

Тестирование интеллектуальных вычислительных сетей





Вычислительные ресурсы, Хранение данных и Сеть — Фундамент центра интеллектуальных вычислений

Network requirements for large model training clusters

- Low Latency
- Massive Bandwidth
- Stable Operation
- Large-scale
- Maintainable



Сетевые решения на технологиях RoCEv2, являются развитием интеллектуальных вычислительных сетей нового поколения

Intelligent Computing Center

The development of intelligent computing centers is based on the latest AI theories and leading AI architectures, with **computility technology** and **algorithm models** being the core and key

Computility Technology

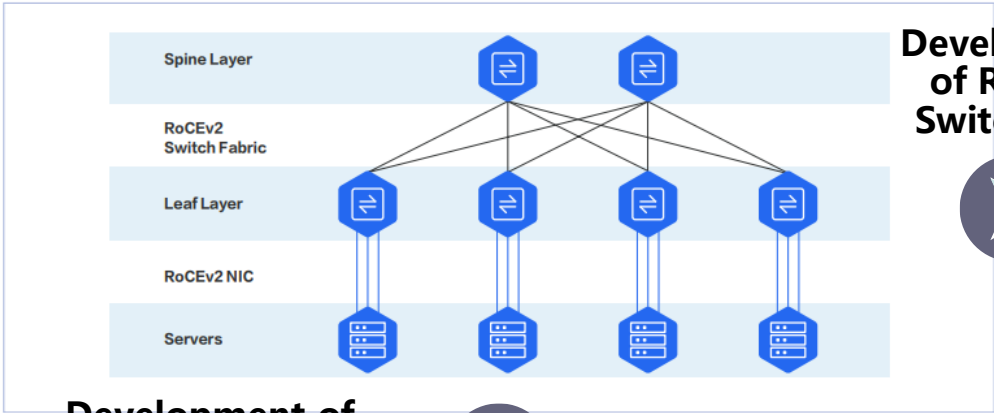
The computility technology is carried by AI chips, AI servers and AI clusters, the current development trend of algorithm models is represented by **AI large models**

Intelligent Computing Network

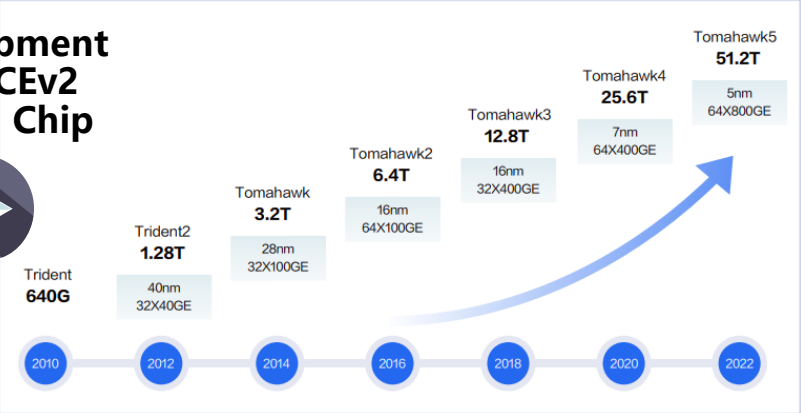
The current intelligent computing network architecture is represented by **IB** and **RoCEv2** technologies. IB currently has superior performance and market share, and Ethernet bearer is the future development direction. It is also a condition for AI to serve thousands of industries and become a universal service



Main Speed of Intelligent Computing Network



Development of RoCEv2 Switch Chip



Development of RoCEv2 Physical Network Card Speed



NIC Speed: 100G/400G/800Gbps
Switch Speed: 100G/400G/800Gbps



Key Technologies and Challenges of Intelligent Computing Networks

The increasing size of GPU cluster networks and clusters brings enormous pressure on GPU demand

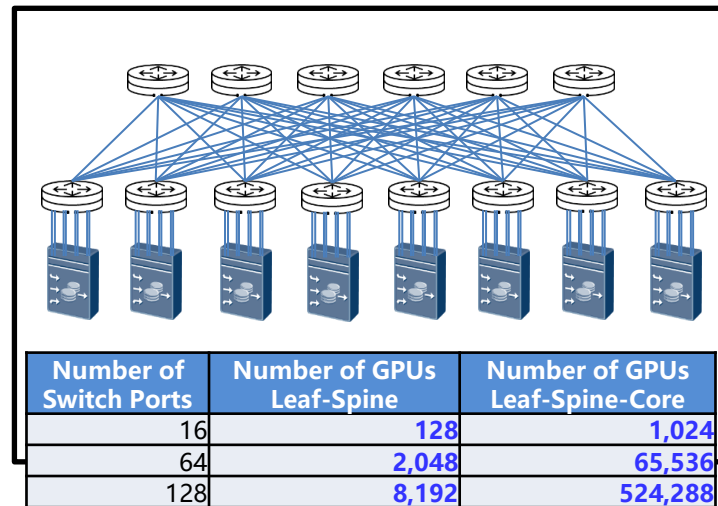
- Two-Tier architecture (Leaf-Spine): the maximum number of GPUs in a GPU cluster is $P^2/2$
- Three-Tier architecture (Leaf-Spine-Core): the maximum number of GPUs in a GPU cluster is $P^3/4$

GPU and power costs are soaring

- The number of GPUs increases exponentially with the size of the network
- Electricity consumption caused by the increase of network servers
- Increased risk of connection failures in transceivers, cables, etc.

Network interconnection and energy conservation and emission reduction

- Network verification of network reliability and stability
- Network assessment of network fault convergence level
- Evaluate the energy-saving and emission reduction level of networking equipment





Key Technologies and Challenges of Intelligent Computing Networks

The loss of network performance can lead to a serious decline in computility performance

The linear acceleration ratio of GPU clusters is influenced by many factors, including the peak computility of GPU cards, video memory capacity, video memory bandwidth, inter card interconnection method, network bandwidth between servers, network architecture, network switches, software and algorithms, and so on.

01 Network lossless and high-speed interconnection are key

- Network packet loss of 0.1% can result in a 50% loss of computility
- The communication required for gradient synchronization within a single calculation iteration is on the order of hundreds of GB

02 High speed RoCE interface testing and ensemble communication simulation

- Selection of network equipment: 100G/400G interface testing, RFC/ECN testing, overlay testing, reliability testing
- GPU NCCL communication simulation, training process network waterline optimization capability evaluation, latency, jitter and out of order evaluation



Challenges for AI/CCL Ensemble Communication

Scalability: Need to support an increasing number of CCL

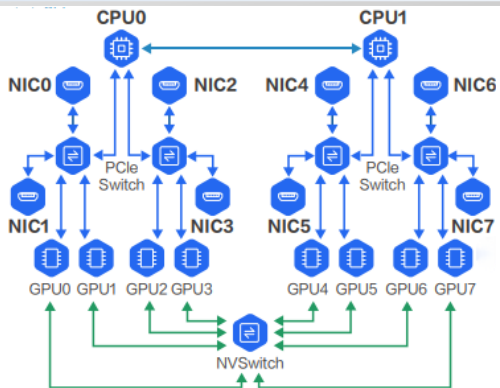
Heterogeneous environment support: supports different types of software and hardware platforms

Compatibility and interoperability: Compatible with other network communication protocols and systems

Fault tolerance capability: ensuring the continuity and reliability of communication

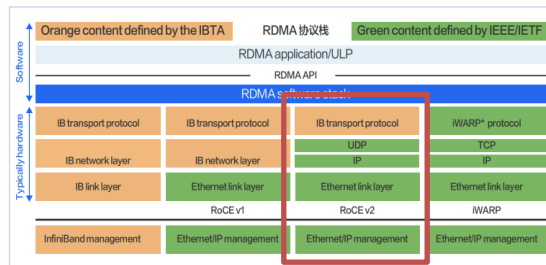
Network latency and bandwidth limitations: high bandwidth and low latency are required

Standardization and normalization: lack



all-gather
all-reduce
broadcast
reduce

...





How To Evaluate Intelligent Computing Network



Lossless Network

As RoCEv2 becomes the main direction of future development, the benchmark verification of flow control mechanisms—including PFC, ECN, and DCQCN—serves as a fundamental prerequisite for constructing intelligent computing networks.

High-Bandwidth Network

Intelligent computing networks exhibit deterministic high-bandwidth requirements. The adoption of 400GE/800GE Ethernet has progressed beyond initial expectations, making high-bandwidth hardware evaluation a clear and essential necessity.

