LATENCY-BASED ROUTING SOLUTION BRIEF

Monetize Differentiated Latency-Critical Connectivity And SLAs While Reducing Costs, With Automated, Real-Time Routing Based On True, End-To-End Latency

Challenge
Low-latency connectivity requires real-time, end-to-end network and service performance intelligence throughout the service chain, even in domains that you don't control. Visibility blind spots and lack of automation make SLA-backed latency guarantees neither technically nor commercially viable.

Solution
Latency-based routing is an integrated, end-to-end solution for automating latency-guaranteed service delivery. It monitors and correlates end-to-end data-plane latency, for full service-chain visibility. In seconds, it automates latency anomaly detection and remediation, re-routing and continuously maintaining optimal latency across all live services.

Benefits
Fully autonomous for lower OPEX: no human intervention needed
Visualize and control end-to-end and hop-by-hop latency through the entire service chain
Simple and intuitive, lowering the skills-barrier for managing latency-critical services

Consistent low-latency connectivity is a critical factor across an increasing number of use cases, and the end user is willing to pay for the improved performance. Today, gamers, FX traders, and live broadcasters need this connectivity; tomorrow, autonomous vehicle drivers, factory operators, and metaverse users will desire it. The latency you can provide your customers makes the difference between winning and losing new subscribers. And whether it's 5G slicing, fiber to the x (FTTX), or any other access technology, guaranteed low-latency services can't be delivered without latency-aware Autonomous Transport Networking.

Low-latency service delivery is commercially lucrative—and technically challenging. You need real-time visibility and control across the end-to-end service chain, even if you don't control every domain. Unlike traditional services, latency issues get immediately noticed. So, for operators with SLAs, the stakes are high. If you get it right, the rewards are massive. Latency-based routing combines market-leading path computation, streaming telemetry, and real-time, data-plane latency monitoring. Proven in the world's most demanding networks, it's a simple, intuitive, and easy-to-deploy service that delivers rapid time-to-value.

The Challenge

Low-latency connectivity is a high-risk, high-reward value proposition. While it can command a premium price for service providers, it leaves no room for error. In low-latency scenarios such as gaming, trading, conferencing, and AR/VR, latency can make the difference between success or failure for the end user. Delivering low latency services therefore requires real-time, automated, end-to-end, latency-aware routing, because it is not possible for a human to react quickly enough to changing network conditions.

Three key challenges stand in the way of achieving real-time, automated, end-to-end, latency-aware routing: lack of integrated end-to-end and hop-by-hop latency awareness; manual network operations; and ineffective DIY network automation systems.

You can't act on what you can't see. End-to-end and hop-by-hop visibility is vital to detecting and resolving transport network issues, and this is particularly critical for low-latency connectivity. An accurate, real-time view of latency, both end-to-end (from the end-user to the service) and hop-by-hop, is essential for monitoring latency as perceived by the end user, and quickly isolating, addressing, and resolving the root-causes of latency spikes. This can't be done solely with end-to-end visibility, nor solely with hob-by-hop visibility. It requires an integrated, correlated view of both. For example, if you can detect a latency spike in a gaming service, you're nevertheless powerless to do anything about it if you can't identify the offending node(s) and link(s). Meanwhile, you could detect a latency spike in a particular hop, but have no visibility of whether this is actually impacting the customer.

With flexibility comes complexity. Today, transport networks are programmable to such an extent that the range of configurations, both active in the network and available to your operations team, is so great as to render network optimization through manual operations alone impossible. And yet today, still 82% of networking activities are largely manual[1]. That leads to sub-optimal configurations, errors, slow response times, and ultimately unreliable networks. This is particularly mission-critical for delivering low-latency services because customer-impacting problems are instantly noticed, so require instant action.

Network automation systems are difficult to get right. They take months, or even years, to build, and 70% of DIY network automation projects fail. Latency-based routing is a casualty of this, as it is a use case that relies heavily on modern network automation technology.

Latency-based routing makes the monetization of low-latency services possible at scale by automating network operations for low-latency connectivity services, removing the dependency on manual operations, and reducing the time to detect and resolve latency issues to seconds or less. It combines real-time, integrated end-to-end (active data plane) and hop-by-hop (device telemetry) visibility, with powerful path computation and automated routing. It leverages the TWAMP-Light protocol to produce periodic measurements of latency between network nodes, and an advanced path computation engine (PCE) to automatically re-route Label Switch Paths (LSPs) when a change in latency is reported via BGP-LS protocol. Such a change in latency can have numerous root causes, including fiber cuts, a re-routing of an underlying wholesaler network, or unexpected bandwidth congestion in a router interface.

