Arista LANZ Overview

Overview

Arista Networks' Latency Analyzer (LANZ) represents the next step in the revolution in delivering real-time network performance and congestion monitoring. For the first time, administrators and applications can now achieve real-time visibility into network traffic patterns and their effect on congestion, application latency and performance. Rather than wait for the effects of network congestion to trickle up to the application layer, LANZ can pro-actively detect impending congestion events allowing preemptive capacity planning before the effect is seen at the application layer. The network administrator's dreaded support call of "the network is slow", can now be answered by the forensic analytic data provide by Arista's LANZ.

The benefits Arista's LANZ functionality will provide to network administrators:

- Real time visibility of congestion hotspots at the microbursts level
- Pre-empt network conditions before they induce latency or packet loss
- Isolate potential bottleneck early, enabling pro-active capacity planning
- Historically trending of network congestion events and their correlation to specific application traffic patterns.

Arista's Latency Analyzer (LANZ) is a pro-active event driven solution, purposely designed to provide real-time visibility of congestion hot spots and their effect on application performance and latency at a nanosecond resolution.

The Drive for LANZ

Network engineers design High Performance Computing (HPC), Data Center (DC) and High Frequency Trading (HFT) environments, with deterministic characteristics with regard to throughput, performance and latency. These characteristics will only be theoretical if the network infrastructure can't provide real time monitoring of changing traffic patterns and their affect on the design characteristics.

Traditional tools used to monitor network traffic patterns, such as RMON and SNMP, have been based on a polling model were data is typically collected at one second or longer intervals. What about the events that will occur within these polling intervals? With the evolution to 10Gbe attachment in the Data Center, within even a one second interval a 10Gbe interface could go from idle to forwarding over 28 million packets and back again. In a polling model this 28 million packet burst can become invisible. Decreasing the polling interval can be seen as a potential solution, however while also increasing the processing overhead the underlying hardware will typical only retrieve data at one second intervals, so the actual data collected would provide no extra visibility.

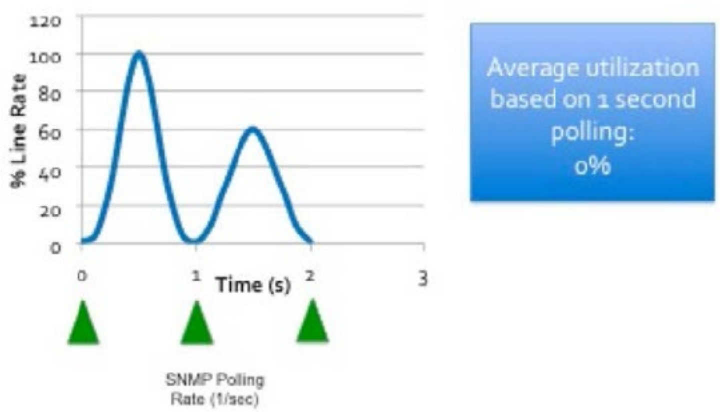

Figure 1: Polling model, fails to provide the correct level of granularity

The significance and relevance of the final data collected can also be questioned, understanding the number of in/out octets of an interface and therefore it's utilization, only provides basic information on the activity of the interface. It fails to provide the necessary information with regards to the affects this activity had on the network. Did the activity result in congestion and consequential an increase in latency? Was the activity the result of a temporary period of oversubscription and consequently packets being dropped?

To truly understand the traffic patterns of today's Data Center and their affect on network congestion and the consequential effect this may have on application throughput and latency a far more granular and pro-active model is required.

Arista LANZ

Arista's Latency Analyzer (LANZ) is a pro-active event driven solution, purposely designed to provide real-time visibility of congestion hot spots and their effect on application performance and latency at a nanosecond resolution. The level of

Detecting congestion events at the queue level rather than the egress ports as a whole, LANZ+ provides visibility of congestion not simply at the network level but at a more granular traffic class or application level.

granularity LANZ delivers is made possible by it's unique event driven architecture, rather than the traditional polling model that only provides visibility at discrete interval. LANZ reports on congestion events as they occur.