AI Collective Communication

AI high-performance networks demand extreme levels of performance. In AI workloads dominated by large-scale "elephant flows," network capabilities alone are insufficient to ensure efficient throughput. Key approaches such as end-host and network co-optimization, multi-path traffic optimization, and specialized Ethernet enhancements are essential. Consequently, simulating collective communication patterns to evaluate network throughput has become especially critical.

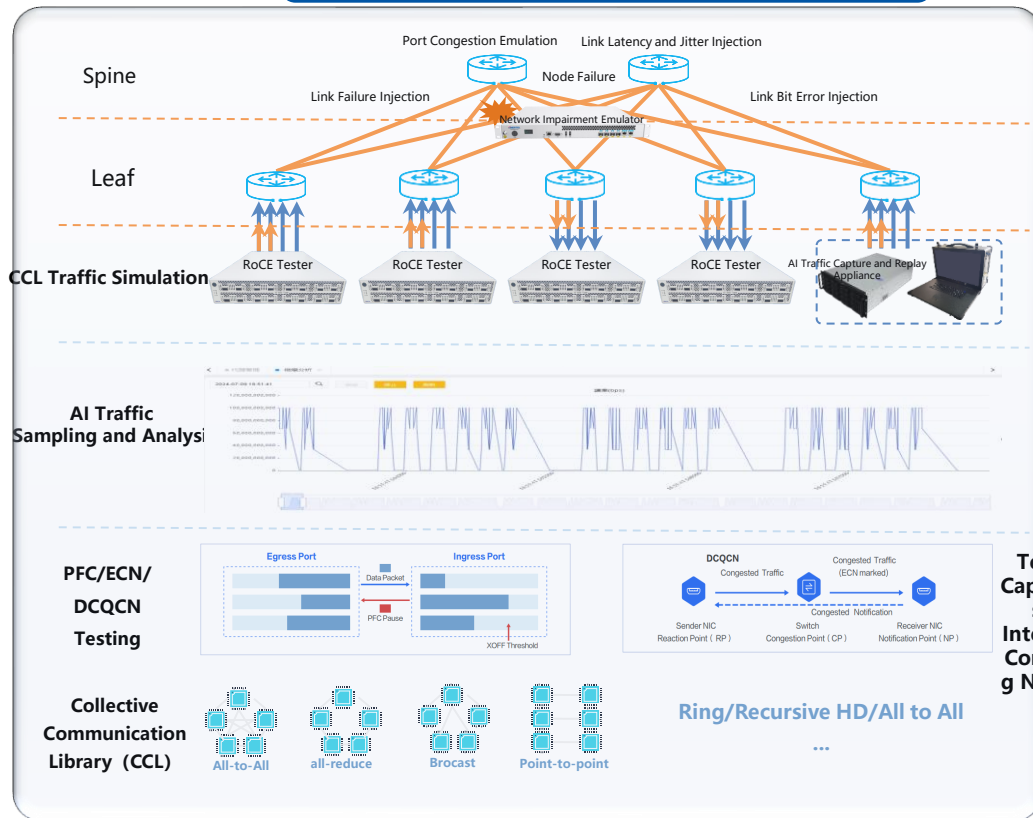
Stability and Reliability Assessment

Intelligent computing networks place stringent demands on service latency, requiring the network to achieve sub-millisecond convergence in the event of failures—ensuring that service continuity remains unaffected.



Xinertel Intelligent Computing Network Testing Solution

RoCEv2 CCL Emulation + AI Traffic Capture and Replay



RDMA AI Performance Testing

AAI Training Server+DPU NIC

100G/25G RoCE

Application and Security Tester

One-Arm / Two-Arm RoCE Testing

- Testing with a security/application emulation tester includes evaluating key performance metrics such as throughput, latency, and multi-client performance under different RDMA protocols (RC/UD).
- Test scenarios include:
 - Two-arm testing of RoCE switches
 - One-arm testing with RoCE-capable DPU NICs
- Test case suite covers:
 - ib_send_bw,ib_send_lat,ib_write_bw,ib_write_lat,ib_read_bw,ib_read_lat,ib_atomic_bw,ib_atomic_lat

Key Testing Capabilities for Intelligent Computing Network

RDMA AI Performance Testing

- Supports both RC and UD service
- Supports Client Single-Arm, Server Single-Arm, and Dual-Arm Testing topologies
- Supports traffic patterns of One-to-Many, Many-to-One, and Many-to-Many
- Conducts key performance tests such as RoCE traffic generation and AI server throughput measurement

Lossless Network + CCL Traffic Emulation + Impairment Simulation

- Supports RoCEv2 Feature Emulation
 - Includes PFC, ECN, DCQCN, and other key mechanisms.
- Collective Communication Library (CCL) Emulation & Network Impairment Simulation
 - Supports algorithms such as All Reduce-DH, All Reduce-Ring, All to All, etc., combined with controlled network impairments for robustness testing.
- Key Performance Metrics Testing
 - Measures critical indicators including Algorithm Bandwidth, Latency, Jitter, and more.

Network Traffic Capture and Replay

- AI Traffic High-Precision Capture
- AI Traffic High-Precision Analysis
- AI Traffic High-Precision Replay



ROCEv2 Software of Xinertel



DarYu 12000



X2-100G-12QSFP28



DarYu 3000



X5-400G

功能	关键特性
PFC	<ul style="list-style-type: none"> • PFC enable and priority configuration • PFC packet response. • PFC packet statistics
ECN	<ul style="list-style-type: none"> • ECN packet detection and CNP packet response • CNP interval configuration • CNP packet priority configuration • ECN/CNP packet statistics
RoCEv2 Traffic	<ul style="list-style-type: none"> • RoCE unidirectional and bidirectional traffic configuration and transmission • Support traffic endpoint selection based on QP • QP quantity: no less than 8K; frame size: no less than 16,384 Bytes • Statistics: <ul style="list-style-type: none"> ➢ Throughput, latency, and packet loss; ➢ Per-QP queue-based statistics.
CCL	<ul style="list-style-type: none"> • Support Ring AllReduce/Having-Doubling AllReduce/Pair Wise AlltoAll;
Software	Renix



ROCEv2 Hardware of Xinertel

-X2-100G-12QSFP28



X2-100G-12QSFP28

Key Features

- Native QSFP28 100G interface, support 12 x 100G L2-3 test ports, or support 6 100G RoCE test ports
- Support the generation and transmission of RoCEv2 traffic
- Supports QOS settings for L2 (VLAN) and L3 (DSCP)
- Support ECN/PFC enabling and priority setting
- Support the selection of traffic endpoints based on QP
- Support the performance test of routing, multicast, access, MPLS, VXLAN, segmented routing (SR) and other protocols
- FPGA based 100% line speed traffic generation, statistics and capture
- Support RFC2544, RFC2889, RFC3918 and other benchmark test suites



ROCEv2 Software of Xinertel

Тестовый моноблок-X5-400G



Моноблок X5-400G-16QDD

Ключевые параметры

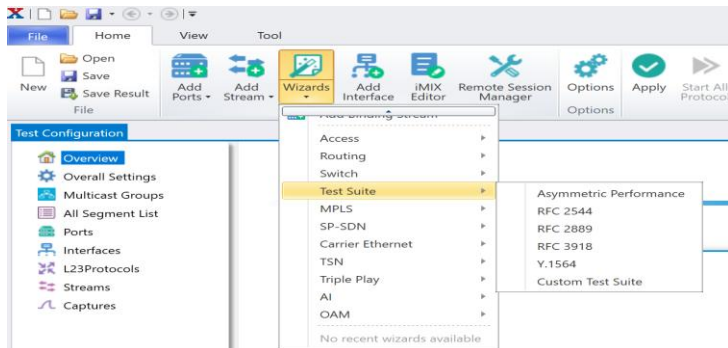
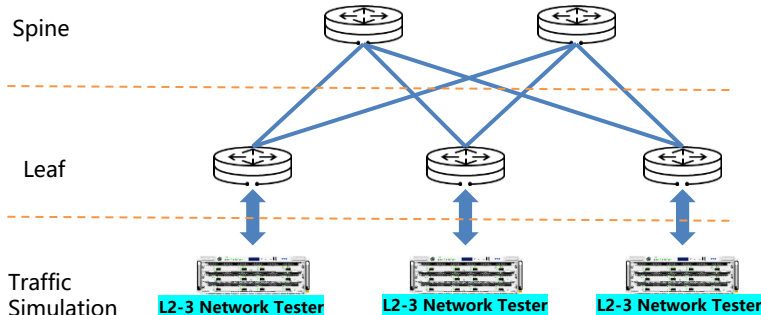
- Native QSFP-DD 400G interface, support 8/16 400G ports, and 400G/200G/100G
- Support large-scale 2-3 layer traffic and routing switching protocol simulation
- Support the performance test of routing, multicast, access, MPLS, VXLAN, segmented routing (SR) and other protocols
- FPGA based 100% line speed traffic generation, statistics and capture
- Support RFC2544, RFC2889, RFC3918 and other benchmark test suites L2-3

Пример кода заказа: X5-400G-16QDD (Multi-Speed RoCE Test Package) - включает 16 портов 400G/200G/100G и лицензию RoCEv2 на все скорости и 6 портов.