Juniper® Paragon Automation enables latency-based routing that provides your operations teams with simple-to-navigate and color-coded dashboards and reports, including graph views of historical latency through time-series data, to guide them through the root-cause identification process, right down to the most granular level. Misconfigurations, misbehaving processes, power outages, failed software updates, overheating; wherever the problem is, your team is empowered to quickly identify and fix it.

When the problem is resolved, latency-based routing can automatically identify this, triggering another re-computation and re-routing, so that traffic can again pass through the now fully operational network elements.

[1] Heavy Reading, 2023
[2] Analysys Mason, 2022

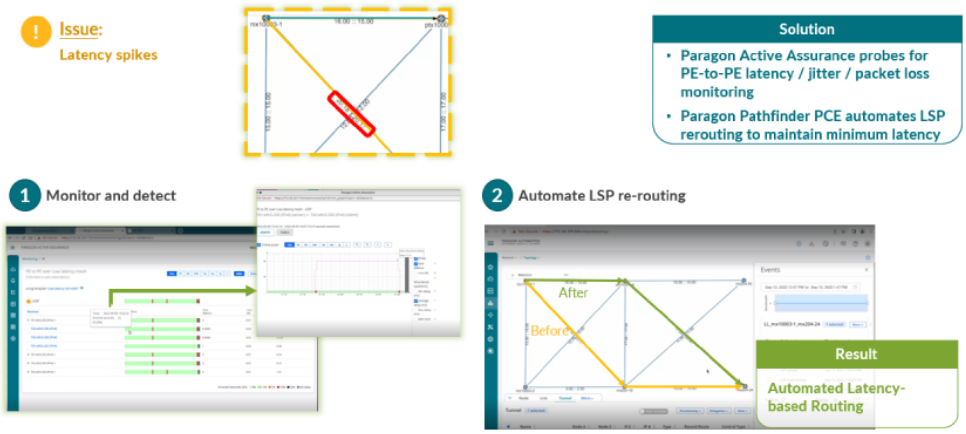

Figure 1: An overview of the processes and capabilities involved in latency-based routing.

Features and Benefits

Table 1: Features and Benefits of latency-based routing

Features	Benefits
Real-time, end-to-end network observability based on device telemetry and active performance metrics. Active performance is measured by sending synthetic traffic on the data plane, from Layer 2-7, to simulate end-user sessions, for continuous end-to-end and per-segment monitoring of customer experience KPIs like delay, jitter, and throughput.	Eliminate blind spots by detecting user-plane performance degradations, not just network resource problems. Proactively detect service quality degradations and SLA violations to resolve issues before customers are impacted.
Extensive use of standardized protocols, including PCEP, gRPC, and BGP-LS for collection of topology data, LSP states, and traffic loads.	Manage multivendor networks with a single solution, based on modern, industry-accepted standards, proven by rigorous, independent multivendor interoperability testing[3]
Powerful, SLA-driven, PCE, that leverages the "active stateful PCE" concept as defined by IETF RFC 5440, supports dynamic topology discovery, and provides automated global optimization of end-to-end network paths (LSPs).	Maintain service integrity during network outages by quickly re-calculating all paths for optimal SLA performance. Enhance network reliability by taking into account "Shared Risk Link Groups" (SRLGs) in routing computation.
Powerful, SLA-driven, PCE, that leverages the "active stateful PCE" concept as defined by IETF RFC 5440, supports dynamic topology discovery, and provides automated global optimization of end-to-end network paths (LSPs).	Maintain service integrity during network outages by quickly re-calculating all paths for optimal SLA performance. Enhance network reliability by taking into account "Shared Risk Link Groups" (SRLGs) in routing computation.

Features	Benefits
Multivendor support with open programmability based on standardized OpenConfig and YANG models.	Implement closed-loop remediation in multivendor networks.
Support for multiple routing protocols, including RSVP and segment routing protocols such as SR-MPLS and SRv6.	Future-proof your closed-loop remediation capability as your routing infrastructure evolves.

[1] EANTC Multi-Vendor Interoperability test report, 2023.

Solution Components

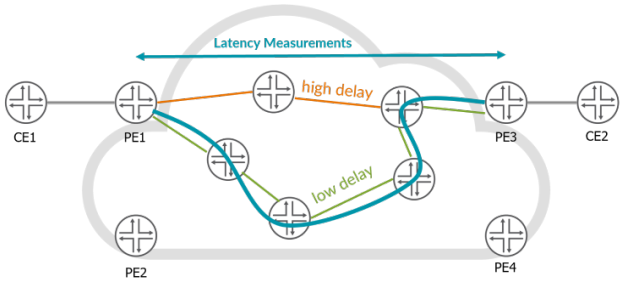

Figure 2: Juniper Paragon Pathfinder re-routes traffic in response to changing latency conditions.

