I. INTRODUCTION

A. RESEARCH BACKGROUND AND SIGNIFICANCE

1) Large-Scale Deployment of IPv6 Networks and Demands for Differentiated Service Bearing

China's IPv6 networks have completed full-domain transformation of backbone networks, metropolitan area networks and edge access networks. Various services such as the Internet of Things, computing power scheduling, high-definition video and industrial control are transmitted via IPv6 protocol networking.

Different from the limited label forwarding system of IPv4, IPv6 carries SRv6 segment labels by virtue of its extension header, which can realize end-to-end customized forwarding paths and support differentiated bandwidth, delay and reliability service scheduling [1].

Among them, industrial control and government dedicated lines are high-priority services with low delay requirements, while Internet of Things traffic is low-priority services with high concurrency [2], [3]. The entire network requires a refined and dynamic label allocation mechanism to adapt to the transmission requirements of graded services. The fixed static label distribution mode can no longer adapt to dynamic traffic fluctuation scenarios.

2) Performance Bottlenecks of Traditional Label Distribution and Fault Path Reconstruction Mechanisms

At present, operators mainly adopt the OSPF-linked static SRv6 label distribution mechanism [3], where network-wide nodes are pre-allocated with fixed segment labels. This approach has the advantage of simple forwarding logic, but its shortcomings are extremely prominent:

First, idle labels are occupied for a long time, and the utilization rate of network-wide label resources is less than 65%, making the label pool prone to exhaustion in large-scale networking.

Second, when link load is unbalanced, labels cannot be dynamically recycled and reallocated, which easily causes local link congestion.

Third, after link or node failures, route recalculation relies on routing protocol flooding, and the reconstruction delay is generally higher than 50 ms, which cannot meet the millisecond-level self-healing requirements of industrial services.

Meanwhile, traditional reconstruction mechanisms only consider connectivity without taking service priority into account, so high-value services are prone to long-term disconnection.

3) Research Value of Deep Reinforcement Learning Empowering Intelligent Network Management and Control

Deep Reinforcement Learning (DRL) has the capabilities of autonomous decision-making in complex environments and iterative optimization of time-series states. It can independently output optimal label allocation strategies based on real-time network topology, traffic and fault data without relying on manually preset routing rules [4]–[6].

Introducing DRL into IPv6 label management scenarios can realize adaptive label distribution for traffic fluctuations and deduction of alternative paths before faults, breaking through the technical limitation of passive response of traditional routing protocols, and building an active and intelligent IPv6 network self-healing system, which has dual values of engineering implementation and theoretical research.

B. RESEARCH STATUS AT HOME AND ABROAD

1) Research Progress of IPv6 Label Distribution Technology

In terms of foreign research, the IETF working group has successively released standards for SRv6 static labels and controlled dynamic label distribution. Filsfils et al. proposed a semi-automatic label distribution scheme based on link bandwidth threshold, which only supports single-node local optimization [7].

In terms of domestic research, Huawei and China Mobile jointly proposed an intra-domain SRv6 label aggregation distribution scheme, which focuses on label compression to save resources [8]. However, its distribution strategy relies on manual threshold configuration with insufficient intelligence, and cannot adapt to network-wide dynamic collaborative scheduling.

2) Research Status of Fast Reconstruction Technology for Network Fault Paths

Existing IPv6 path reconstruction is divided into two categories: pre-calculated reconstruction and real-time calculated reconstruction [9]. Pre-calculated reconstruction backs up network-wide alternative paths in advance, which brings extremely high resource overhead. Real-time reconstruction relies on routing protocol flooding with high computing overhead.

3) Application Status of Deep Reinforcement Learning in Routing Optimization

DQN and PPO algorithms are widely used in traffic scheduling. Native PPO adopts a clipped advantage function to update strategies, which is suitable for continuous network decision-making with optimal stability. However, it does not distinguish the value of time-series traffic samples, and outdated traffic samples will interfere with the accuracy of label decision-making [10], [11].

Most existing multi-agent DRL routing studies focus on IPv4 networks and are not adapted to the IPv6 extension header label bearing rules, so there are defects in algorithm scenario adaptability.

C. RESEARCH CONTENTS

First, this paper analyzes the adaptation shortcomings of IPv6 SRv6 label forwarding mechanism and mainstream DRL algorithms, and constructs a network multi-dimensional state decision model.