Performance Testing of Intelligent Computing Network Devices and Network Architecture

100G/400G/800G Compute Network Test Platform



Comprehensive Protocol Coverage

Stateless Traffic Testing

The 100G/400G/800G hardware platform simulates stateless traffic to test forwarding capability.

Network Capacity Testing

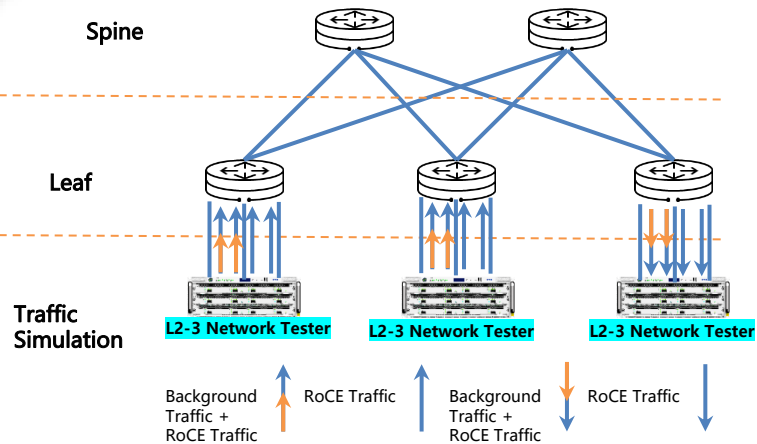
Renix platform combined with the Daryu system offers massive interface access, enabling comprehensive evaluation of traffic ingress, egress, and forwarding latency, jitter, and packet loss metrics under Spine-Leaf architectures.

Test Suites

Renix platform offers an RFC 2544 test suite that combined with 100G/400G/800G hardware, enables precise evaluation of network latency, jitter, and packet loss metrics, while generating professional-grade test reports.



Lossless Network RoCEv2 Testing



RoCEv2 Testing Capabilities for Compute Networks

- PFC/ECN Functionality Testing
- Bandwidth and Latency Testing in N-to-1 RoCE Traffic Scenarios
- Bandwidth and Latency Testing in N-to-N RoCE Traffic Scenarios
- Bandwidth and Latency Testing in 2-to-1 Mixed Traffic Scenarios
- Buffer Loss Analysis, PFC Frame Statistics, and Traffic Counters

Congestion Control Testing

The 100G/400G hardware platform emulates congestion scenarios with the DUT to assess its congestion control performance by enabling PFC and ECN.

Lossless Network Test Suite

Renix platform provides a comprehensive RoCEv2 test suite. Combined with 100G/400G hardware, it enables accurate evaluation of key network metrics such as latency, jitter, and packet loss, and generates professional test reports.

Stateful Traffic Testing

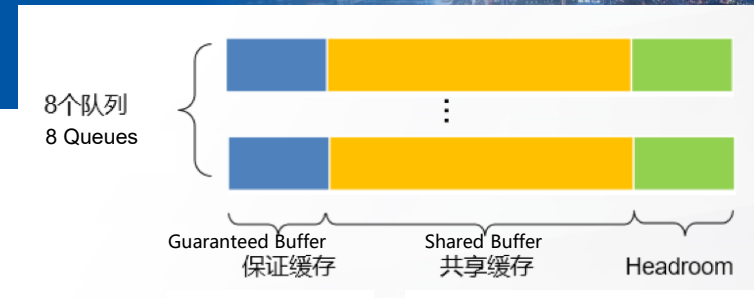
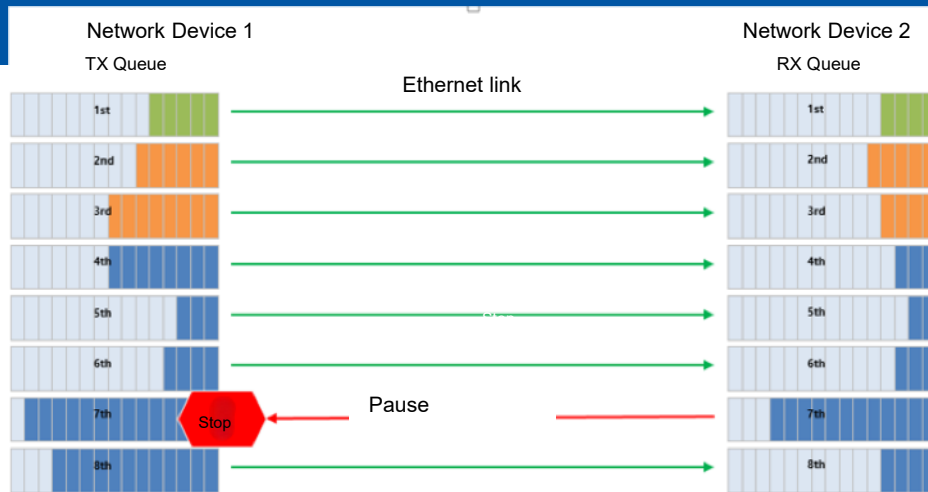
Using a stateful RoCE tester (with full RDMA protocol state implementation), typical communication operations can be simulated by defining jobs. By employing a many-to-many communication pattern, the traffic characteristics of large-scale AI models are emulated. This enables precise measurement of network bandwidth utilization, packet latency, and job completion latency.



PFC (Priority-based Flow Control)

PFC:

PFC provides per-hop priority-based flow control for various types of traffic. When forwarding packets, the device maps the packet's priority using a priority mapping table and assigns the packet to the corresponding queue for scheduling and forwarding. If the sending rate of packets with a certain 802.1p priority exceeds the receiving rate and the receiving end's buffer is insufficient, the receiver sends a PFC pause frame to the sender. Upon receiving the PFC pause frame, the sender stops transmitting packets with the specified 802.1p priority until it receives a PFC XON frame or the aging timer expires. When PFC is configured, congestion of specific types of packets does not affect the normal forwarding of other types of packets.



At the queue level, the buffer is divided into three parts based on the usage scenario: guaranteed buffer, shared buffer, and headroom.

Guaranteed Buffer: Dedicated buffer allocated to each queue to ensure that every queue has a minimum amount of buffer space for basic forwarding. **Shared Buffer:** A common buffer pool shared by all queues, which can be dynamically allocated during traffic bursts to improve buffer utilization. **Headroom:** Reserved buffer space that can still be used after the PFC threshold is triggered and before the server responds to the pause frame and reduces the sending rate. This helps prevent packet loss during this response interval.



PFC Testing

Wizards → Switch → RoCEv2 wizard

- Reserve the test port and switch to RoCE mode
- Configure the port parameters and enable PFC. Taking Priority 6 as an example, set the PFC priority trust mode to Trust-L2 PCP, and configure VLAN ID and IP address information to ensure ARP resolution is successful.
- Configure the RoCEv2 Server, and set the VLAN priority (or configure DSCP in Trust-L3 DSCP mode).

RoCEv2

Configure RoCEv2 Ports Parameters

- Configure RoCEv2 Servers
- Select Queue Pair Traffic Endpoints
- Summary

Select RoCEv2 Ports

Enable two or more ports to configure RoCEv2 servers. The RoCEv2 server will be emulated on the

IP Version: IPv4 Overwrite Non-RoCE Traffic

Select Port		Enable	Port Name	Enable PFC	Custom Priority	PFC Priority Mode	Enable ECN
<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	PortConfig1	<input checked="" type="checkbox"/>	Priority 6	Trust-L2 PCP	<input type="checkbox"/>
<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	PortConfig2	<input checked="" type="checkbox"/>	Priority 6	Trust-L2 PCP	<input type="checkbox"/>

RoCEv2

Configure RoCEv2 Ports Parameters

- Configure RoCEv2 Servers
- Select Queue Pair Traffic Endpoints
- Summary

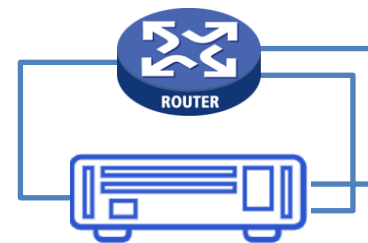
Configure RoCEv2 Server Parameters

Configure RoCEv2 Server Parameter. Emulated devices that are created on the selected ports.