Latency-based routing is supported by an intuitive user interface. It combines streaming telemetry with user-plane active testing and monitoring; AI-powered analytics and automation; and real-time path computation. These include the following capabilities to enable the latency-based routing use case.

Real-time path computation: Provides a powerful IETF-compliant path computation element, which calculates and deploys optimized end-to-end traffic paths in your transport network to completely avoid problematic routers when triggered by an outage, or a precursor anomaly or event. The multi-vendor, cross-domain path computation engine supports optimization policies including latency, bandwidth, packet-loss, and Traffic Engineering metrics and disjoint paths. Other key capabilities include automation of Shared Risk Link Groups (SRLG) path diversity, and bandwidth calendaring.

Active assurance for data-plane testing and monitoring: Provides active data-plane KPI monitoring and on-demand measurements, by deploying software-based test agents across routers in your network, as well as in third-party domains, such as cloud service providers. Alongside streaming telemetry, it provides enhanced, real-time network observability and forms the basis for remediation based on perceived customer experience. It complements telemetry data by providing visibility of customer-impacting issues where telemetry data alone may not. It also enables proactive monitoring: measuring performance using synthetic traffic, to simulate network usage even if there is no live traffic.

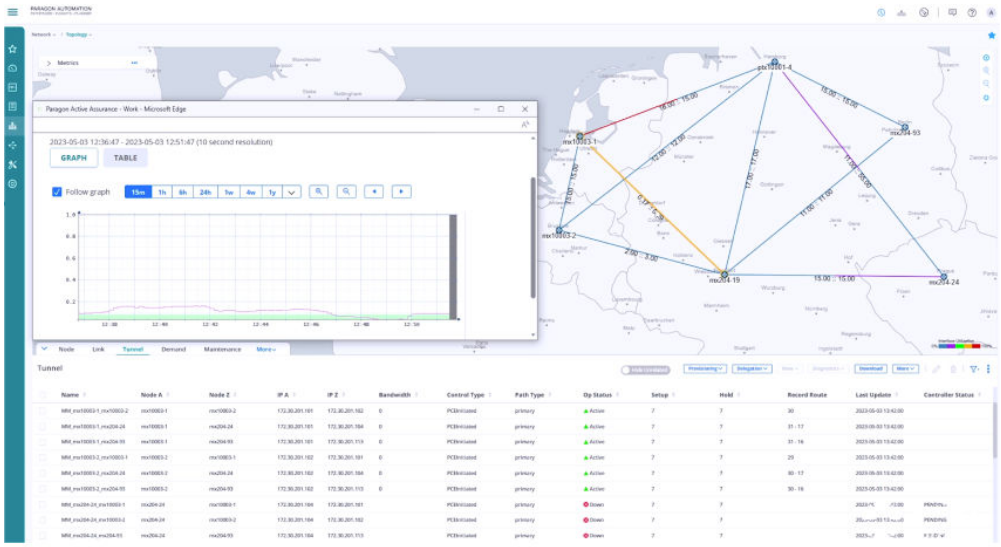

Figure 3: An example of routing based on correlated end-to-end latency and streaming telemetry.

Summary–Better Experiences, Without the Complexity

Based on Paragon Automation's fully cloud-native architecture, latency-based routing provides you the confidence to offer and monetize guaranteed low-latency services. It can be rapidly configured to cater to specific network requirements, such as managing unique enterprise SLAs, from a single, intuitive user interface. It empowers staff, so you can grow your business without compromising network performance.

Latency-based routing is proven to significantly enhance network performance and reliability, unlocking new, lucrative low-latency services while reducing costs and reliance on skilled staff. Ask us for a demo today.

Next Steps

- Watch our Tech Field Day demo on latency-based routing
- Read this blog post describing how the solution applies in an example scenario
- Contact your Juniper account representative to schedule a demo today!

Corporate and Sales Headquarters

Juniper Networks, Inc.

1133 Innovation Way

Sunnyvale, CA 94089 USA **Phone:**

888.JUNIPER (888.586.4737)

or +1.408.745.2000

www.juniper.net

APAC and EMEA Headquarters

Juniper Networks International B.V.

Boeing Avenue 240

1119 PZ Schiphol-Rijk

Amsterdam, The Netherlands **Phone:**

+31.0.207.125.700