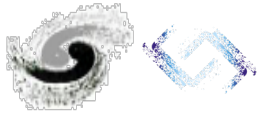
国家高能物理科学数据中心
National HEP Science Data Center

高能所計算中心
IHEP Computing Center

perfSONAR@IHEP Data Center

Mengyao Qi, Shan Zeng, Tao Cui
IHEP Computing Center
2024-10-11

Outline



1 Introduction and Challenges

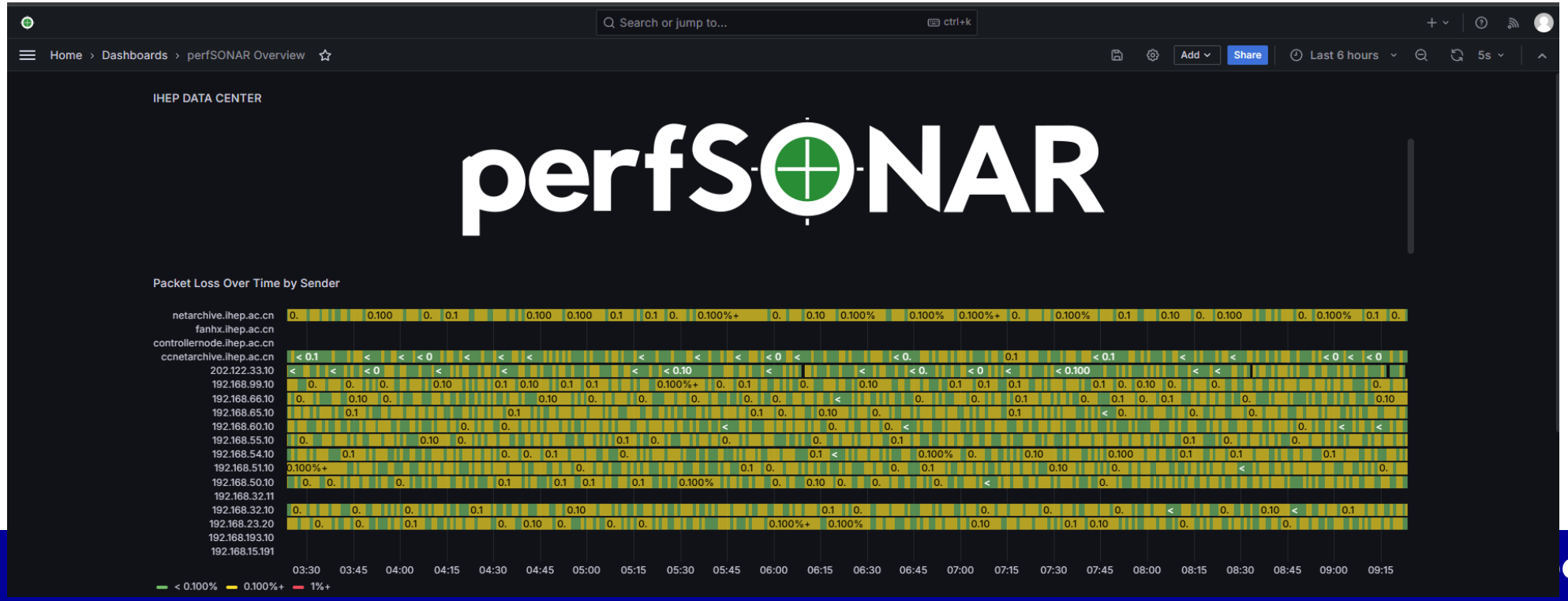
2 Current status

3 Future Plan

4 Summary

Introduction

- Importance of network performance measurement in data centers
  - High-energy physics research relies heavily on data transfer and network stability
- Key metrics: latency, jitter, bandwidth, packet loss, throughput
- To ensure optimal performance and reliability of services



Challenges

- **Technical** :
  - Compatibility issues, unexpected test failures
- **Learning Curve**:
  - New features and configurations in the latest version
- **Challenges**:
  - Limitations of older versions, scalability issues
  - A vast number of data center switches and IP subnets.

Current status

- **Initial Deployment**:
  - Overview of the starting point, including version and scale

IHEP NetWork Dashboard

☰ Dashboards ⚠ Topology ⚙ Setting

Last page refresh time: 10月

IHEP Dashboard

IHEP - DataCenter Latency Testing

Loss rate is <= 0 | Loss rate is >= 0 | Loss rate is >= 1 | Unable to retrieve data | Check has not yet run

DC-13-1-CR60-10, DC-13-2-CR61-254, DC-18-1-DMZ37-82, DC-23-1-IG33-254, DC-28-1-CR55-10, DC-33-1-CR54-9, DC-53-1-CS50-10, DC-53-2-CR50-30, DC-53-3-CR50-31, DC-58-1-CR65-10, DC-58-2-CR51-10, DC-63-1-DMZ37-99, DC-68-1-CR55-9

IHEP - Campus Latency Testing

Loss rate is <= 0 | Loss rate is >= 0 | Loss rate is >= 1 | Unable to retrieve data | Check has not yet run

CP-HAL03-38-9, CP-HAL04-204-10, CP-HAL05-205-10, CP-HAL06-206-10, CP-HAL14-0-10, CP-MBA209-28-6, CP-MBA510-35-10, CP-MBB222-37-8, CP-MBC227-36-10, CP-MSup-208-10, CP-Report-207-10, CP-WKS_II-220-10

IHEP - International Latency Testing

Loss rate is <= 0 | Loss rate is >= 0 | Loss rate is >= 1 | Unable to retrieve data | Check has not yet run

Beijing-IHEP

IHEP - BES3 - IPv4 Latency Testing

Loss rate is <= 0 | Loss rate is >= 0 | Loss rate is >= 1 | Unable to retrieve data | Check has not yet run