Server Config	Port Name	Addr.	MAC Address	MAC Address Step	Enable VLAN	First VLAN ID	VLAN ID Step	VLAN Priority	DSCP Value	IPv4 Address	IPv4 Address Step	IPv4 Gateway Address	IPv4 Gateway
PortConfig1	1	00:10:94:00:00:01	00:00:00:00:00:01	<input checked="" type="checkbox"/>	1	1	0	0	0	192.168.1.2	0.0.0.1	192.168.1.1	0.0.0.1
PortConfig2	1	01:10:94:01:00:01	00:00:00:00:00:01	<input checked="" type="checkbox"/>	2	1	0	0	0	192.168.2.2	0.0.0.1	192.168.2.1	0.0.0.1

PFC Statistics Under Congestion

Stream/Port Stream Statistic		Select Result View		1/1		Record Per Page: 25	
Basic	PFC	RoCEv2					
Port Name	↑	↑	↑	↑	↑	↑	↑
	↑	↑	↑	↑	↑	↑	↑
Port_1	530,265	8,446	0	0	0	0	0
Port_2	0	0	0	0	0	530,267	8,446



Tester



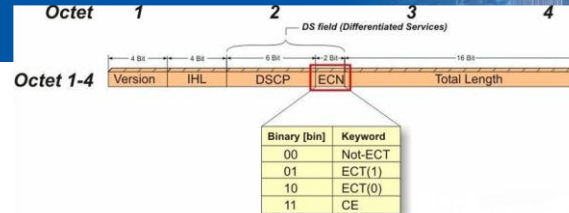
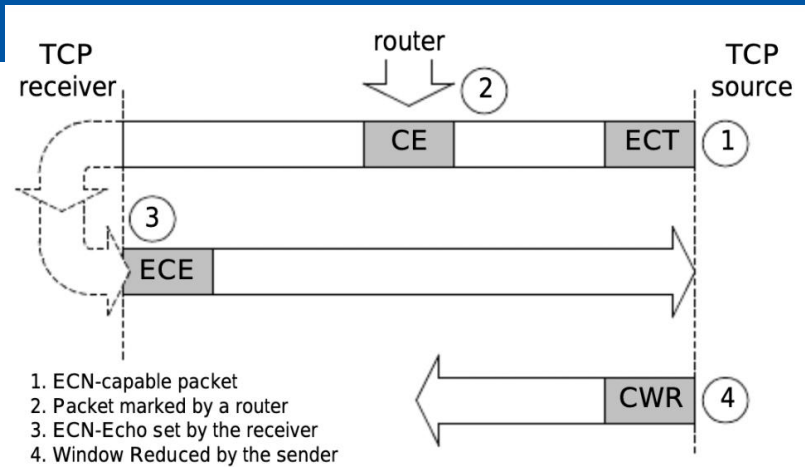
When the transmission rate of packets with a certain priority exceeds the receiving rate, resulting in insufficient available buffer space at the receiver, a Pause frame is sent to the upstream device to pause the transmission of packets with that specific priority. This mechanism ensures lossless transmission.



ECN (Explicit Congestion Notification)

ECN :

- ECN (Explicit Congestion Notification) defines a flow control and end-to-end congestion notification mechanism based on the IP and transport layers. When congestion occurs, the device marks the ECN field in the IP header of the data packet. The receiver, upon detecting the ECN mark, generates a CNP (Congestion Notification Packet) to notify the sender to reduce its transmission rate.
- By enabling end-to-end congestion management, ECN helps to mitigate and prevent the propagation of congestion across the network.



MAC Header	
IPv4/IPv6 Header	
UDP Header	
BTH	
DestQP set to QPN for which the RoCEv2 CNP is generated	
Opcode set to b'10000001	
PSN set to 0	
SE set to 0	
M set to 0	
P_Key set to the same value as in the BTH of the ECN packet marked	
(16 bytes) - Reserved. MUST be set to 0 by sender. Ignored by receiver	
ICRC	
FCS	

ECN Testing

Wizards→Switch→RoCEv2 Wizard

- Reserve the test port and switch it to RoCE mode.
- Configure port parameters and enable ECN. Set the ECN field to 11 (Congestion Experienced, CE).
- Set the CNP (Congestion Notification Packet) priority trust mode to Trust-L2 PCP, and configure VLAN ID and IP address information to ensure successful ARP resolution.
- Configure the RoCEv2 Server, and set the VLAN Priority (or configure DSCP under Trust-L3 DSCP mode).

RoCEv2

Configure RoCEv2 Ports Parameters

Select RoCEv2 Ports
Enable two or more ports to configure RoCEv2 servers. The RoCEv2 server will be emulated on the port.

IP Version: Overwrite Non-RoCE Traffic

Enable	Port Name	Enable PFC	Enable ECN	ECN Value	CNP Priority Mode	L2 PCP Priority	Enable Automatic Rate Adjustment
<input checked="" type="checkbox"/>	PortConfig1	<input type="checkbox"/>	<input checked="" type="checkbox"/>	01 (ECT)	Trust-L2 PCP	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	PortConfig2	<input type="checkbox"/>	<input checked="" type="checkbox"/>	01 (ECT)	Trust-L2 PCP	0	<input type="checkbox"/>

RoCEv2

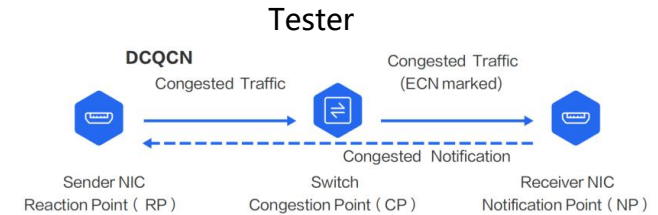
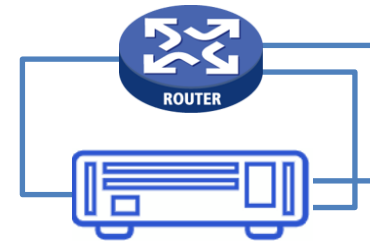
Configure RoCEv2 Server Parameters
Configure RoCEv2 Server Parameter. Emulated devices that are created on the selected ports.

Port Name	Addr.	MAC Address	MAC Address Step	Enable VLAN	First VLAN ID	VLAN ID Step	VLAN Priority	DSCP Value	IPv4 Address	IPv4 Address Step	IPv4 Gateway Address	IPv4 Gateway
PortConfig1	1	00:10:94:00:00:01	00:00:00:00:00:01	<input checked="" type="checkbox"/>	1	1	0	0	192.168.1.2	0.0.0.1	192.168.1.1	0.0.0.1
PortConfig2	1	01:10:94:01:00:01	00:00:00:00:00:01	<input checked="" type="checkbox"/>	2	1	0	0	192.168.2.2	0.0.0.1	192.168.2.1	0.0.0.1

CNP Statistics Under Congestion

Stream/Port Stream Statistic | Select Result View | | 1/1 | Record Per Page: 25

Basic	PFC	Tx Priority	RoCEv2			
Port Name	Rx RoCEv2 Frames	Rx ECN Frames	Rx ECN Rate (fps)	Rx CNP Frames	Tx RoCEv2 Frames	Tx CNP Frames



1. The sender sets the ECT codepoint to 10 in the IP header, indicating ECN capability.
2. When intermediate devices experience congestion, they mark the ECN field in affected packets as CE (11).
3. Upon receiving a packet with the CE mark, the receiver generates and sends a CNP (Congestion Notification Packet) back to the sender.
4. After receiving the CNP, the sender reduces the transmission rate of the corresponding Queue Pair (QP). The original rate is restored either after a certain time interval or after sending a predefined amount of data.



