Real-time Network Visibility

Data Center Visibility and Analysis in Real Time using Mellanox's Spectrum™-based Open Ethernet Switches

AN OUNCE OF PREVENTION IS WORTH A POUND OF CURE

When network capacity fails to meet demand, frustrated IT managers and customers complain of slow network responses to web-based applications and the Internet. A sluggish network also affects critical IT operations and can result in slow backups and data corruption resulting from dropped packets. If the issue isn't resolved, business continuity and system availability may be compromised. If mission-critical customer-facing applications are impacted, performance issues can have an adverse effect on brand perception and customer loyalty.

While the symptoms of a poorly performing network are easy to see, pinpointing the underlying causes of network issues is not nearly so straightforward. The size, complexity, speed and dynamic nature of network traffic can make it difficult to identify bottlenecks or other issues.

When it comes to network downtime, there are numerous figures floating around the industry:

According to Ponemon Institute mean downtime cost is around $8,800 for each minute (derived from 63 data centers located in the United States). That means that a network outage lasting only a single hour can cost a business more than half a million dollars.

A separate report from IHS Inc., a leader in global information and analysis, found that midsize to large companies typically experience five minutes of downtime every month, which translates to a cost of about $1 million annually for midsize firms and $60 million a year for large enterprises.

Only by putting a solution in place to monitor network activity and measure application response times can network managers and administrators hope to proactively address poor application performance before it impacts the end user. **Simply put, you can't manage what you can't measure.** A network with increased and real-time visibility enables better reliability and real-time control.

KEY USE CASES

- Link and fabric-level latency analysis
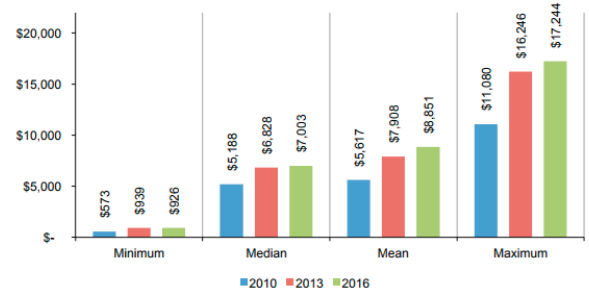- End-to-end network debugging
- Real-time congestion detection

Figure 1. Ponemon Report on Network Downtime Cost

†For illustration only. Actual products may vary.

PUSH, PUSH, PUSH

For the last decades, network operators have heavily relied on SNMP polling and CLI screen-scraping to extract operational data from the network. An application polls the host CPU to gather aggregate telemetry every few seconds or minutes. Such a method doesn't scale well or supply the full picture of today's network behavior. And as we move to higher density platforms, the amount of important operational data becomes truly staggering. Instead of pulling data off the network, sit back and let the network **push** it to you.

TELEMETRY OVER SPECTRUM

Network telemetry provides the linkage between the network infrastructure and critical business application performance that ensures visibility into critical real-time information. It dramatically reduces application downtime and network operations costs through improved real-time system and network visibility (streaming), improved debug capabilities and real time alerts correlating to application behavior and advanced end-to-end path monitoring tools.

Organizations using a Mellanox Spectrum switch as a telemetry vehicle, together with an analytic tool such as Cumulus NetQ or Grafana, can truly scale to the next level of visibility and debug-ability of their fabrics. The Spectrum switch enhanced visibility capabilities include:

Real-time monitoring using 8 smart mirror engines

- Allows moving from reactive to proactive mode using mirror engines, which enables user flexibility and maximizes fabric optimizations.
- Spectrum switches can mirror in **real-time** with **nanosecond granularity** the following events: congestion control (WRED, ECN), drop occurrences, high latency, ACL flows, and even buffer levels.
- Alerts based on buffer levels can pro-actively signal the IT manager that a certain queue is about to be congested and packets might drop. Such critical alerts can make a big difference to users who are sensitive to packet loss.

Enhanced debug capabilities using statistics, triggered watermarks and buffer status snapshot

- Histograms summarize buffer dynamics and reduce the size of a raw input data to a set of counts, which are divided across a set of buckets. The Spectrum switch measures the queue depth up to 16 million times per second (rate of 64ns) and reflects the occurrences counter in 10 different pre-defined buckets.