The first and earliest indicator of a network congestion event, is the depth or occupancy level of a switch's queue. During periods of congestion, where the offered load is beyond the available capacity of an egress port, a switch will attempt to manage the congestion by temporary buffering/queuing the frames that cannot be transmitted immediately. While buffering can prevent packet drops being observed by an application, the frames enqueued will introduce additional latency while extensive congestion will result in buffer exhaustion and eventual packet drops.

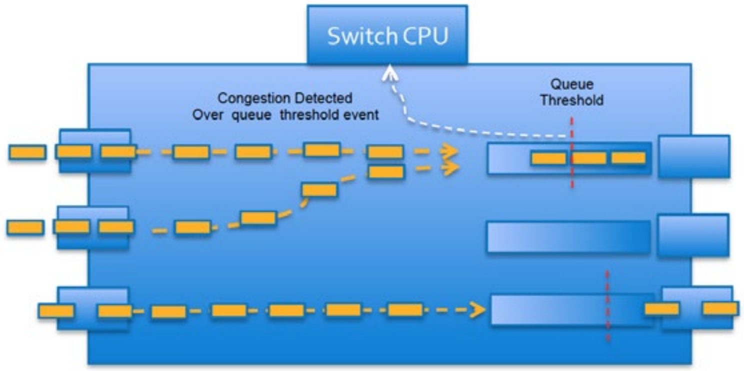

Figure 2: LANZ+ Event driven Congestion detection

The Arista LANZ functionality operates by monitoring and exporting in real time the queue length data on the switch based on a user configurable high and low threshold. At the point a queue threshold is crossed, indicating a period of congestion, LANZ+ generates a time stamped congestion event. This unique event driven model, allows congestion periods as short as a few 100ns to be detected and reported. Detecting congestion events at the queue level rather than the egress ports as a whole, LANZ+ provides visibility of congestion not simply at the network level but at a more granular traffic class or application level.

While providing granular visibility of application congestion events at the nanosecond level, LANZ delivers time stamped analytic data on the affects the event had on network latency and application throughput for accurate correlation with external monitoring tools.

LANZ Granularity

The key parameter that defines the precision at which congestion events can be observed and their affect on network latency is the granularity at which queue occupancy is measured. The level of granularity is determined by the segment size used to define the occupancy level of a queue. Packets in a queue are stored across one or multiple segments, the exact number being dependent on the size of the packets and the segment size itself. A large segment size in KiloBytes (e.g. 384KB) could mean the same level of congestion being detected regardless of whether one or a 1000 packets are stored in the segment. With a smaller byte sized segment

To provide fine grain visibility into congestion periods as short as a few 100ns, LANZ detects queue congestions events in small byte size segments (480 bytes) with the ability to detect when just two segment are enqueued in a buffer. This provides the ability to detect congestion when just a single packet is queued in the buffer.

a far greater level of granularity can be achieved, as this will allow detection of congestion when only one or two packets are queued.

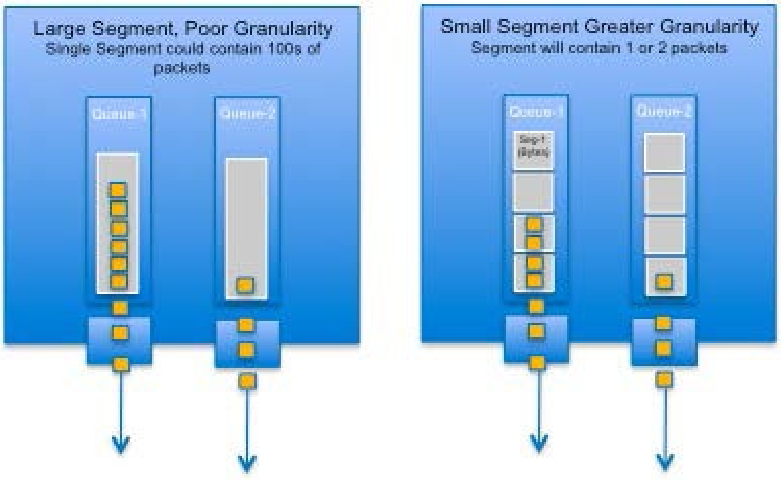

Figure 3: The smaller Segment size provides more granular view of the queue occupancy level