IHEP - BES3 - IPv4 Bandwidth Testing

Throughput >= 600Mbps | Throughput < 600Mbps | Throughput <= 30Mbps | Unable to retrieve data | Check has not yet run

Upgrading

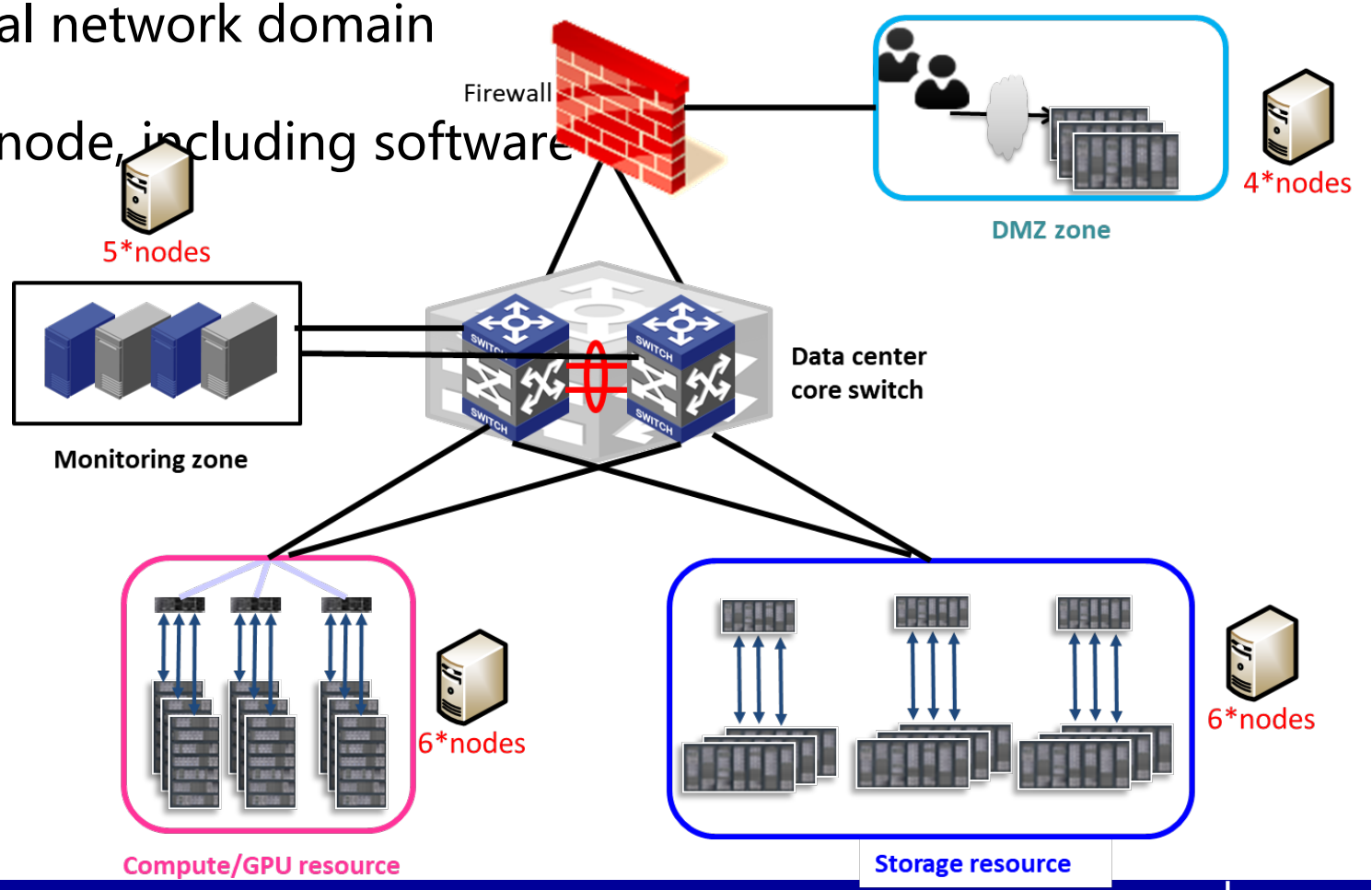
```
[root@perf-99-10 ~]# df -hl
Filesystem                Size  Used Avail Use% Mounted on
devtmpfs                  4.0M     0  4.0M   0% /dev
tmpfs                     1.8G  440K  1.8G   1% /dev/shm
tmpfs                     712M   66M  647M  10% /run
/dev/mapper/almalinux-root  26G  9.6G   17G  38% /
/dev/sda1                 960M  412M  549M  43% /boot
tmpfs                     356M     0  356M   0% /run/user/0
```

- **Upgrade Process**:
  - Centos 6.x to Almalinux9.4
  - The version of perfSONAR from 4.x to 5.1.3
  - Use a high-performance server as the core node(central opensearch instance).
- **Benefits of Upgrading**:
  - Enhanced features, better performance, improved user interface, and security patches

```
[root@perf-99-10 ~]# top
top - 09:24:29 up 20 days, 17:17,  1 user,  load average: 0.43, 0.37, 0.41
Tasks: 326 total,   1 running, 322 sleeping,   0 stopped,   3 zombie
%Cpu(s):  2.9 us,  5.9 sy,  0.0 ni, 91.2 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
MiB Mem :   3558.2 total,    800.0 free,   2567.3 used,    568.6 buff/cache
MiB Swap:   3056.0 total,   1324.1 free,   1731.9 used.    990.9 avail Mem
```