Second, it improves the sample sampling mechanism of the PPO algorithm, designs the RW-PPO special label decision algorithm, and constructs a multi-objective weighted reward function.

Third, it designs a linked dynamic label distribution process to realize on-demand allocation, recycling and aggregation of network-wide labels.

Fourth, it constructs a fault-predicted three-level path reconstruction architecture, and connects the whole process of fault perception, alternative path retrieval and label switching.

Fifth, it builds a simulation network and completes comparative experimental verification from the dimensions of distribution efficiency, reconstruction delay and resource overhead.

II. RELATED TECHNOLOGIES AND THEORETICAL BASIS

A. CORE TECHNOLOGIES OF IPV6 NETWORK AND LABEL FORWARDING

1) IPv6 Protocol Architecture and Extension Header Mechanism

A basic IPv6 packet consists of a fixed 40-byte basic header, a variable extension header and payload data. Compared with IPv4, it cancels the checksum field and has higher forwarding efficiency.

SRv6 segment routing carries Segment ID (SID labels) through the IPv6 Routing Extension Header (RH). Each SID corresponds to a forwarding instruction of a network node, and nodes forward packets hop by hop according to the label sequence. It allows flexible editing of forwarding paths without modifying the underlying routing forwarding table [12], [13].

2) Technical Principle of SRv6 Segment Routing

SRv6 includes three types of labels: node SID, link SID and service SID. The node SID identifies backbone forwarding nodes, the link SID identifies interconnected links, and the service SID is bound to service transmission levels.

The intra-domain network management system uniformly manages the SID label pool. In the traditional mode, the network management system issues SIDs statically, and nodes solidify label forwarding paths, which cannot independently change the label sequence when traffic fluctuates.

3) Basic Protocols for Label Distribution and Path Forwarding

Existing IPv6 label distribution is completed in linkage with IS-IS and OSPFv3 protocols. The protocol neighbor negotiation completes the processes of label binding, delivery and revocation. The protocol convergence duration determines the label update speed, and protocol iterative calculation will occupy node CPU computing power. The label distribution delay increases linearly in large-scale networking.

B. BASIC THEORY OF DEEP REINFORCEMENT LEARNING

1) Core Framework and Key Elements of Reinforcement Learning

Reinforcement learning consists of five core elements: agent, environment, state, action and reward [14].

In network management and control scenarios, the agent is the intra-domain management and control node, the environment is the IPv6 network-wide topology link, the state includes load, fault and service traffic data, the action refers to label allocation and path adjustment instructions, and the reward is the quantitative benefit of network performance.

2) Typical Deep Reinforcement Learning Algorithms (DQN, PPO, DDPG)

DQN is suitable for discrete label allocation actions but has high computing overhead [15]; DDPG is prone to gradient explosion [16]; native PPO adopts a clipped advantage function to update strategies, which is suitable for continuous network decision-making with optimal stability. However, it does not distinguish the value of time-series traffic samples, and outdated traffic samples will interfere with the accuracy of label decision-making [11].

3) Adaptation Characteristics of Multi-Agent Reinforcement Learning in Distributed Scenarios

Multi-node distributed agents in the domain make decisions collaboratively. The core agent coordinates the network-wide label pool and issues global scheduling instructions; edge agents collect local link status and report it, and are responsible for local label update execution. Neighboring nodes synchronize label change information in a lightweight incremental manner, which fundamentally avoids the bandwidth loss caused by network-wide routing packet flooding.

C. ANALYSIS OF PERFORMANCE DEFECTS OF EXISTING MECHANISMS

1) Insufficient Dynamic Adaptability of Static Label Distribution Mechanisms

Static SID labels are bound to fixed forwarding paths. When service traffic fluctuates in a tidal manner, labels on high-load links cannot be migrated and scheduled, and labels on idle links cannot be recycled and reused, so the problem of network-wide load imbalance cannot be repaired autonomously.

2) Time Delay and Resource Overhead Problems of Traditional Fault Path Reconstruction

After a link fault, OSPFv3 floods topology packets across the network, recalculates the optimal path iteratively, and updates the network-wide label forwarding table simultaneously. The whole-domain interaction packets are large in quantity with high reconstruction delay, and the reconstruction success rate drops significantly when sudden multi-node failures occur.

3) Limitations of Existing Intelligent Optimization Schemes in IPv6 Scenario Adaptation

Existing DRL routing algorithms are adapted to the IPv4 label system and are not compatible with the IPv6 extension header label encapsulation format. The reward function only considers transmission delay without incorporating label resource overhead and service priority indicators, so the engineering implementation adaptability is poor.