To provide fine grain visibility into congestion periods as short as a few 100ns, LANZ detects queue congestions events in small byte size segments (480 bytes) with the ability to detect when just two segment are enqueued in a buffer. This provides the ability to detect congestion when just a single packet is queued in the buffer.

Effects on Latency

As queued packet must wait for the buffer to clear before they can be transmitted, any buffering will introduce an extra level of latency into the application. The accuracy, with which this latency can be calculated, is also defined by the segment size used to determine queue occupancy.

The effect on application latency is determined by both the number of frames buffered and their individual size. For example a single 480 byte frame, due to serialisation delay will result in 400ns of additional latency on a 10Gbe interface or 4us on a 1Gbe interface. The latency increasing as more packets are buffered and with larger packet sizes.

```
480 byte segment + 20 byte overhead  = 500 bytes

Packet Overhead = Preamble (7 bytes)+ Start frame
Delimiter (1 byte) and Inter-Frame Gap (12 bytes)

10Gbe interface will transmits 1 byte of data in
0.8ns

Thus 500bytes will be take 0.8ns * 500 = 400ns to be
transmitted
```

As queue occupancy is detected by the number of segments held in the buffer rather than the number of packets, which could be of variable size, a smaller segment size will provide a more accurate estimate of the actual data bits stored in the queue and therefore it’s latency. With a larger segment size (e.g. KB) a single 64 byte packet could occupy a whole KB segment, resulting in a latency estimate in microsecond rather than the actual nanoseconds value.

LANZ provides real-time visibility into network congestion events as short as a few 100 nanoseconds.

LANZ Time Stamping

LANZ provides real-time visibility into network congestion events as short as a few 100 nanoseconds. To provide an audit trail each event generated and stored by LANZ is time-stamped with 10us resolution. This allows accurate historical trending of network congestion events and precise correlation to specific application traffic patterns at the microsecond level.

LANZ Implementation

Monitoring of congestion operates by allowing user configurable high and low thresholds to be set on each of the switch's queues. Providing threshold setting at the queue rather than the port level allows the monitoring of specific applications and traffic classes within the network. The high threshold defines the occupancy level before LANZ monitoring should start on the specific queue. The low threshold defining the level at which the occupancy of the queue needs to drop, before monitoring stops on the queue.

When a queue high threshold is first crossed a congestion "Start" record is generated, recording the time of the events occurrence, the queue length and the packet drop count at the start of the event. While the queue remains over the high threshold, periodic "Update" records for the queue are created, reporting, the new queue length and any effect the congestion has on the packet drop count. At the point the queue falls below the defined low threshold a congestion "End" record is created. The congestion "End" record providing a summary of the congestion event; calculating highest queue length over the congestion period, when it occurred and the duration of the overall congestion event.

```
Arista-7150#show queue-monitor length csv

Report generated at 2013-01-16 20:48:42
Type,       Time,                  Interface, Duration, Queue-Length,  Time-Of MaxQueue, Latency
                                              (usecs)                       (usecs)      (usecs)
 S,  2013-01-16 20:44:24.40391  Et46(1)    N/A        677               N/A          60.659
 U,  2013-01-16 20:44:24.40400  Et46(1)    N/A       1664               N/A         149.094
 U,  2013-01-16 20:44:24.40409  Et46(1)    N/A       2585               N/A         231.616
 U,  2013-01-16 20:44:24.40417  Et46(1)    N/A       3119               N/A         279.462
 U,  2013-01-16 20:44:24.40425  Et46(1)    N/A       3118               N/A         279.372
 U,  2013-01-16 20:44:24.40533  Et46(1)    N/A       3118               N/A         279.372
 U,  2013-01-16 20:44:24.40542  Et46(1)    N/A       3119               N/A         279.462
 U,  2013-01-16 20:44:24.40550  Et46(1)    N/A       3121               N/A         279.641

-----------------------Truncated Output--------------------------

 U,  2013-01-16 20:45:29.40736  Et46(1)    N/A       3119               N/A         286.948
 U.  2013-01-16 20:45:34.40746  Et46(1)    N/A       3118               N/A         286.856
 E,  2013-01-16 20:45:37.51457  Et46(1)  73110667    3121*             2257         287.132
```