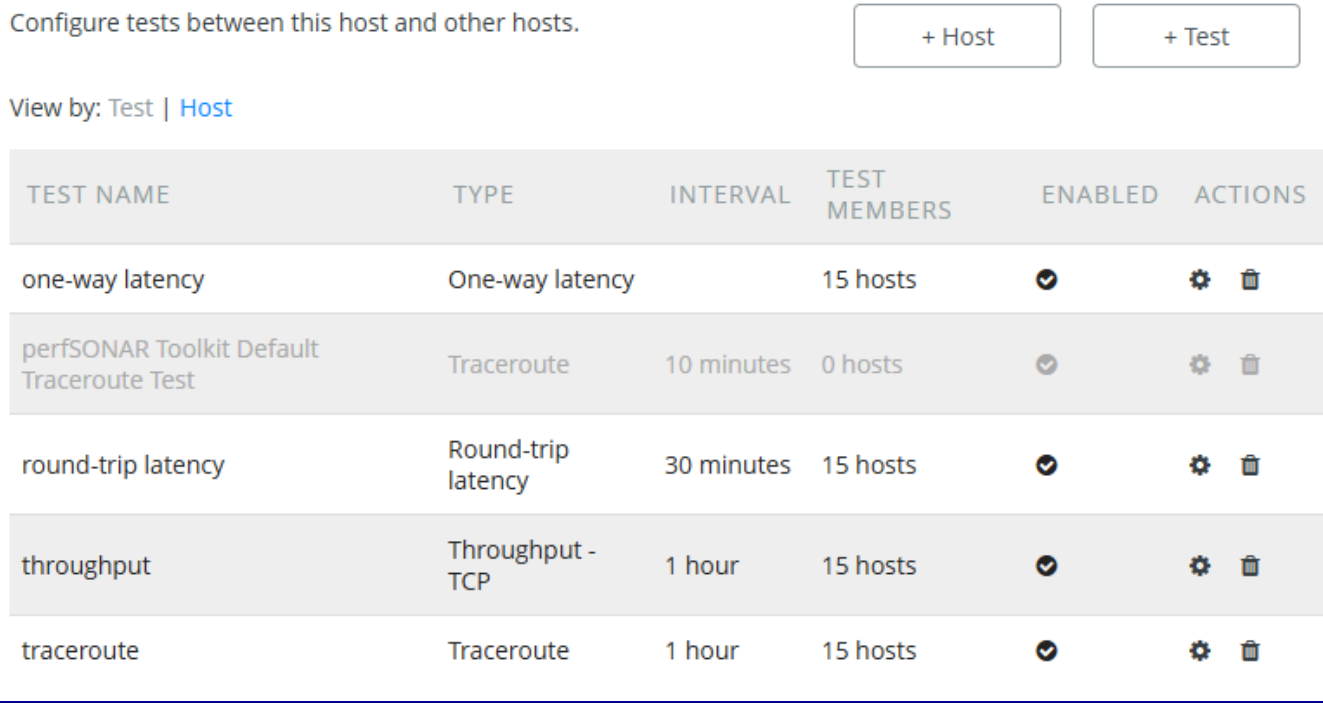
perfSONAR Topology@IHEP DC

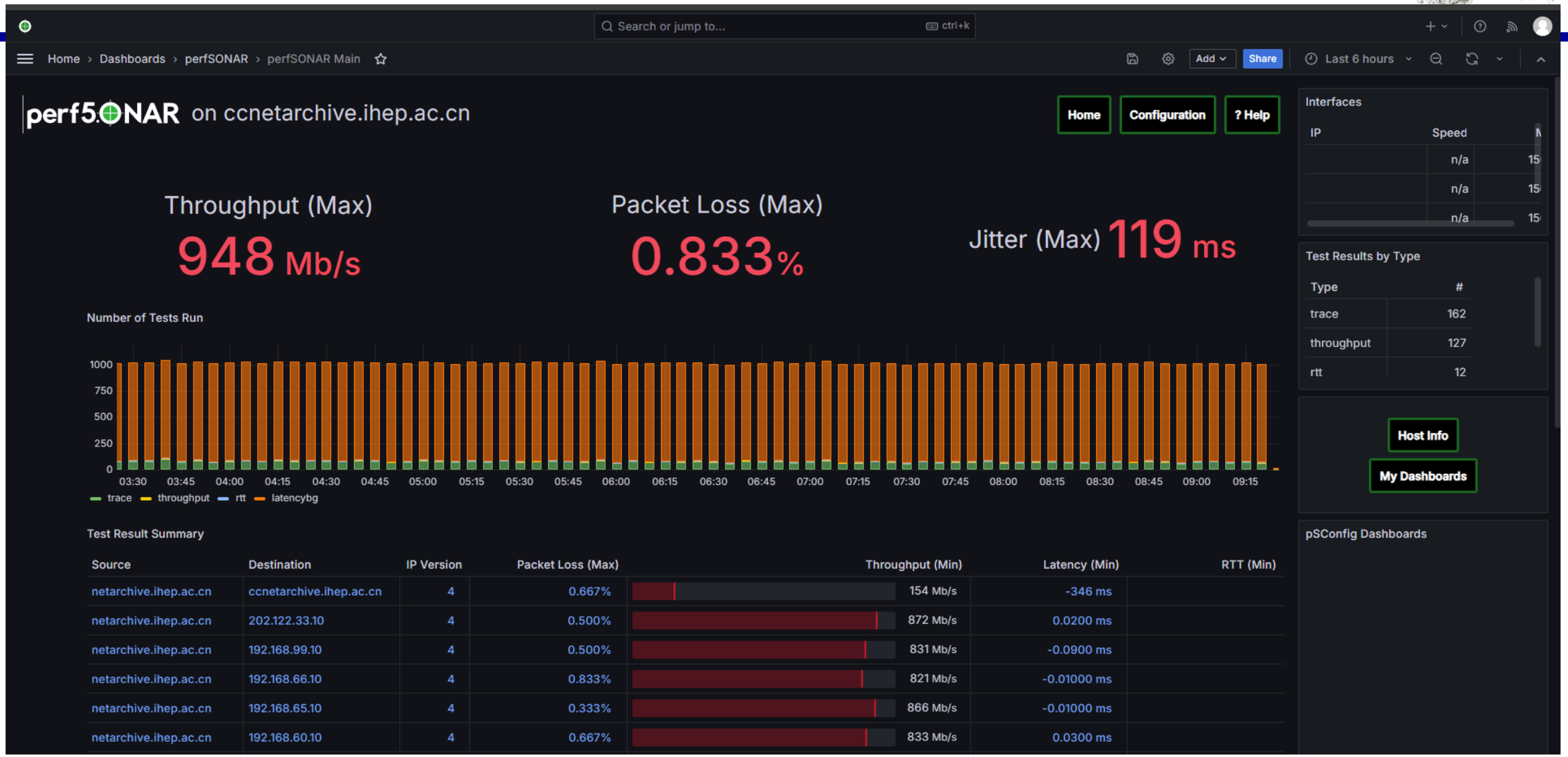
- Creation of over 20+ test nodes across the data center network
- Strategic distribution to cover all critical network domain
- Configuration: Detailed setup of each node, including software installation and test configurations