III. IPV6 DYNAMIC LABEL DISTRIBUTION MECHANISM BASED ON DEEP REINFORCEMENT LEARNING

A. OVERALL ARCHITECTURE OF DYNAMIC LABEL DISTRIBUTION SYSTEM

1) System Function Module Division (State Perception, Intelligent Decision-Making, Label Execution)

This paper designs a three-layer distributed label distribution architecture, namely the network-wide state perception layer, the RW-PPO intelligent decision-making layer and the edge label execution layer.

The perception layer collects four types of data in real time: link transmission delay, port load, service level and remaining available labels.

The decision-making layer completes multi-objective optimization calculation and outputs scheduling strategies.

The execution layer modifies the SID sequence in the IPv6 extension header.

The three-layer modules interact at the millisecond level to complete the whole process of dynamic label scheduling in a closed loop.

2) Cooperative Interaction Logic of Distributed Nodes

A hierarchical architecture of core management and control agents in the domain matched with edge forwarding agents is adopted. The core agent uniformly manages the domain-wide SID label pool and issues global scheduling instructions.

Edge agents collect and report local link status, and are responsible for local label update execution. Neighboring nodes synchronize label change information in a lightweight incremental manner, which fundamentally avoids the bandwidth loss caused by network-wide routing packet flooding.

3) Unified Management Mechanism of IPv6 Label Space

The global label pool is divided into three zones: fixed government service labels, dynamic service labels and reserved emergency fault labels.

Government dedicated lines are bound with fixed labels that cannot be scheduled arbitrarily; ordinary Internet and Internet of Things traffic uses recyclable dynamic labels; emergency labels are only enabled in fault scenarios. The zoned management rules are clear, which can effectively avoid cross-service label address conflicts from the source.

B. DESIGN OF DEEP REINFORCEMENT LEARNING DECISION MODEL

1) State Space Definition

A multi-dimensional normalized state space $S = \{L_d, L_b, Q_f, S_n\}$ is constructed, where L_d is the link transmission delay, L_b is the link bandwidth utilization rate, Q_f is the service priority coefficient, and S_n is the number of remaining available labels of the node. All parameters are normalized to the interval $[0, 1]$ to reduce the interference of data with different dimensions on model training and improve the accuracy of agent state recognition.

2) Action Space Design

Three groups of continuously executable scheduling actions are defined to form a complete action space $A = \{A_1, A_2, A_3\}$. A_1 is responsible for allocating dedicated forwarding labels for newly accessed services; A_2 completes label aggregation and idle recycling for low-load links; A_3 performs label path migration adjustment for congested links, fully covering all label scheduling service scenarios under network steady state.

3) Construction of Multi-Objective Reward Function

A reward function is constructed by combining delay, label overhead and congestion degree with weights to avoid the one-sidedness of single-index optimization. The formula is as follows:

$$R = \alpha(1 - \bar{L}_d) + \beta(1 - S_r) - \gamma C_b \tag{1}$$

Where α, β and γ are weight coefficients, L_d is the average transmission delay, S_r is the label occupancy rate, and C_b is the link congestion cost.

For high-priority services, the delay weight is automatically amplified to balance the utilization of network-wide resources and the differentiated service transmission quality.

C. DYNAMIC LABEL DISTRIBUTION ALGORITHM BASED ON IMPROVED PPO

1) Sample Reweighting Improvement Idea for Network Dynamics

Native PPO assigns equal training weights to old and new traffic samples. In dynamic traffic scenarios, expired invalid samples are likely to cause decision deviation.

This paper proposes the RW-PPO time-series weighting strategy, which assigns higher weights to real-time high-load samples according to the similarity of traffic time series, attenuates the weights of long-term idle historical samples, and eliminates invalid training samples, significantly accelerating the algorithm convergence speed in dynamic network environments.

2) Algorithm Training Process and Convergence Optimization Strategy

The algorithm is divided into two stages: offline topology pre-training and online real-time iteration.

In the offline stage, the network-wide topology is imported for pre-training to generate a basic label allocation strategy.

In the online stage, the strategy is updated iteratively in real time, and the loss function clipping threshold is set to limit the amplitude of strategy update, avoiding algorithm gradient oscillation caused by sudden changes in network traffic, and reducing the number of convergence iterations by 32%.