Figure 4: LANZ report for a congestion event detected on Interface 46 traffic class 1

The effect each congestion records has on latency is automatically calculated from the number of segments enqueued during each event and is viewable from the CLI.

```
Arista-7150# show queue-monitor length tx-latency

Report generated at 2013-01-16 20:47:57
Time                           Intf( TC )     Tx-Latency (usecs)
-----------------------------------------------------------------
0:02:20.08527 ago              Et46(1)        287.132
0:02:23.19239 ago              Et46(1)        286.856
0:02:28.19249 ago              Et46(1)        286.948
0:02:33.19259 ago              Et46(1)        286.948
0:02:38.19270 ago              Et46(1)        286.948
0:02:43.19280 ago              Et46(1)        286.948
------------- Truncated output ------------------------
0:03:33.19451 ago              Et46(1)        279.372
0:03:33.19542 ago              Et46(1)        279.462
0:03:33.19550 ago              Et46(1)        279.552
0:03:33.19559 ago              Et46(1)        279.372
0:03:33.19567 ago              Et46(1)        279.462
0:03:33.19576 ago              Et46(1)        231.616
0:03:33.19584 ago              Et46(1)        149.094
0:03:33.19594 ago              Et46(1)        60.659
```

Figure 5: LANZ output calculating the network latency effects for each congestion event detected on Interface 46 traffic class 1

LANZ Data Collection

LANZ provides congestion data events in real-time, making it possible to monitor and alert on congestion at microsecond resolution rather than seconds. This level of resolution enables accurate historical trend analysis of congestion events. Operating in real-time makes pro-active capacity management a reality. Rather than wait for applications to report a drop in performance, capacity management tools can now have the relevant data present to them in real-time, enabling them to react pre-emptively on transient hotspots before congestion accumulates to affect the application layer.

To make this possible and provide maximum versatility with regard to how the data maybe consumed, the LANZ data is presented in a number of open standard formats for both real-time and historical usage.

- **CLI Output:** All congestion data is available instantaneous and continuously via the Command line Interface (CLI) of the switch for quick and easy analysis of congestion data by network administrator
- **Syslog Messaging:** For automatic alerts on the occurrence of a congestion event, LANZ will generate a Syslog message when queue threshold is exceeded.
- **CSV format:** For storing LANZ data for historical trending and third party analysis, the congestion data saved in a CSV which can be stored on the flash, USB, SSD or an external file system (FTP, TFTP, NFS)
- **Congestion Data Stream*:** To provide real-time visibility of congestion events for pro-active monitoring and potential capacity management, congestion events can be streamed in real-time to external third-party monitoring tools. The streamed data utilizes the industry standard format of Google Protocol Buffers (GPB).

Santa Clara—Corporate Headquarters
5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500
Fax: +1-408-538-8920
Email: info@arista.com

Ireland—International Headquarters
3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office
9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

San Francisco—R&D and Sales Office
1390 Market Street, Suite 800
San Francisco, CA 94102

India—R&D Office
Global Tech Park, Tower A & B, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Nashua—R&D Office
10 Tara Boulevard
Nashua, NH 03062