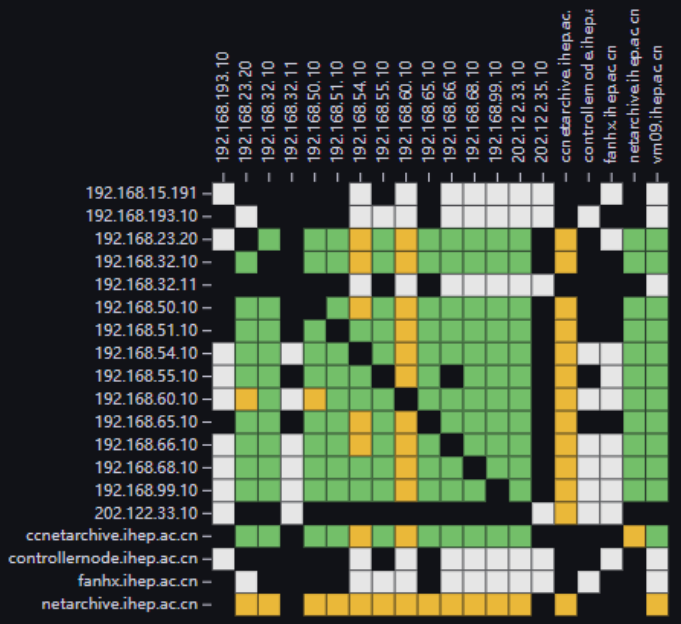
Network Performance Measurement Tests

- **Types of Tests**: Ping, traceroute, iperf3, owamp, bwctl, throughput
- **Test Frequency**: Continuous monitoring with periodic in-depth tests
- **Data Collection**: Real-time and historical data storage for analysis
- **Data Sources**: OpenSearch




Search or jump to... ctrl+k
Home › Dashboards › perfSONAR › perfSONAR Main
Add
Share
Last 6 hours
perf5ONAR on ccnetarchive.ihep.ac.cn
Home
Configuration
? Help
Throughput (Max)
948 Mb/s
Packet Loss (Max)
0.833%
Jitter (Max) 119 ms
Interfaces
IP
Speed
n/a
n/a
n/a
Test Results by Type
Type
#
trace
162
throughput
127
rtt
12
Host Info
My Dashboards
pSConfig Dashboards
Number of Tests Run
1000
750
500
250
0
03:30
03:45
04:00
04:15
04:30
04:45
05:00
05:15
05:30
05:45
06:00
06:15
06:30
06:45
07:00
07:15
07:30
07:45
08:00
08:15
08:30
08:45
09:00
09:15
trace
throughput
rtt
latencybg
Test Result Summary
Source
Destination
IP Version
Packet Loss (Max)
Throughput (Min)
Latency (Min)
RTT (Min)
netarchive.ihep.ac.cn
ccnetarchive.ihep.ac.cn
4
0.667%
154 Mb/s
-346 ms
netarchive.ihep.ac.cn
202.122.33.10
4
0.500%
872 Mb/s
0.0200 ms
netarchive.ihep.ac.cn
192.168.99.10
4
0.500%
831 Mb/s
-0.0900 ms
netarchive.ihep.ac.cn
192.168.66.10
4
0.833%
821 Mb/s
-0.01000 ms
netarchive.ihep.ac.cn
192.168.65.10
4
0.333%
866 Mb/s
-0.01000 ms
netarchive.ihep.ac.cn
192.168.60.10
4
0.667%
833 Mb/s
0.0300 ms

Maximum Packet Loss



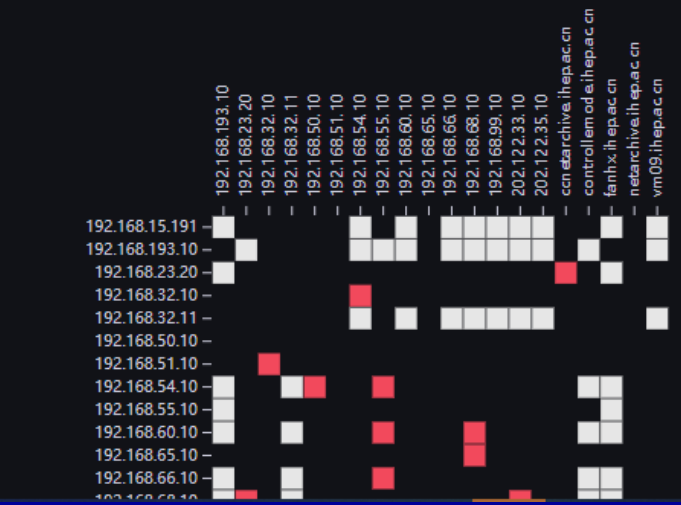
Highest Average Packet Loss by Sender

Sender Host	Max Packet Loss
192.168.55.10	0.500%
192.168.66.10	0.500%
192.168.23.20	0.333%
192.168.32.10	0.333%
192.168.50.10	0.333%
192.168.51.10	0.333%
192.168.54.10	0.333%
192.168.60.10	0.333%
192.168.65.10	0.333%
192.168.68.10	0.333%