3) Multi-Node Distributed Cooperative Decision-Making Mechanism

A hierarchical collaborative decision-making mode is adopted. Edge agents give priority to local link label path optimization to reduce scheduling response delay.

When local optimization cannot alleviate the network-wide imbalance, the core management and control agent intervenes to complete the global label scheduling correction, taking into account both the real-time response speed of label distribution and the balanced allocation effect of domain-wide label resources.

D. LABEL DISTRIBUTION SIGNALING AND IMPLEMENTATION PROCESS

1) IPv6 Extension Header Label Bearing Scheme

This paper fully reuses the standard IPv6 routing extension header to carry SRv6 labels, and only adds a 1-bit scheduling flag bit in the reserved field of the extension header to distinguish between statically configured and dynamically issued SIDs. It does not need to modify the underlying forwarding chip logic, and existing commercial IPv6 switches and routers can be directly compatible with parsing, greatly reducing the cost of engineering deployment and transformation of the scheme.

2) Dynamic Label Update Trigger Threshold Mechanism

The system sets 75% link bandwidth utilization as the congestion reallocation trigger threshold and 30% utilization as the idle label recycling threshold. When the link load is within the interval, the original label forwarding path is maintained, and RW-PPO re-decision is only started after the threshold is reached, which can effectively reduce meaningless signaling interaction and relieve the pressure of node signaling processing.

3) Cross-Node Label Consistency Synchronization Protocol

Abandoning the network-wide full label table synchronization mode, an adjacent node incremental synchronization mechanism is adopted, which only transmits the changed SID label information, greatly compressing the length of synchronization packets. A label check field is attached during synchronization, which can quickly identify label conflicts and loop risks, and ensure real-time consistency of label forwarding tables of all nodes in the domain.

IV. FAST PATH RECONSTRUCTION MECHANISM FOR IPV6 IN FAULT SCENARIOS

A. OVERALL FRAMEWORK OF FAST PATH RECONSTRUCTION MECHANISM

1) Fast Detection and Localization Mechanism for IPv6 Network Faults

Relying on lightweight BFD detection packets, three typical faults including link interruption, port downtime and node offline are identified at the millisecond level. Linked with the local topology snapshot of edge agents, the faulty link number can be quickly located. The whole fault perception delay is controlled within 1.1 ms, and the detection sensitivity and response speed are much higher than traditional routing convergence schemes such as OSPFv3.

2) Reconstruction Architecture Integrating Pre-Calculation and Real-Time Decision-Making

A three-stage integrated reconstruction architecture of pre-fault prediction, in-fault switching and post-fault optimization is built. In the offline stage, the RW-PPO model is used to identify high-risk links and pre-store multiple alternative label paths. When a fault occurs, the cached paths are directly called for fast switching. After the faulty link is restored, the network-wide load and label resource distribution are uniformly optimized.

3) Perception-Free Switching Logic of Label Paths

Various upper-layer service terminals do not need to perceive the change of underlying forwarding paths. Only the SID label field in the IPv6 extension header is replaced at intermediate forwarding nodes, and the source and destination IP addresses of packets remain unchanged throughout the process, completely retaining the original TCP session connection and realizing perception-free path self-healing switching at the service layer.

B. PRE-GENERATION OF ALTERNATIVE PATHS BASED ON DEEP REINFORCEMENT LEARNING

1) Risk Link Identification and Fault Scenario Prediction

A fault prediction sub-model is trained based on multi-dimensional data such as historical link jitter, long-term load fluctuation and flash failure frequency. Links with sustained high load and frequent flash failures are marked as primary risk links, and 2 to 3 non-overlapping, low-congestion differentiated alternative forwarding paths are preferentially generated for them.

2) Pre-Training Model for Alternative Label Path Set

A multi-scenario sample data set covering single-link and dual-link concurrent faults is constructed offline. After inputting the network fault state, the optimal SID forwarding path is solved and stored in the local high-speed cache table. When a fault is triggered, it can be directly retrieved by table lookup, saving the high computing time-consuming of online real-time iterative solution.

3) Path Resource Reservation Strategy in Multi-Fault Scenarios

15% of emergency dynamic label resources are uniformly reserved for primary risk links, and no additional global label pool space is occupied for low-risk links. Through the graded differentiated reservation strategy, while ensuring the fault redundancy switching capability, the overall utilization efficiency of network-wide labels is improved as much as possible.