  In Figure 2, the IT manager can **quickly identify** a microburst at time 19:19:49. Some of the samples show a momentary high level of buffer occupancy. Starting from 19:19:54 the queue is fully utilized all the time, implying congestion.
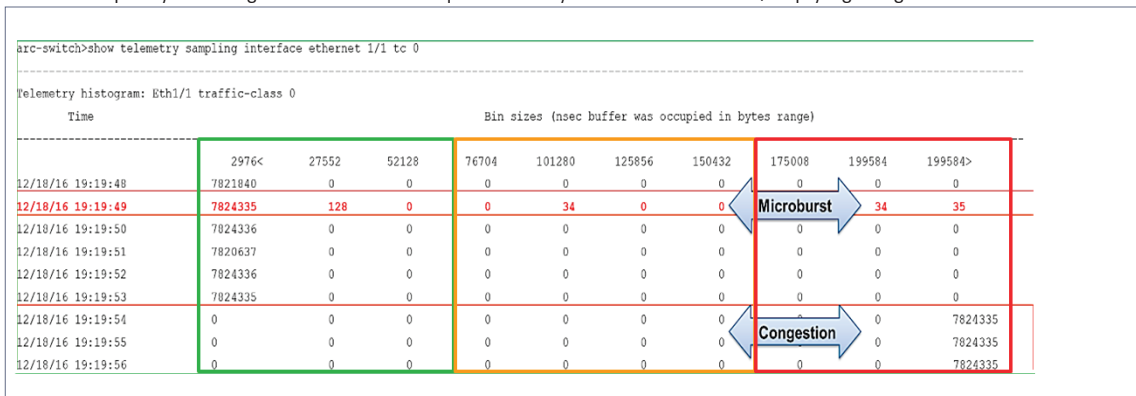
```
arc-switch>show telemetry sampling interface ethernet 1/1 tc 0

Telemetry histogram: Eth1/1 traffic-class 0
     Time                              Bin sizes (nsec buffer was occupied in bytes range)

                2976<    27552    52128   76704   101280  125856  150432   175008   199584   199584>
12/18/16 19:19:48  7821840     0        0       0       0       0       0      0        0        0
12/18/16 19:19:49  7824335    128       0       0       34      0       0   |Microburst|  34    35
12/18/16 19:19:50  7824336     0        0       0       0       0       0      0        0        0
12/18/16 19:19:51  7820637     0        0       0       0       0       0      0        0        0
12/18/16 19:19:52  7824336     0        0       0       0       0       0      0        0        0
12/18/16 19:19:53  7824335     0        0       0       0       0       0      0        0        0
12/18/16 19:19:54     0        0        0       0       0       0       0  |Congestion|  0    7824335
12/18/16 19:19:55     0        0        0       0       0       0       0               0    7824335
12/18/16 19:19:56     0        0        0       0       0       0       0      0         0    7824335
```

Figure 2. Telemetry histogram example

  The microburst histogram monitoring allows the IT manager to specify absolute limit values both for the data and bursts size where simple monitoring tools, such as SNMP or CLI polling, are just not accurate or quick enough to detect.

- Events may be triggered once crossing a predefined threshold or low/high thresholds (hysteresis). An event could be a mirrored packet, snapshot of the buffers' occupancy or pure traffic log streaming.
- **A snapshot of the entire switch buffer's status** may be generated according to predefined events, such as a packet drop. Shared buffer is the best way to utilize the switch's resources, however it also introduces new challenges to network debugging as resource allocation is now more fluid. For example, assume that an application suffering from low performance, due to packet drops, ends up with an unhappy customer. This application doesn't consume much traffic and uses short messages, which don't pass any buffer watermark. How can the IT manager debug the fabric?

  A snapshot of the buffer queue allocation will allow the IT manager to quickly determine that another flow abused the shared buffer and consumed all resources, causing starvation to the poor-performing application. Modifying shared buffer rules can bring to an immediate performance boost.

Smart flow estimation engines

- Network telemetry can be considered a gold mine for researchers who are performing data mining, QoS, security forensics or malware spread. The Spectrum-2 switch supports unique smart flow estimation engines which continuously monitor application behavior and can **identify any deviation in the communication pattern** once it occurs. One such example could be an anomaly in the fabric due to a DDoS attack.