DCQCN (Data Center Quantized Congestion Notification)

DCQCN:

- DCQCN (Data Center Quantized Congestion Notification) is currently the most widely used congestion control algorithm in RoCEv2 networks. It integrates concepts from both the QCN (Quantized Congestion Notification) and DCTCP (Data Center TCP) algorithms.
- DCQCN offers good fairness, achieves high bandwidth utilization, and ensures low queue buffer occupancy with minimal buffer jitter.
- The DCQCN algorithm relies on ECN marking at the switch side. ECN support is a common feature in commercial data center switches. Two bits in the Differentiated Services Field of the IP header are used to indicate congestion. When congestion occurs at the switch, these two bits are set to "11" (Congestion Experienced, CE).

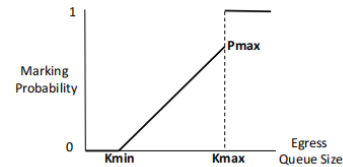
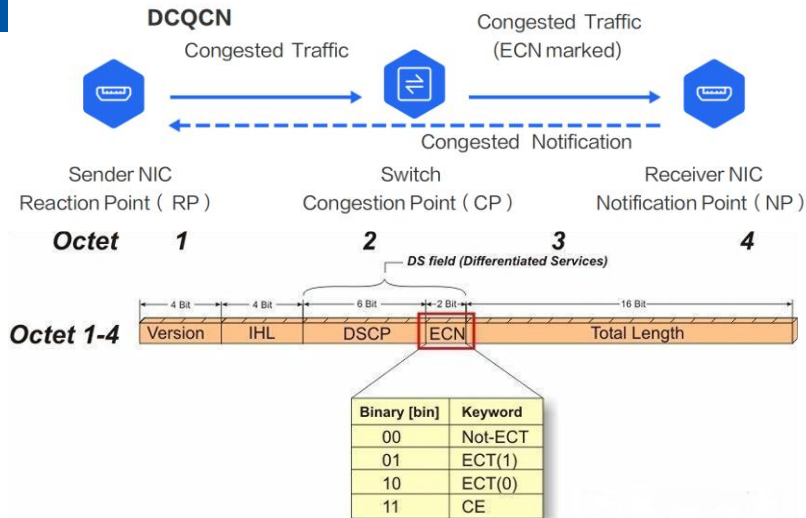


Figure 5: Switch packet marking algorithm

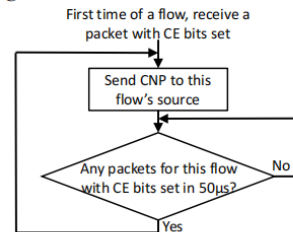


Figure 6: NP state machine

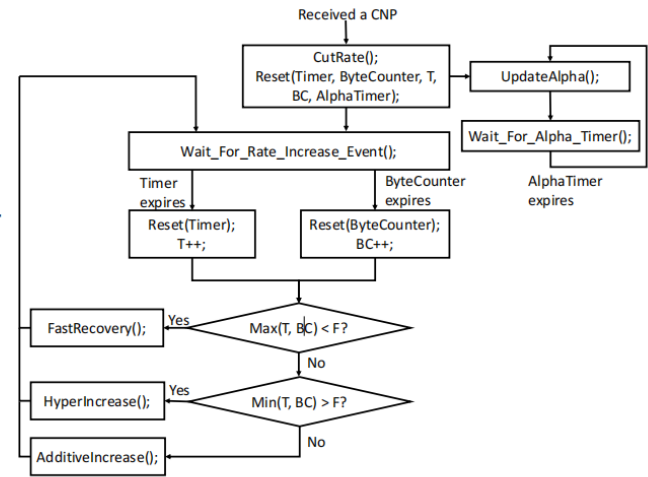


Figure 7: Pseudocode of the RP algorithm



DCQCN Testing

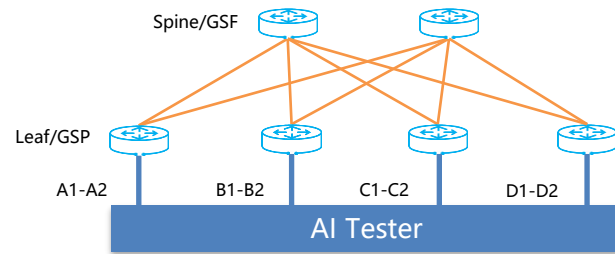
Wizards → Switch → RoCEv2 wizard

- Reserve the test port and switch it to RoCE mode.
- Configure port parameters as follows: Enable ECN, and set the ECN value to 11 (CE); Enable automatic rate adjustment, and select the DCQCN algorithm; Set the CNP priority trust mode to Trust-L2 PCP; Configure VLAN ID and IP address information to ensure ARP resolution is successful.
- Then configure the RoCEv2 Server, and set the VLAN Priority. (Under Trust-L3 DSCP mode, configure the corresponding DSCP value.)

Enable	Port Name	Enable PFC	Custom Priority	PFC Priority Mode	Enable ECN	ECN Value	CNP Priority...	L2 PCP Priority	Enable Automatic Rate Adjustme...	DCQCN Profile
<input checked="" type="checkbox"/>	PortConfig1	<input checked="" type="checkbox"/>	Priority 6	Trust-L2 PCP	<input checked="" type="checkbox"/>	01 (ECT)	Trust-L2 PCP	0	<input checked="" type="checkbox"/>	DCQCN
<input checked="" type="checkbox"/>	PortConfig2	<input checked="" type="checkbox"/>	Priority 6	Trust-L2 PCP	<input checked="" type="checkbox"/>	01 (ECT)	Trust-L2 PCP	0	<input checked="" type="checkbox"/>	DCQCN

Server Config	Port Name	Addr...	MAC Address	MAC Address Step	Enable VLAN	First VLAN ID	VLAN ID Step	VLAN Priority	DSCP Value	IPv4 Address	IPv4 Address Step	IPv4 Gateway Address	IPv4 Gateway
▶	PortConfig1	1	00:10:94:00:00:01	00:00:00:00:00:01	<input checked="" type="checkbox"/>	1	1	0	0	192.168.1.2	0.0.1	192.168.1.1	0.0.1
	PortConfig2	1	01:10:94:01:00:01	00:00:00:00:00:01	<input checked="" type="checkbox"/>	2	1	0	0	192.168.2.2	0.0.1	192.168.2.1	0.0.1

Port Name	Rx RoCEv2 Frames	Rx ECN Frames	Rx ECN Rate (fps)	Rx CNP Frames	Tx RoCEv2 Frames	Tx CNP Frames



Name	CNP Generation Interval (ms)	Time Reset (ms)	Byte Reset (MB)	Phase Threshold	Minimum Rate (Mbps)	Additive Increase Rate (Mbps)	Hyper Increase Rate (Mbps)	Minimum Alpha Value
Default	0.05	0.055	2	5	2	5	40	0.001
DCQCN_0.05	0.05	0.055	10	5	1	40	100	0.001

Key Metric:

DCQCN must take into account the following two conflicting constraints when setting buffer thresholds on the switch:

- PFC (Priority-based Flow Control) should not be triggered too early—it must not occur before ECN marking.
- PFC should not be triggered too late, or it may result in packet loss.

The trigger point for PFC is primarily determined by the buffer threshold configured on the switch.

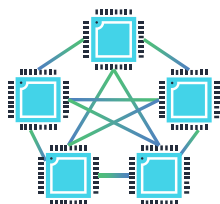
To ensure that ECN is triggered before PFC, the system must account for worst-case scenarios—for example, when all egress queue traffic originates from the same ingress queue. In this case, egress pressure is minimal, while ingress pressure is maximal.