C. FAST PATH RECONSTRUCTION ALGORITHM TRIGGERED BY FAULTS

1) Millisecond-Level Diffusion Mechanism of Fault Information

The network-wide topology packet flooding mode of the traditional reconstruction mechanism is cancelled, and lightweight fault instructions are only pushed to the adjacent forwarding nodes upstream and downstream of the fault, which greatly narrows the scope of information diffusion, reduces the conflict of signaling interaction among multiple nodes across the network, further compresses the pre-reconstruction time consumption, and significantly reduces the network-wide signaling transmission load.

2) Local Path Recalculation and Fast Label Update

Only the SID label sequences of nodes directly connected to the fault upstream and downstream are modified, and the routing forwarding tables of other forwarding nodes across the network remain unchanged. Local incremental update replaces domain-wide synchronization, which greatly reduces the computing overhead of label update and shortens the iterative effective time of the forwarding table.

3) Path Switching Scheduling with Service Priority Awareness

In scenarios of concurrent multiple faults and limited network forwarding resources, high-priority services such as government services and industrial services are preferentially scheduled to switch to reserved alternative paths, while low-priority Internet of Things traffic suspends the switching operation. The self-healing priority of core services is strictly guaranteed, and the loss of service packet loss and transmission interruption is minimized.

D. NETWORK STATE OPTIMIZATION AFTER PATH RECONSTRUCTION

1) Link Load Balancing Adjustment after Reconstruction

After the path switching process is fully completed, the RW-PPO agent continuously monitors the bandwidth load of the new forwarding path in real time. If the link utilization rate exceeds the standard rapidly, the branch service label path is fine-tuned immediately, which effectively avoids secondary congestion of alternative paths and maintains the steady state of network-wide link load balance for a long time.

2) Redundant Label Resource Recycling and Optimization

After the faulty link hardware equipment returns to normal and goes online, and the link connectivity is stable, the system automatically recycles the emergency reserved labels temporarily occupied during the fault period, verifies the format compliance of the label fields, and then puts them back into the dynamic label pool for recycling, avoiding long-term idle occupation of limited domain-wide label resources.

3) Long-Term Network Stability Guarantee Mechanism

The network state, decision-making actions and performance data of each fault reconstruction process are fully recorded, the internal training parameters of the model are iteratively updated, and the list of network-wide risk links is dynamically refreshed, which continuously improves the reconstruction decision-making accuracy under similar fault scenarios in the future, and realizes the long-term iterative upgrading of network self-healing capability.

V. SIMULATION EXPERIMENTS AND PERFORMANCE ANALYSIS

A. EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTINGS

1) Simulation Platform Construction

Mininet 2.3 is used to build the IPv6 backbone network, and TensorFlow 2.7 is used to build the DRL algorithm training environment. The server is equipped with Ubuntu 20.04, an Intel Core i7-12700H CPU and 32 GB of running memory. The network is equipped with the SRv6 protocol module to support SID label editing functions.

2) Network Topology and Service Traffic Model

A mesh IPv6 topology with 12 backbone nodes and 24 edge nodes is built, and 3 primary risk links are set. Services are divided into three categories: high-priority industrial services, medium-priority video services and low-priority Internet of Things traffic. The traffic conforms to the Poisson distribution, and the service packet sending rate dynamically fluctuates from 10 Mbps to 100 Mbps.

3) Comparison Schemes and Performance Evaluation Indicators

Three groups of comparison schemes are set: Scheme 1: static OSPF-SR label distribution + passive reconstruction; Scheme 2: native PPO intelligent distribution and reconstruction; Scheme 3: the RW-PPO joint mechanism proposed in this paper. The evaluation indicators include five categories: algorithm convergence rounds, label occupancy rate, fault reconstruction delay, service interruption duration and signaling overhead (see Table 1).

B. PERFORMANCE VERIFICATION OF DYNAMIC LABEL DISTRIBUTION MECHANISM

1) Comparison of Basic Algorithm Convergence Performance

The algorithm convergence efficiency and decision-making delay of label distribution are tested to verify the improvement effect of time-series sample weighting optimization on algorithm performance. The core characteristics and performance comparison of the three schemes are shown in Table 2.