In-band telemetry or in-situ OAM

- Spectrum-2 in-band telemetry defines a mechanism for adding telemetry data to the data plane packets without consuming any host CPU. This metadata info can include Switch-ID, per queue latency and more. It complements out-of-band probe mechanisms such as ping or traceroute (see Figure 3).
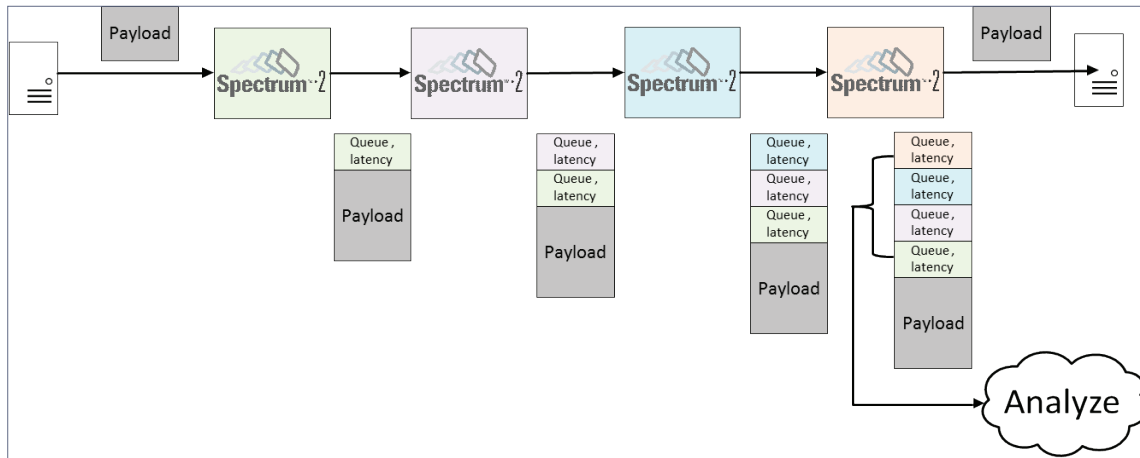

Figure 3. In-band telemetry

- Using in-band telemetry, it becomes easier to **analyze when a packets enters and exists a switch**, at what rate packets arrive at a particular switch and how long a packet spent at each hop. The embedded information can be added in various layers supporting encapsulation options such as Geneva and VXLAN GPE.

REAL-LIFE USE CASES

Low Latency Trading

With the advent of algorithmic trading, there is a critical need for low latency trading platforms, where, for example, when executing arbitrage strategies, every millisecond lost could result in ~$100M lost opportunity. Hence, speed becomes the key differentiator, making it absolutely necessary to establish the lowest possible latency between processing environments. Low latency trading firms require an agile and high-performing infrastructure that addresses the enormous volumes and surges of high rate data speeds which are typically associated with periods of market volatility and latency spikes that must be monitored and detected in real-time.

Figure 4. Volumes and surges of high rate data speeds

A microburst is a traffic pattern that causes short-lived network congestion. This pattern often occurs during a phase of high activity, such as a highly volatile period in the market, which causes network endpoints to send traffic bursts into the network. Dropping even a single packet of information due to a microburst exceeding the network bandwidth, can dramatically increase trading time, requiring retransmissions and possibly resulting in complete data loss and, therefore, transaction loss. Congestion characterization at micro-second granularity, such as understanding the arrival time and duration of a burst, can help the operators to better optimize their network and guarantee higher SLA.
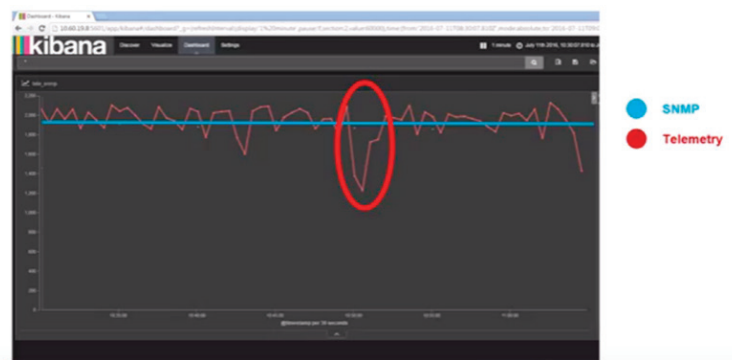
Media Streaming

Quality of Experience (QoE) is a measure of the delight or annoyance of the customer's experience with a service (web, VoIP or broadcast TV). QoE focuses on the entire service experience; it is a holistic concept, similar to the field of User Experience, but with its roots in telecommunication. Poor bandwidth, delays and microbursts glitches are 'persona non grata' for these kind of solutions. IT managers need ability to detect, analyze and react in real time to ensure high quality whenever needed.

In-band Telemetry to Enable End-to-End Network Debugging

Overlay technologies and XaaS markets have increased management complexity while reducing the visibility into traffic flows. Operators have been demanding the ability to monitor packet flows across the fabric, and to isolate underlay or overlay troubleshooting to improve SLA compliance.

In-band telemetry can provide service and quality assurance with service/path verification to ensure KPIs for overlay networks. Using the in-band metadata Switch ID per hop that is aggregate through the end-to-end path gives the operator a true story of the underlay data flow.