Therefore, it's essential to ensure: $t_{pfc} \leq n * t_{ECN}$

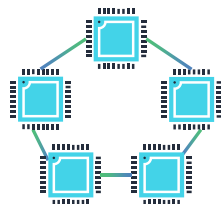


Collective Communication Traffic Model Simulation

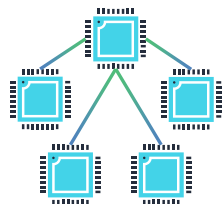
Collective Communication Traffic Simulation(CCL)



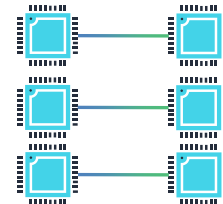
All-to-All



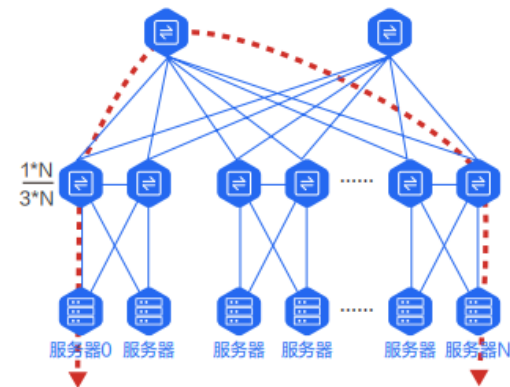
All-reduce



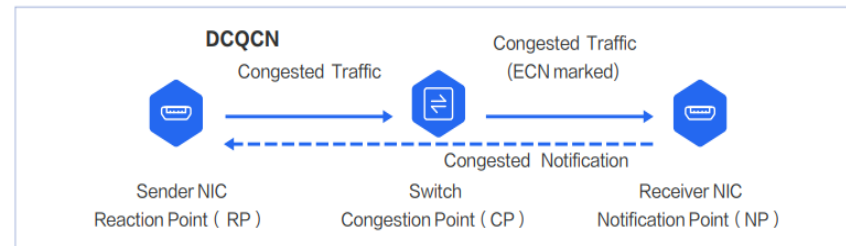
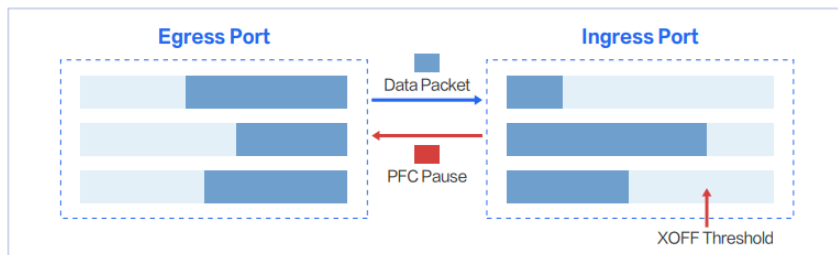
Broadcast



Point-to-point



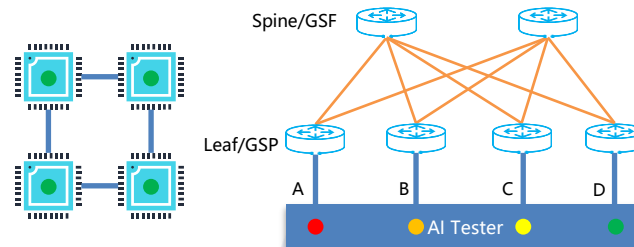
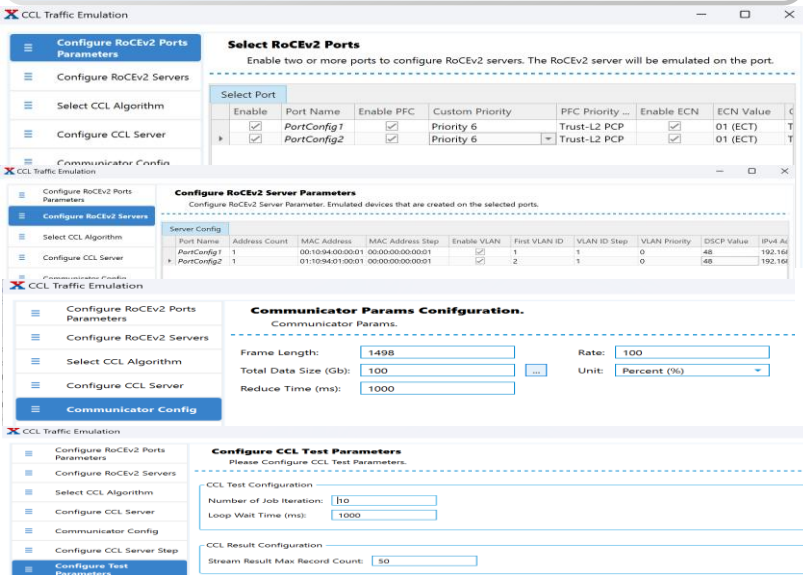
PFC/ECN Testing



CCL(Ring All Reduce) Over RoCE/GSE Testing

Wizards→AI→CCL Traffic Emulation Wizard

- Reserve the test port and switch it to RoCE mode.
- Configure the port parameters as follows: Enable PFC, using Priority 6 as an example; Set the PFC priority trust mode to Trust-L3 DSCP; Configure the IP address information to ensure ARP resolution is successful.
- Then configure the RoCEv2 Server: Set the appropriate DSCP value.
- For collective communication testing: Select Ring as the CCL algorithm; Choose AllReduce as the communication primitive; Arrange the CCL GPU communication ring on demand.



Simulate communication among 4 GPUs, divided into 3 Reduce Scatter and 3 All Gather phases. As shown in the diagram on the right, the traffic is broken down per step, with all GPUs transmitting simultaneously in each step.

Step number	op description	rate unit	Waiting time(ms)
1	Reduce Scatter 300	Percent (%)	10
2	Reduce Scatter 300	Percent (%)	10
3	Reduce Scatter 300	Percent (%)	10
4	All Gather 300	Percent (%)	0
5	All Gather 300	Percent (%)	0
6	All Gather 300	Percent (%)	0

Send port	Send port address	receiving port	sequence
A1	/0.0.0.0/1	A2	Step [1]-RoCEv2 Server 1-Qp Block (100-100)->RoCEv2 Server 2-Qp Block (200-200)
A2	/0.0.0.0/2	A3	Step [1]-RoCEv2 Server 2-Qp Block (200-200)->RoCEv2 Server 3-Qp Block (300-300)
A3	/0.0.0.0/3	A4	Step [1]-RoCEv2 Server 3-Qp Block (300-300)->RoCEv2 Server 4-Qp Block (400-400)
A4	/0.0.0.0/4	A1	Step [1]-RoCEv2 Server 4-Qp Block (400-400)->RoCEv2 Server 1-Qp Block (100-100)
A1	/0.0.0.0/1	A2	Step [2]-RoCEv2 Server 1-Qp Block (100-100)->RoCEv2 Server 2-Qp Block (200-200)
A2	/0.0.0.0/2	A3	Step [2]-RoCEv2 Server 2-Qp Block (200-200)->RoCEv2 Server 3-Qp Block (300-300)
A3	/0.0.0.0/3	A4	Step [2]-RoCEv2 Server 3-Qp Block (300-300)->RoCEv2 Server 4-Qp Block (400-400)
A4	/0.0.0.0/4	A1	Step [2]-RoCEv2 Server 4-Qp Block (400-400)->RoCEv2 Server 1-Qp Block (100-100)
A1	/0.0.0.0/1	A2	Step [3]-RoCEv2 Server 1-Qp Block (100-100)->RoCEv2 Server 2-Qp Block (200-200)
A2	/0.0.0.0/2	A3	Step [3]-RoCEv2 Server 2-Qp Block (200-200)->RoCEv2 Server 3-Qp Block (300-300)
A3	/0.0.0.0/3	A4	Step [3]-RoCEv2 Server 3-Qp Block (300-300)->RoCEv2 Server 4-Qp Block (400-400)
A4	/0.0.0.0/4	A1	Step [3]-RoCEv2 Server 4-Qp Block (400-400)->RoCEv2 Server 1-Qp Block (100-100)
A1	/0.0.0.0/1	A2	Step [4]-RoCEv2 Server 1-Qp Block (100-100)->RoCEv2 Server 2-Qp Block (200-200)
A2	/0.0.0.0/2	A3	Step [4]-RoCEv2 Server 2-Qp Block (200-200)->RoCEv2 Server 3-Qp Block (300-300)
A3	/0.0.0.0/3	A4	Step [4]-RoCEv2 Server 3-Qp Block (300-300)->RoCEv2 Server 4-Qp Block (400-400)
A4	/0.0.0.0/4	A1	Step [4]-RoCEv2 Server 4-Qp Block (400-400)->RoCEv2 Server 1-Qp Block (100-100)
A1	/0.0.0.0/1	A2	Step [5]-RoCEv2 Server 1-Qp Block (100-100)->RoCEv2 Server 2-Qp Block (200-200)
A2	/0.0.0.0/2	A3	Step [5]-RoCEv2 Server 2-Qp Block (200-200)->RoCEv2 Server 3-Qp Block (300-300)
A3	/0.0.0.0/3	A4	Step [5]-RoCEv2 Server 3-Qp Block (300-300)->RoCEv2 Server 4-Qp Block (400-400)
A4	/0.0.0.0/4	A1	Step [5]-RoCEv2 Server 4-Qp Block (400-400)->RoCEv2 Server 1-Qp Block (100-100)
A1	/0.0.0.0/1	A2	Step [6]-RoCEv2 Server 1-Qp Block (100-100)->RoCEv2 Server 2-Qp Block (200-200)
A2	/0.0.0.0/2	A3	Step [6]-RoCEv2 Server 2-Qp Block (200-200)->RoCEv2 Server 3-Qp Block (300-300)
A3	/0.0.0.0/3	A4	Step [6]-RoCEv2 Server 3-Qp Block (300-300)->RoCEv2 Server 4-Qp Block (400-400)
A4	/0.0.0.0/4	A1	Step [6]-RoCEv2 Server 4-Qp Block (400-400)->RoCEv2 Server 1-Qp Block (100-100)