Highest Average Packet Loss by Receiver

Receiver Host	Max Packet Loss
192.168.60.10	0.500%
ccnetarchive.ihep.ac.cn	0.500%
192.168.54.10	0.333%
192.168.55.10	0.333%
vm09.ihep.ac.cn	0.333%
192.168.23.20	0.167%
192.168.32.10	0.167%
192.168.50.10	0.167%
192.168.51.10	0.167%
192.168.65.10	0.167%

Minimum Throughput



Lowest Minimum Throughput by Sender

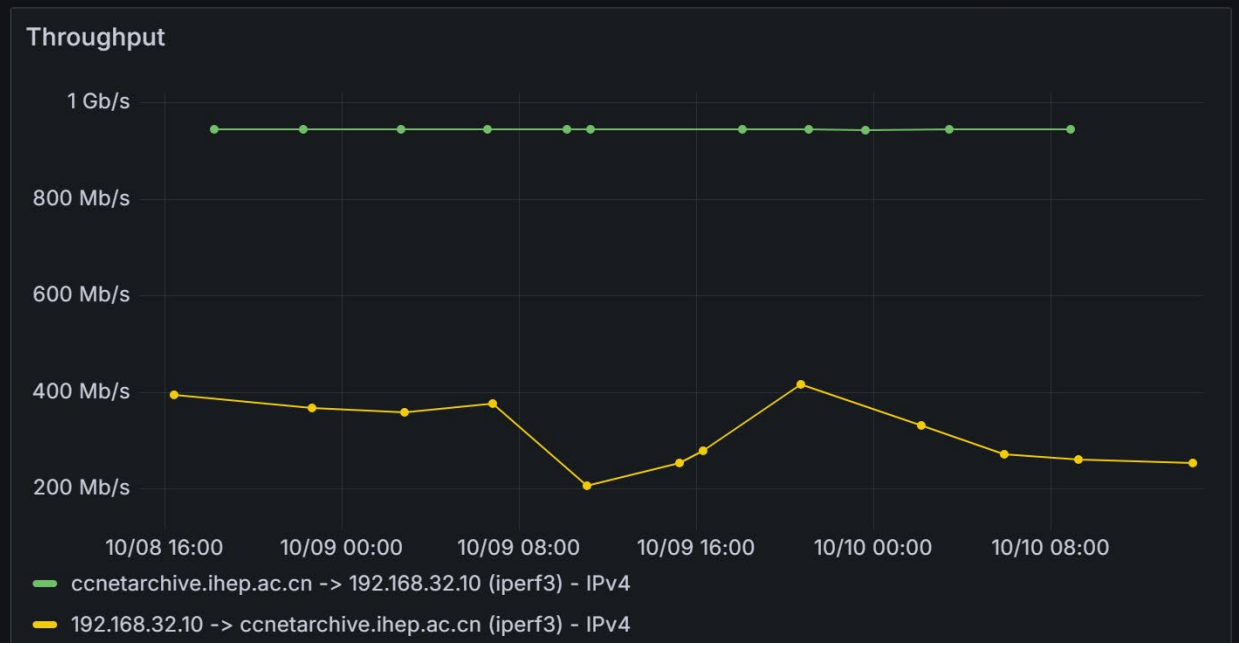
Sender Host	Receiver Throughput
192.168.23.20	287 Mb/s
192.168.68.10	932 Mb/s
192.168.60.10	932 Mb/s
192.168.54.10	934 Mb/s
192.168.66.10	935 Mb/s
ccnetarchive.ihep.ac.cn	935 Mb/s
192.168.65.10	937 Mb/s
192.168.99.10	937 Mb/s
192.168.51.10	937 Mb/s