According to the test results, the RW-PPO algorithm proposed in this paper reduces the convergence rounds by 32.2% compared with the native PPO through the optimization of time-series sample weights. Relying on the hierarchical distributed decision-making architecture, edge nodes give priority to local decision-making response, and the single-step label decision delay is only 8.7 ms, which is much lower than the manual configuration update delay of the traditional static protocol. The native PPO is not adapted to the IPv6 label zoned management mechanism and can only schedule general dynamic labels, so there are limitations in scenario adaptability.

2) Comparison of Label Resource Utilization

Three levels of network-wide service load (low, medium and high) are set to test the average network-wide label occupancy rate and average link delay under steady state, so as to verify the intensive label distribution capability. The experimental results are shown in Table 3.

Combined with the data analysis in Table 3, in the full load range, the proposed mechanism can actively recycle labels on idle links based on the link load threshold, and reduce redundant label occupation through label aggregation. The average network-wide label occupancy rate is reduced by 18.72% compared with static OSPF-SR.

In high-load scenarios, relying on label path migration to balance network-wide traffic, the average link transmission delay is only 12.47 ms, which adapts to the scheduling requirements of tidal dynamic service traffic.

3) Guarantee Effect of Service Classified QoS

For three types of services with different priorities, the differentiated service guarantee effect of the label distribution mechanism is tested. The experimental results show that high-priority industrial services always occupy the optimal label path, and the delay fluctuation is controlled within ±1.2 ms.

Low-priority Internet of Things traffic actively yields to high-priority services during congestion periods, and the label path is dynamically adjusted, which not only ensures

TABLE 1. Basic configuration parameters of simulation experiments

Parameter Name	Specific Configuration Value	Parameter Name	Specific Configuration Value
IPv6 prefix length	128 bit	DRL learning rate	0.0003
Single-node label pool capacity	512	Delay weight of reward function α	0.55
BFD fault detection cycle	1 ms	Overhead weight of reward function β	0.30
Basic link bandwidth	1000 Mbps	Congestion weight of reward function γ	0.15
Maximum algorithm iteration rounds	1200	Proportion of reserved emergency labels	15%

TABLE 2. Comparison of core characteristics of three comparison algorithms

Algorithm Scheme	Sample Processing Method	Convergence Iteration Rounds	Single-step Decision Delay	IPv6 Adaptability
Static OSPF-SR	No intelligent decision-making	No iteration required	42.6 ms	Fully adapted
Native PPO	Equal-weight training of samples	862 rounds	16.3 ms	Partially adapted
Proposed RW-PPO	Time-series weighted sample training	584 rounds	8.7 ms	Fully adapted domain-wide

TABLE 3. Comparison of label distribution performance indicators under different service loads

Network-wide Service Load Level	Evaluation Indicator	Static OSPF-SR	Native PPO	Proposed RW-PPO
Low load (30%)	Label occupancy rate / %	62.41	51.36	43.69
Medium load (60%)	Label occupancy rate / %	74.58	60.72	55.86
High load (90%)	Average transmission delay / ms	28.74	19.25	12.47

the transmission stability of core services, but also realizes the efficient utilization of network-wide label resources.

C. PERFORMANCE VERIFICATION OF FAST PATH RECONSTRUCTION MECHANISM

1) Single-Link Fault Reconstruction Performance

Three scenarios including ordinary single-link fault, risk link fault and dual-link concurrent fault are simulated respectively to test the total delay and service interruption duration of the whole process of fault detection, path switching and label update, so as to verify the speed-up effect of the pre-generated alternative path mechanism. The experimental data are shown in Table 4.

In the ordinary single-link fault scenario, relying on BFD millisecond-level fault detection and direct call of locally cached alternative paths, the proposed mechanism does not require network-wide route recalculation, and the total reconstruction delay is only 9.2 ms, which is 82.8% shorter than that of the traditional OSPF passive reconstruction mechanism.

Since multiple sets of alternative label paths are pre-generated for risk links, the service interruption duration is controlled at 14.5 ms, which meets the self-healing requirement of within 50 ms for industrial control services.

2) Multi-Link Concurrent Fault Reconstruction Performance

The extreme fault scenario of simultaneous interruption of dual links is simulated to test the reconstruction success rate and fault tolerance capability of the mechanism. The experimental results show that the traditional OSPF mechanism has a reconstruction success rate of only 79.3% due to network-wide flooding conflicts and increased complexity of route calculation.

The proposed mechanism relies on graded fault diffusion and multi-alternative path resource reservation, and achieves a reconstruction success rate of 98.4% for dual-link concurrent faults. Only a few isolated edge nodes have transient disconnection, and core backbone services are not affected throughout the process.