In the Figure 5 example:

1. The performance of the original path from switch A to switch C through switch B, was good enough to meet application SLA.
2. At time 't', the path was re-routed through switch D; now the application suffers from poor performance.
3. The link from switch A to switch D suffers from low bandwidth or a bad cable, or even congestion in switch D.

Using end-to-end in-band telemetry and Switch-ID metadata, we are able to quickly identify the performance degradation and ensure SLA performance.
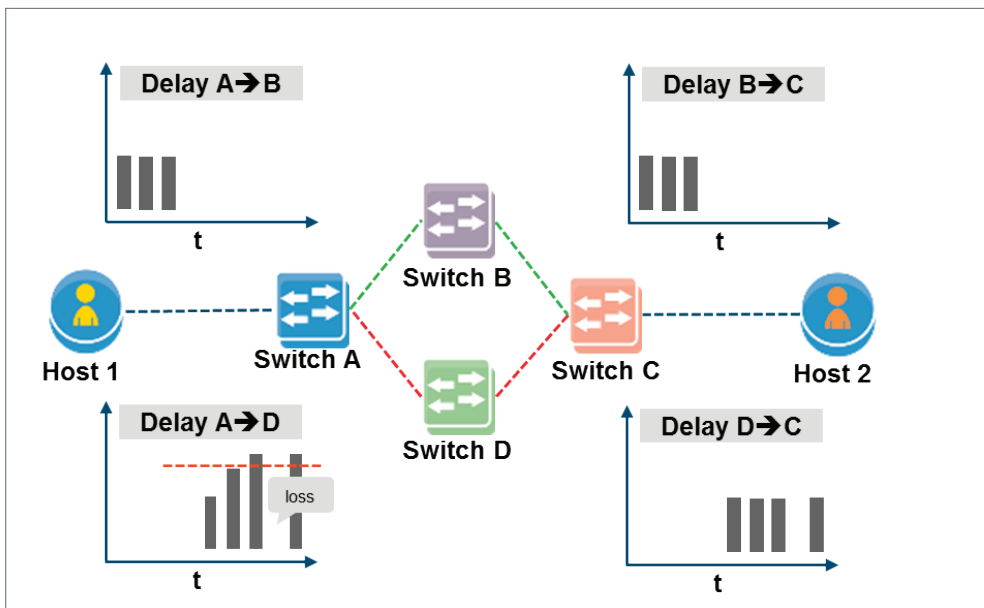

Figure 5. In-band telemetry to enable end-to-end network debugging

WHY MELLANOX SPECTRUM?

Mellanox Spectrum has several advantages over the competition which no other switch solution or vendor has.

First, Mellanox's Spectrum-based open Ethernet switches are the only switches that support a full telemetry feature-set on all system speeds, from 10GbE to 200GbE.

Second, Spectrum's unique benefits include:

- Histogram counters that can count up to 10 minutes
    - The 8 bits used by the competitors at a 64ns sampling rate saturates after a few micro-seconds
- Support for both unicast and multicast flows, with no compromise
- Only switch in the market to SPAN on ECN
- Only switch in the market to support watermarks and histograms, not just per port but also per traffic control (TC) Atomic buffers snapshot driven by hardware events
- Smart flow estimation engines

YOU CAN'T MANAGE WHAT YOU CAN'T MEASURE

IT security and analytic tools are only as good as the data they see. IT's core challenge is to ensure that the infrastructure supporting these tools delivers applications that are reliable, fast and secure. This means that IT needs total visibility into the network. Business continuity with high QoE requires a real-time, fast, accurate overlay and underlay, as well as distinct and quickly debugged fabrics to ensure the network is fully covered. Mellanox Spectrum switches are the only solution in the market to offer a full telemetry package to support speeds from 10GbE up to 200GbE with a full set of features and unique capabilities, with no compromise on performance.

You need real-time network visibility with no compromise on performance!

You need a Spectrum switch!

NEXT STEPS...

Learn more about: Open Ethernet

Learn more about: Mellanox Ethernet Switches in 3 minutes

Join the Mellanox Community

About Mellanox

Mellanox Technologies (NASDAQ: MLNX) is a leading supplier of end-to-end Ethernet and InfiniBand intelligent interconnect solutions and services for servers, storage, and hyper-converged infrastructure. Mellanox intelligent interconnect solutions increase data center efficiency by providing the highest throughput and lowest latency, delivering data faster to applications and unlocking system performance. Mellanox offers a choice of high performance solutions: network and multi-core processors, network adapters, switches, cables, software and silicon, that accelerate application runtime and maximize business results for a wide range of markets including high performance computing, enterprise data centers, Web 2.0, cloud, storage, network security, telecom and financial services. More information is available at www.mellanox.com.

350 Oakmead Parkway, Suite 100, Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com