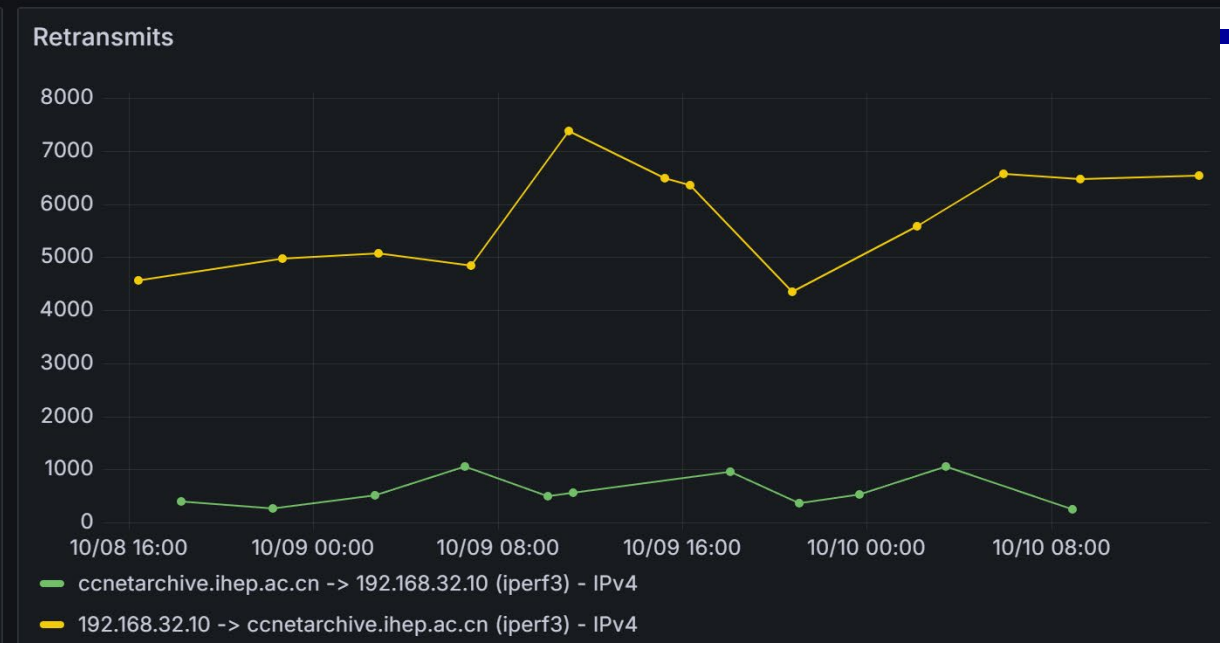
Lowest Minimum Throughput by Receiver

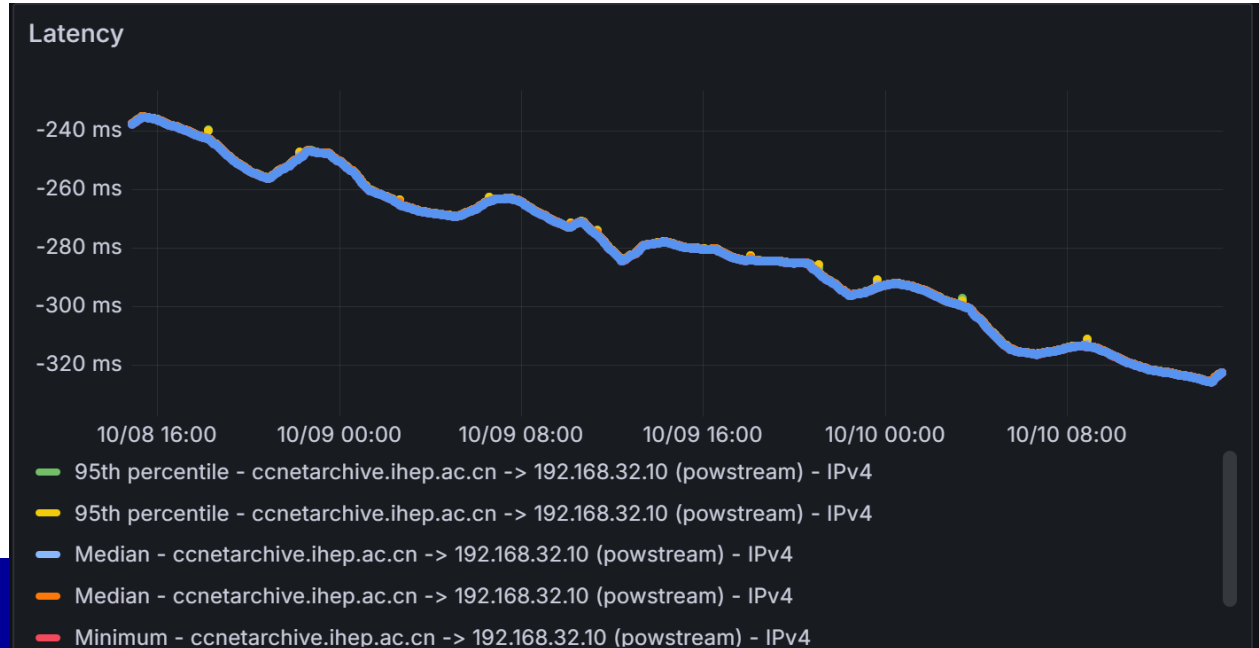
Receiver Host	Min Throughput
ccnetarchive.ihep.ac.cn	287 Mb/s
192.168.23.20	932 Mb/s
192.168.68.10	932 Mb/s
192.168.50.10	934 Mb/s
192.168.55.10	935 Mb/s
netarchive.ihep.ac.cn	935 Mb/s
202.122.33.10	936 Mb/s
192.168.32.10	937 Mb/s
192.168.54.10	941 Mb/s