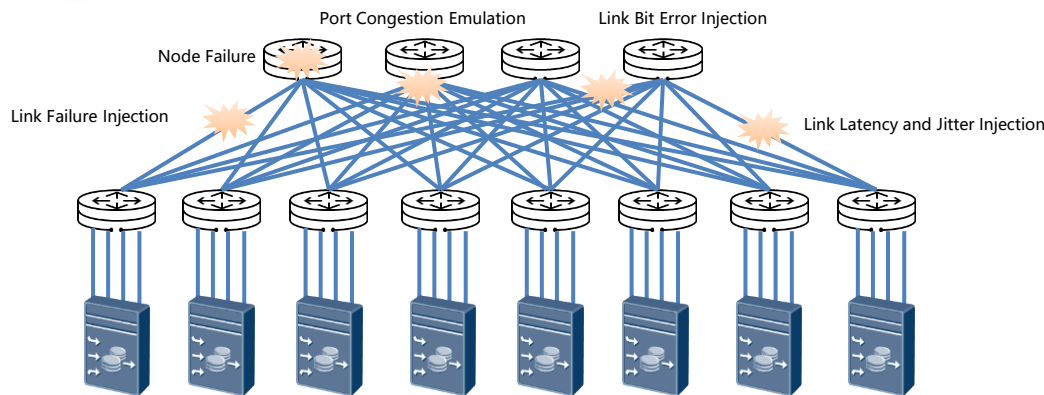
Task Breakdown (Taking 2GB as an Example)

In a 4-GPU communication scenario using Ring AllReduce, the data is divided into 16 equal parts —each GPU handles 4 parts, and each part is 128MB.

- Step 1: A → B, B → C, C → D, D → A: all GPUs simultaneously send their respective data segments
- Steps 2 and 3: Same pattern as Step 1 — these steps together constitute the Reduce Scatter phase
- Step 4: A → B, B → C, C → D, D → A: all GPUs again simultaneously send data
- Steps 5 and 6: Same pattern as Step 4 — these steps form the All Gather phase



Intelligent Computing Network Simulation and Testing



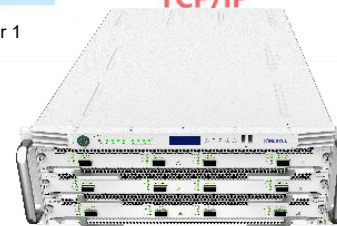
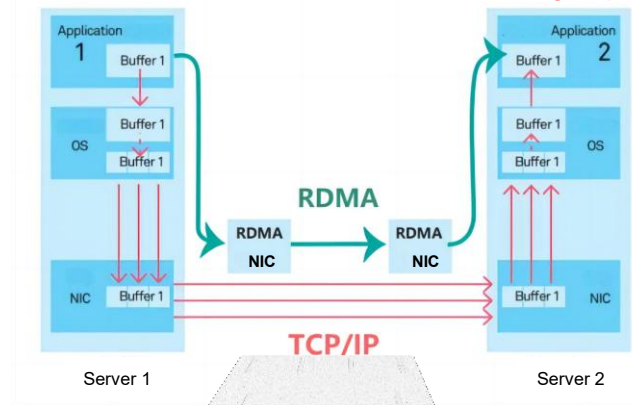
Simulate real-world network environments to validate network architecture and service deployment plans.

- PFC/ECN Testing
- Hash Imbalance, Packet Reordering Simulation, and Reordering Depth Testing
- Service Resilience Testing under Node Failure Scenarios
- Service Resilience Testing under Link Failure Scenarios
- Service Resilience Testing under Latency, Jitter, and Bit Error Conditions
- Verification of SDN-Based Load Balancing and Scheduling Algorithms in Data Center Networks



XCompass Series Network Impairment Emulator

Latency, Jitter, Packet Loss, Bit Error, Packet Reordering, Duplication, etc.



DarYu Series Network Tester

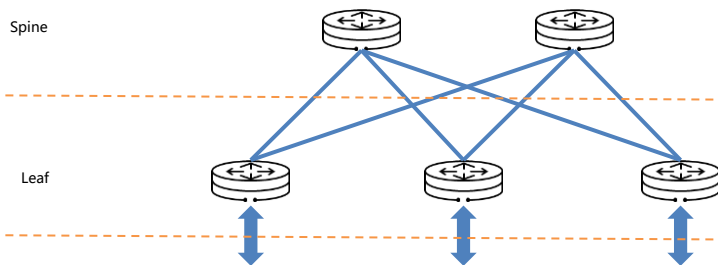
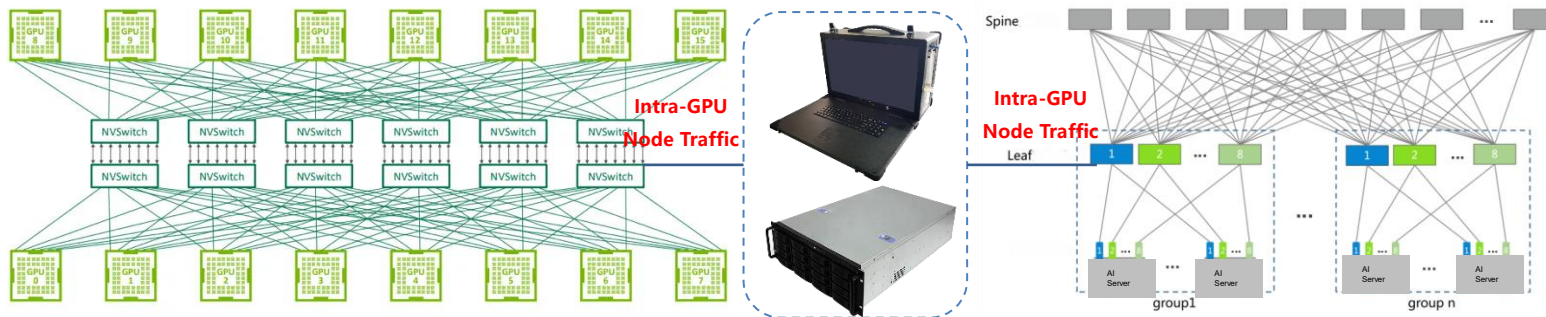
The DarYu series offers massive-scale RoCEv2 traffic emulation and integrated RFC 2544 test suites for standardized performance benchmarking.



AI Large Model Training Traffic Capture & Analysis & Replay

X-Spider Network Traffic Processing

Fast Traffic Search, Original-Speed Replay, Line-Rate Replay, Micro burst Analysis



Traffic Capture / Replay / Analysis



XA-N4200P Network Traffic Processing Device, 2 Ports, Each Port Supports 200G/100G

High-Speed Ethernet Traffic Capture

Supports high-speed extraction of target network packets from massive traffic volumes or full line-rate packet capture via 10GE, 25GE, and 100GE interfaces.