D. COMPREHENSIVE PERFORMANCE AND OVERHEAD ANALYSIS

1) Validity Verification of Algorithm Modules

Ablation experiments are carried out to verify the performance contribution of three core modules: time-series sample weighting, service priority reward and alternative

TABLE 4. Test results of path reconstruction performance in multiple fault scenarios

Fault Type	Performance Indicator	Static OSPF-SR	Native PPO	Proposed RW-PPO
Ordinary single-link fault	Total reconstruction delay / ms	53.6	21.4	9.2
Risk link fault	Service interruption duration / ms	61.2	24.8	14.5
Dual-link concurrent fault	Reconstruction success rate / %	79.3	91.6	98.4

TABLE 5. Results of RW-PPO algorithm module ablation experiments

Ablation Experiment Group	Excluded Module	Convergence Rounds	Single-link Reconstruction Delay / ms
Group 1	None (complete algorithm)	584	9.2
Group 2	Time-series sample weighting module	796	16.8
Group 3	Service priority reward module	612	11.5
Group 4	Alternative path pre-generation module	579	27.3

path pre-generation, and verify the optimization effect of module collaboration. The experimental results are shown in Table 5.

The ablation experiment results show that the alternative path pre-generation module has the most significant optimization effect on fault reconstruction delay. After excluding this module, the reconstruction delay increases to 27.3 ms, which is the core support for realizing millisecond-level self-healing.

The time-series sample weighting module mainly optimizes the algorithm convergence efficiency and also has an indirect impact on the reconstruction delay.

The service priority reward module mainly guarantees the transmission quality of core services and has little impact on the overall convergence speed and reconstruction delay.

The three modules work collaboratively to jointly achieve the optimal dual performance of efficient label distribution and fast fault reconstruction.

2) Signaling and Computing Overhead Analysis

The number of signaling packets and node computing power occupation during network-wide label update and fault reconstruction are counted. The experimental results show that the traditional OSPF mechanism requires flooding all topology packets for each network-wide label update, and the peak CPU occupation of a single node reaches 37%.

The proposed mechanism adopts adjacent node incremental synchronization and local fault diffusion, the peak CPU occupation of a single node is only 12%, and the number of signaling packets is reduced by 68.4%.

The computing overhead in the offline pre-training stage is concentrated on the management and control nodes, and edge nodes only need to perform table lookup execution in the online decision-making stage with extremely low computing overhead, which adapts to the computing power conditions of existing commercial network equipment.

3) Large-Scale Network Scalability Analysis

By gradually increasing the number of network nodes, the performance attenuation of the mechanism under different scale networks is tested. The experimental results show that when the number of nodes increases from 36 to 120, the single-step decision delay of the proposed mechanism only increases from 8.7 ms to 15.3 ms, and the attenuation amplitude of label distribution efficiency is less than 20%.

The fault reconstruction delay increases gently and is always controlled within 20 ms. Relying on the distributed multi-agent architecture, the mechanism has good scalability and can adapt to the deployment requirements of metropolitan-level IPv6 networks.

VI. SUMMARY AND PROSPECT

Aiming at the pain points of resource waste and lagging fault self-healing in SRv6 IPv6 networks, this paper designs an RW-PPO integrated management and control architecture to realize dynamic label scheduling and fast fault path reconstruction. Simulation results show that the mechanism can significantly save label resources, reduce service interruption duration, and guarantee the transmission quality of services at different levels.

Future research will be deepened and expanded from three aspects:

First, optimize the multi-agent collaborative communication mechanism to adapt to ultra-large-scale cross-domain IPv6 networks of provincial operators and reduce the interaction overhead of cross-domain label scheduling.

Second, connect with the northbound interface of the intent-based network to realize the automation of the whole process of service intent parsing, label distribution and fault self-healing, and reduce manual configuration intervention.

Third, introduce network attack scenarios such as route spoofing and label hijacking, add security constraint indicators to the reward function, and build an intelligent forwarding system with both fault self-healing and network attack resistance.

FUNDING

This work was supported by the Science and Technology Project of Southern Power Grid Corporation (No.ZDKJXM20230027).