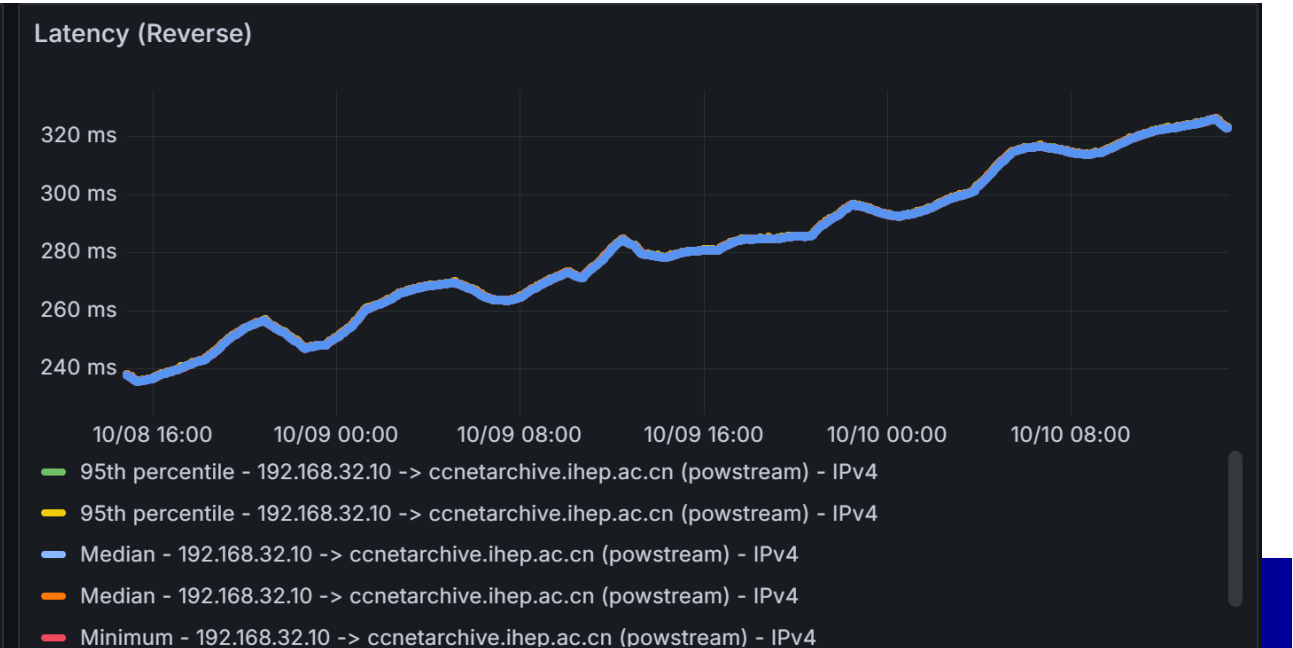
Add | Share | Last 2 days

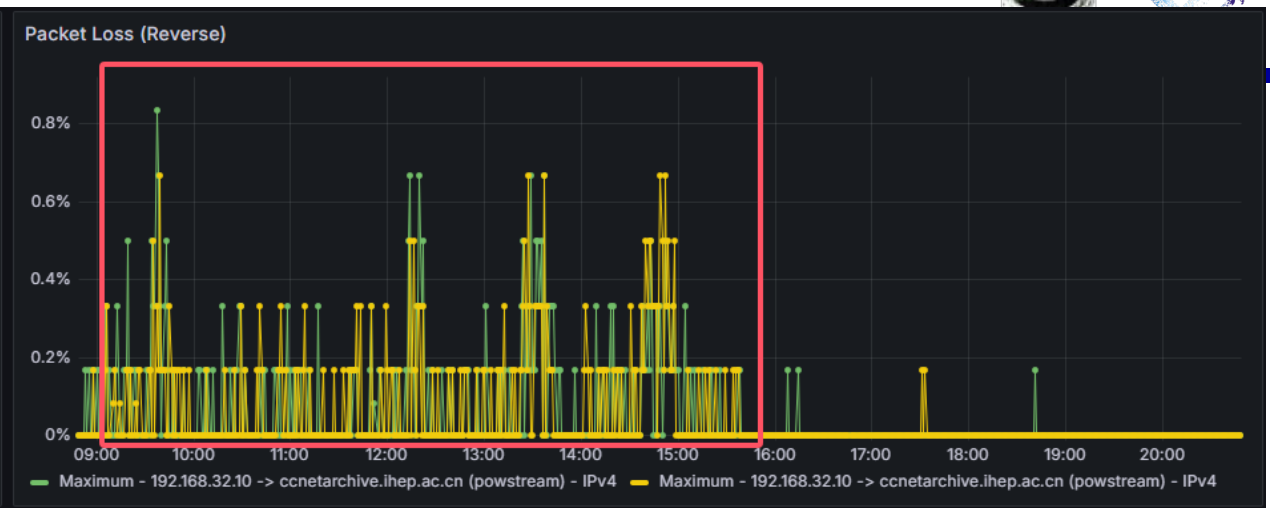
Source: ccnetarchive.ihep.ac.cn | Destination: 192.168.32.10 | Filters