Network Traffic Replay

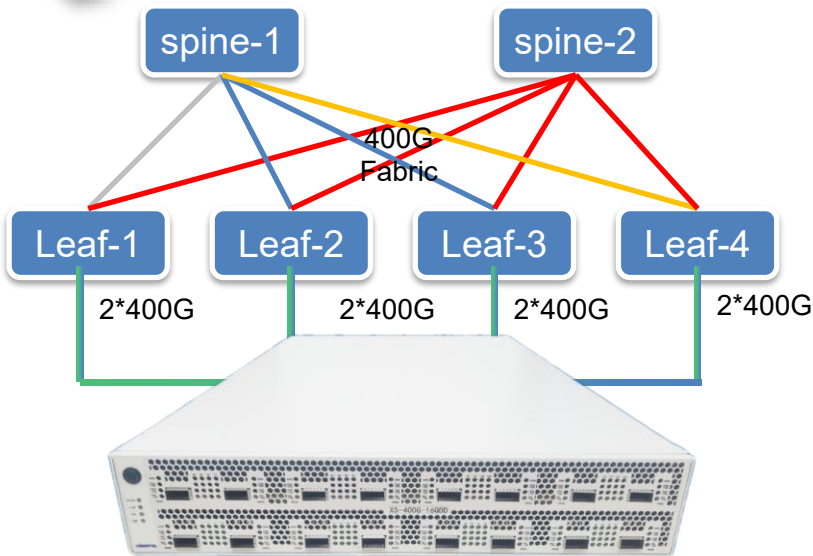
Replay the captured training traffic within and between GPU nodes, achieving lossless restoration of the training process.

Micro burst Traffic Analysis

Implement long-duration micro burst traffic analysis based on sampling intervals of 1 μ s, 10 μ s, and 100 μ s.



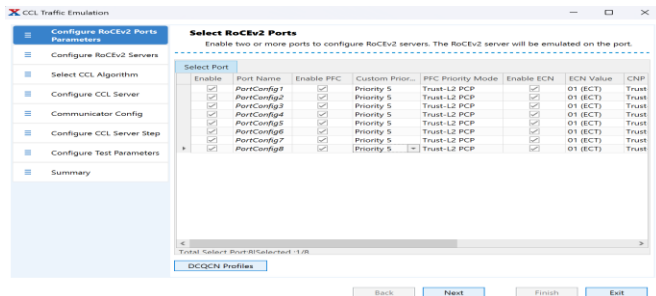
Case Study 1: RoCE Testing of a Telecom Operator



The Renix software platform by Xinertel offers a dedicated CCL Traffic Emulation Wizard, enabling users to simulate and analyze collective communication loads (CCL) under various network conditions. By emulating different AI training packet patterns, the platform supports performance evaluation in both lossless and lossy network environments. Key test items include:

➤ **Performance testing of collective communication in non-blocking scenarios**

Total Data Size (Gbit)	Total Time (ms)	Transmission Time (ms)	Algorithm Bandwidth (Gbps)	Bus Bandwidth (Gbps)	L1 Algorithm Bandwidth (Gbps)	L1 Bus Bandwidth (Gbps)
1	18392.97687	4.725392	211.6226548	370.3527293	228.5643561	400.0012799
2	18281.49651	9.450784	211.6226548	370.3527293	228.5643561	400.0015169
4	18116.90719	18.901162	211.6272005	370.3522465	228.5692657	400.001114
8	18471.32233	37.80196	211.6292383	370.3515936	228.5714666	400.0004681
16	18657.41816	75.603962	211.6291207	370.3513879	228.5713396	400.0002755
32	18815.76082	151.20798	211.6290423	370.3512507	228.571255	400.0001422
64	18342.19819	302.416058	211.6289737	370.3511307	228.5711809	400.00002
128	18535.1001	604.831696	211.6291207	370.3511242	228.5713396	400.0000167
256	18740.6593	1209.662986	211.6291917	370.3511166	228.5714164	400.0000104
512	18503.15487	2419.326028	211.6291868	370.3511081	228.5714111	400.000002

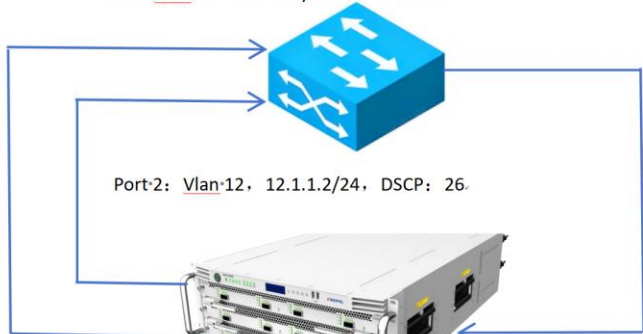


Total Data Size (Gbit)	Tx Frames	Rx Frames	Min Latency (us)	Max Latency (us)	Avg Latency (us)	Min Jitter (us)	Max Jitter (us)	Avg Jitter (us)
1	1228976	1228976	2.503	2.625	2.585946429	0	0.038	0.002473214
2	2457952	2457952	2.503	2.625	2.585848214	0	0.035	0.002392857
4	4915792	4915792	2.503	2.625	2.586580357	0	0.037	0.002223214
8	9831472	9831472	2.503	2.625	2.586464286	0	0.035	0.002107143
16	19662944	19662944	2.507	2.628	2.5869375	0	0.03	0.002017857
32	39325888	39325888	2.503	2.625	2.586875	0	0.033	0.002017857
64	78651776	78651776	2.503	2.628	2.586955357	0	0.032	0.002
128	157303440	157303440	2.5	2.628	2.586732143	0	0.033	0.002
256	314606768	314606768	2.5	2.628	2.586616071	0	0.035	0.002
512	629213536	629213536	2.503	2.628	2.586741071	0	0.033	0.002



Case Study 2: PFC/ECN Testing of a Telecom Operator

Port-1: Vlan-11, 11.1.1.2/24, DSCP: 26.

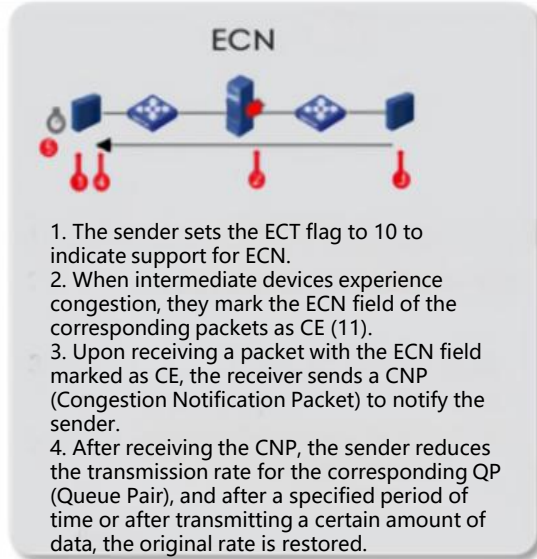
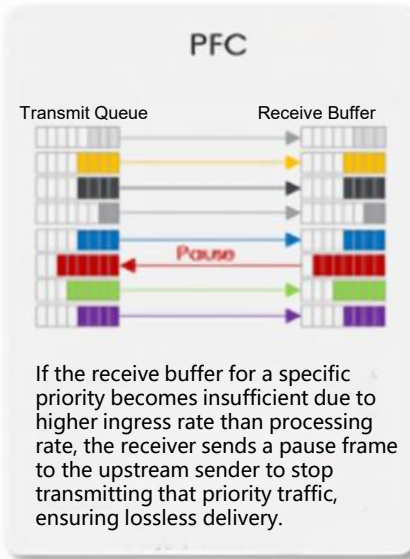
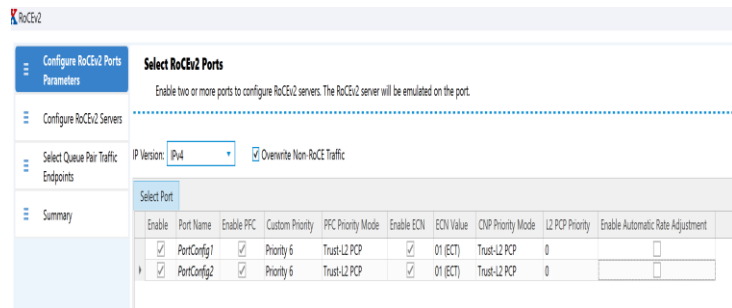


Port-2: Vlan-12, 12.1.1.2/24, DSCP: 26.

Port-3: Vlan-13, 13.1.1.2/24, DSCP: 26.

The Renix software platform from Xinertel includes a dedicated RoCEv2 testing wizard, which enables users to assess the behavior and performance of key congestion management mechanisms such as Priority Flow Control (PFC), Explicit Congestion Notification (ECN), and Data Center Quantized Congestion Notification (DCQCN). Specific test scenarios include:

- **Verifying PFC triggering and lossless transmission based on DSCP values**
- **Simulating congestion conditions to test ECN marking and reaction mechanisms**



Дистрибьютор в РФ и РБ ООО «Комтиню»

☎ Tel: +7-495-937-3609

✉ sales@comtinu.ru

🌐 www.comtinu.ru