REFERENCES

[1] J. Gao, Y. Wang, and R. Liu, "Analysis on deployment and application of data center network based on IPv6," Electron. Technol., vol. 54, no. 9, pp. 54–55, 2025.
[2] M. Mikami, "Segment Routing IPv6 Based Low Latency Mobile Network for Vehicular Communications towards Automated Driving Era: Special Section on Emerging Communication Technologies in Conjunction with Main Topics of ICETC2024," IEICE Trans. Commun., vol. E109-B, no. 2, pp. 147–160, 2026, doi: 10.23919/TRANSCOM.2025CEI0002.
[3] Y. Tian, Z. Wang, X. Yin, et al., "Traffic Engineering in Partially Deployed Segment Routing Over IPv6 Network With Deep Reinforcement Learning," IEEE/ACM Trans. Netw., vol. PP, pp. 1–14, 2020, doi: 10.1109/tnet.2020.2987866.
[4] J. Qu, Y. Yang, L. Gao, et al., "Deep Learning-Based Fast Security Detection Method for Cross-Domain Access Behaviors in IPv6 Networks," J. Circuits Syst. Comput., vol. 35, no. 17, 2026, doi: 10.1142/S0218126626501343.
[5] E. A. Ksimi, C. Leghris, S. Lafraxo, et al., "ICMPv6-based DDoS Flooding-Attack Detection Using Machine and Deep Learning Techniques," IETE J. Res., vol. 70, no. 4, pp. 3753–3762, 2024, doi: 10.1080/03772063.2023.2208549.
[6] E. O. E, A. Mohammed, H. Shady, et al., "Deep-Learning-Based Approach to Detect ICMPv6 Flooding DDoS Attacks on IPv6 Networks," Appl. Sci., vol. 12, no. 12, p. 6150, 2022, doi: 10.3390/APP12126150.
[7] C. Filsfils, P. Camarillo, J. Leddy, et al., "Segment Routing over IPv6 (SRv6)," RFC 8986, pp. 1–40, 2021, doi: 10.17487/RFC8986.
[8] D. Ma, X. Zhao, C. Wang, et al., "Review on application and challenges of SRv6 technology in data forwarding plane," J. Softw., vol. 37, no. 1, pp. 378–397, 2026, doi: 10.13328/j.cnki.jos.007422.
[9] D. Yuan, H. Yu, and C. Jiang, "Cloud-Native SRv6-SDN Architecture for Traffic Inference and Rerouting," in Proc. 2025 IEEE Int. Symp. Parallel Distrib. Process. Appl. (ISPA), 2025, doi: 10.1109/ISPA67752.2025.00176.
[10] Y. Zhao, Q. Guo, Y. Zhao, et al., "An improved deep q-network method for fault diagnosis based on deep reinforcement learning," Measurement, vol. 279, p. 121779, 2026, doi: 10.1016/J.MEASUREMENT.2026.121779.
[11] N. Alyaoui, B. A. Guiloufi, O. Baccouch, et al., "Intelligent Congestion Control and Scheduling in VANETs: An SDN-Based PPO Approach," J. Netw. Syst. Manag., vol. 34, no. 4, p. 104, 2026, doi: 10.1007/S10922-026-10088-9.
[12] B. Liang, F. Li, and N. Zheng, "AutoSRv6: Configuration Synthesis for Segment Routing Over IPv6," IEEE J. Sel. Areas Commun., vol. 43, no. 2, pp. 473–483, 2025, doi: 10.1109/JSAC.2025.3528808.
[13] P. Loreti, A. Mayer, and P. Lungaroni, "Implementation of Accurate Per-Flow Packet Loss Monitoring in Segment Routing over IPv6 Networks," in Proc. IEEE High Perform. Switching Routing Conf. (HPSR), 2020, doi: 10.1109/HPSR48589.2020.9098988.
[14] Z. Ding, Y. Huang, and H. Yuan, "Introduction to Reinforcement Learning," in Lect. Notes Electr. Eng., 2020, doi: 10.1007/978-981-15-4095-0_2.
[15] H. Tran-Dang and D. S. Kim, "DQN-Based Resource Allocation and Offloading (DQN-RAO) Algorithm for Multi-User Multi-Server MEC Network," in Proc. 2026 Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC), 2026, doi: 10.1109/ICAIIC68212.2026.11454139.
[16] Q. Qiu, F. Zeng, and H. Yang, "Memory-Augmented Deep Deterministic Policy Gradient," in Lect. Notes Comput. Sci., pp. 41–52, 2025, doi: 10.1007/978-981-97-8963-4_5.