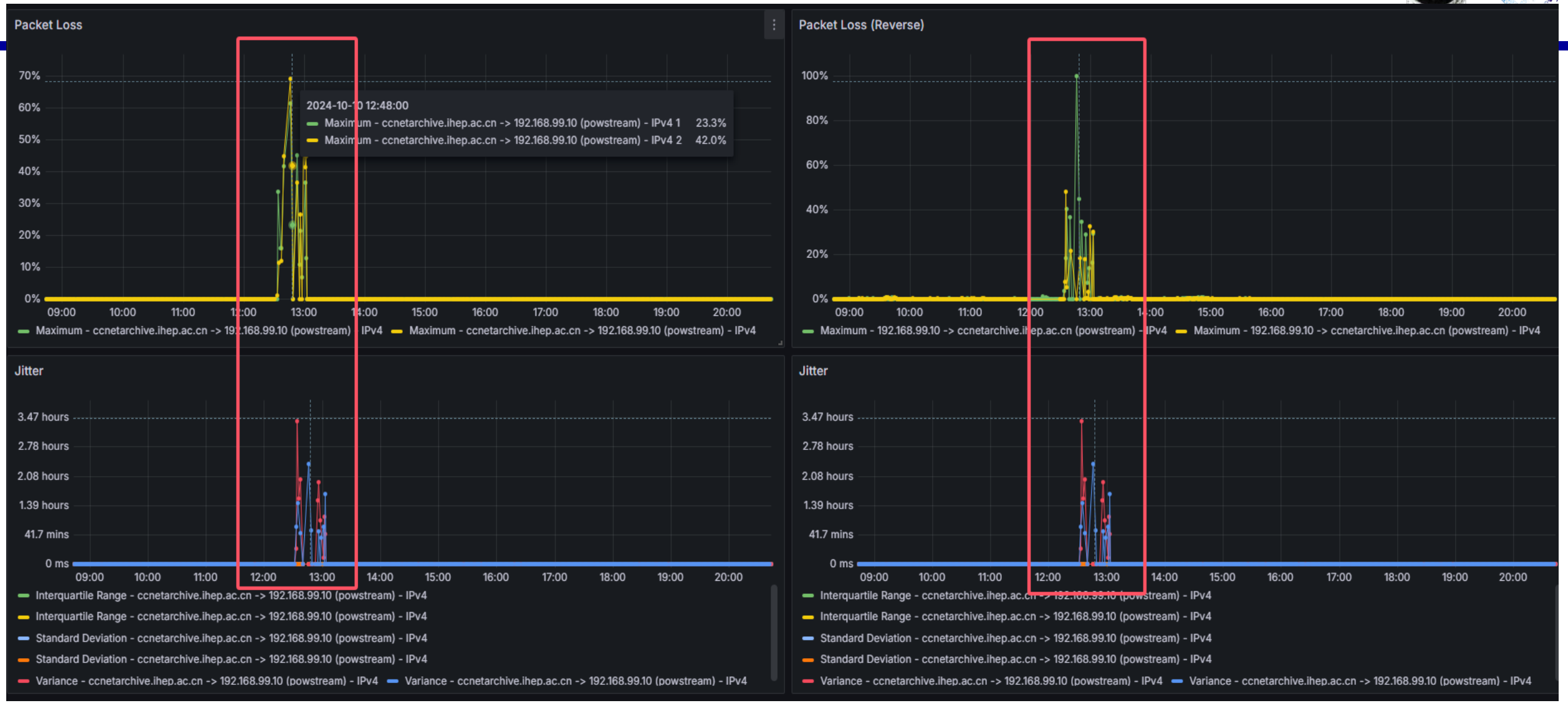





```
[root@perf-32-10 ~]# ifconfig
enp2s0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.32.10  netmask 255.255.255.0  broadcast 192.168.32.255
        inet6 2401:de00:1:2032::9669  prefixlen 128  scopeid 0x0<global>
        inet6 fe80::2e0:4cff:fe69:3cfa  prefixlen 64  scopeid 0x20<link>
        ether 00:e0:4c:69:3c:fa  txqueuelen 1000  (Ethernet)
        RX packets 3626365612  bytes 4785341488435 (4.3 TiB)
        RX errors 0  dropped 414224  overruns 0  frame 0
        TX packets 3843783958  bytes 5161542930266 (4.6 TiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

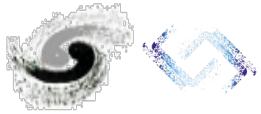
net_ipv4_tcp_rmem_default	87380
net_ipv4_tcp_rmem_max	33554432
net_ipv4_tcp_rmem_min	4096
net_ipv4_tcp_wmem_default	65536
net_ipv4_tcp_wmem_max	33554432
net_ipv4_tcp_wmem_min	4096

204800
33554432
102400
102400
33554432
51200


Packet Loss
70%
60%
50%
40%
30%
20%
10%
0%
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
2024-10-10 12:48:00
Maximum - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4 1 23.3%
Maximum - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4 2 42.0%
Maximum - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Maximum - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Packet Loss (Reverse)
100%
80%
60%
40%
20%
0%
Maximum - 192.168.99.10 -> ccnetarchive.ihep.ac.cn (powstream) - IPv4
Maximum - 192.168.99.10 -> ccnetarchive.ihep.ac.cn (powstream) - IPv4
Jitter
3.47 hours
2.78 hours
2.08 hours
1.39 hours
41.7 mins
0 ms
Interquartile Range - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Interquartile Range - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Standard Deviation - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Standard Deviation - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Variance - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Variance - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Jitter
Interquartile Range - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Interquartile Range - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Standard Deviation - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Standard Deviation - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Variance - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4
Variance - ccnetarchive.ihep.ac.cn -> 192.168.99.10 (powstream) - IPv4

Home › Dashboards › perfSONAR › perfSONAR Host Metrics
Add
3 GB
2 GB
1 GB
0 B
09:00 09:30 10:00 10:30 11:00 11:30 12:00 12:30 13:00 13:30 14:00 14:30 15:00 15:30 16:00 16:30 17:00 17:30 18:00 18:30 19:00 19:30 20:00 20:30
Available
Total
Average CPU Utilization
100%
80%
60%
40%
20%
0%
09:00 09:30 10:00 10:30 11:00 11:30 12:00 12:30 13:00 13:30 14:00 14:30 15:00 15:30 16:00 16:30 17:00 17:30 18:00 18:30 19:00 19:30 20:00 20:30
CPU Utilization

Lessons Learned



- **1: Latency Reduction**
  - **Scenario**: Elevated latency between data servers and analysis workstations
  - **Analysis**: Used perfSONAR data to isolate the problematic segment
  - **Solution**: Upgraded network hardware
- **2: Packet Loss Mitigation**
  - **Scenario**: Frequent packet loss during high-data-rate transfers
  - **Root Cause**: Faulty interface on a tor switch
  - **Resolution**: Replaced the faulty hardware and implemented redundancy
- **Importance of Regular Updates**: Keeps the system secure and functional
- **Comprehensive Testing**: Ensures reliability before full-scale deployment
- **Collaborative Effort**: Engaging with the perfSONAR community for support and best practices

Future Plans

Expanding Network Coverage

- **Network Segment Expansion**: Plan to include more network segments in the monitoring system
- **Enhanced Visibility**: Fuller figure of network health and performance

Scalability Considerations

- **Horizontal Scaling**: Adding more test nodes as the network grows
- **Vertical Scaling**: Upgrading existing nodes with more powerful hardware

Enhancements to Monitoring Capabilities

- New features to be integrated into the monitoring system
- Plans for integrating AI and machine learning for predictive analysis
- Enhancing visualization capabilities

Summary

- Overview of the perfSONAR deployment at IHEP data center
- Network Performance Analysis System works well
- Positive impact on network performance and research productivity
- Optimistic future with ongoing improvements and expansion