Commissioned by
Lightfleet Corporation

Lightfleet Data Distribution System (DDS)

Competitive Performance Evaluation

Lightfleet DDS vs. Arista 7150S-52 10GbE Switch with Solarflare SFN7122F 10GbE Adapters

EXECUTIVE SUMMARY

High-Performance Computing (HPC) and Big Data applications demand the best performance possible with the lowest latency and the greatest throughput. Traditional LAN switch architectures add control plane and other overhead that negatively impact performance.

Lightfleet Corporation commissioned Tolly to evaluate the end-to-end latency, throughput and fabric power consumption of its Lightfleet Data Distribution System (DDS). DDS consists of the Lightfleet Data Distribution Module (DDM) which provides the fabric and the Lightfleet host bus adapter (HBA) and related software for the end-station. This was compared to a traditional switch as represented by the Arista 7150 running 10GbE in conjunction with Solarflare 7122 10GbE network interface cards (NIC) and related software for the end-station.

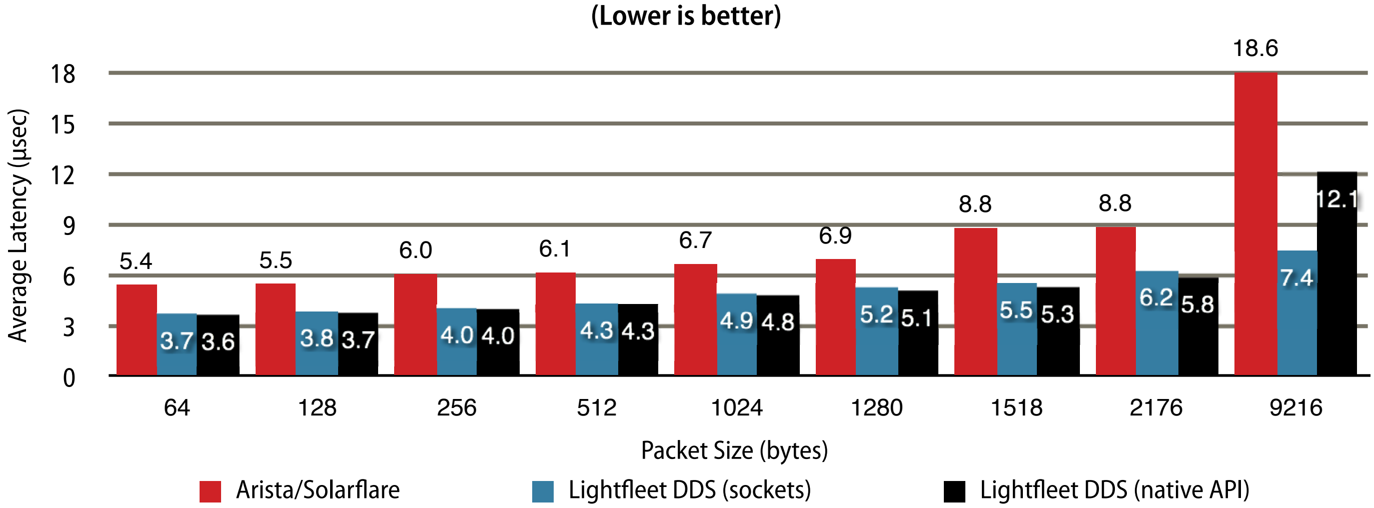
Tests showed that the Lightfleet solution delivered higher throughput, lower latency and lower power consumption than the Arista/Solarflare solution.

THE BOTTOM LINE

Lightfleet's Data Distribution System delivers:

1. Lower end-to-end latency at all packet sizes tested including jumbo packets
2. Higher throughput at all packet sizes tested including jumbo packets
3. Significantly lower power consumption under all load conditions (idle, 10%, 100%)

End-to-End Average Latency Comparison (Two Ports)
Lightfleet DDS vs Arista 7150 10GbE Switch & Solarflare SFN7122F 10GbE Adapters
(Lower is better)



Notes:1) Arista 7150S-52 switch was run in cut-through mode, 2) Arista 7150/Solarflare 7122 NIC latency measured with sockperf network benchmark tool (ping-pong) using UDP protocol 3) Lightfleet DDS latency with sockets measured with sockperf and libsocklf, 4) Lightfleet DDS latency with native API measured using ping-pong. Packet sizes/methodology per RFC 2544 5) Latencies of optical transceivers and fiber cables are included in latency results.

Source: Tolly, August 2016

Figure 1

Importance of End-to-End Performance

For many demanding applications, such as HPC and Big Data, it is not just the data fabric performance characteristics that matter but also the performance "up" through the network adapter and even the software stack to that point where the network meets the application.

For this reason, these tests measure not only the data fabric device (e.g. switch), but also the network adapter and the communications protocol stack. Results represented what was measured by the benchmarking applications on top of the relevant protocol stack. Lightfleet provides a data fabric that is a purpose-built alternative to traditional Ethernet-based LAN switches. End-stations use a Lightfleet host bus adapter (HBA) in place of an Ethernet adapter. Applications can be run either over a standard TCP sockets interface or using the high-performance Lightfleet API. Tests were run using both software stack options in a Red Hat Enterprise Linux environment.

Test Results

Latency

Lightfleet delivered lower (better) latency than the Arista/Solarflare solution at every packet size, from smallest to jumbo, using a standard socket interface and the Lightfleet API in tests of two ports.

Lightfleet Corp.

Lightfleet Data Distribution System (DDS)

Competitive Performance Evaluation

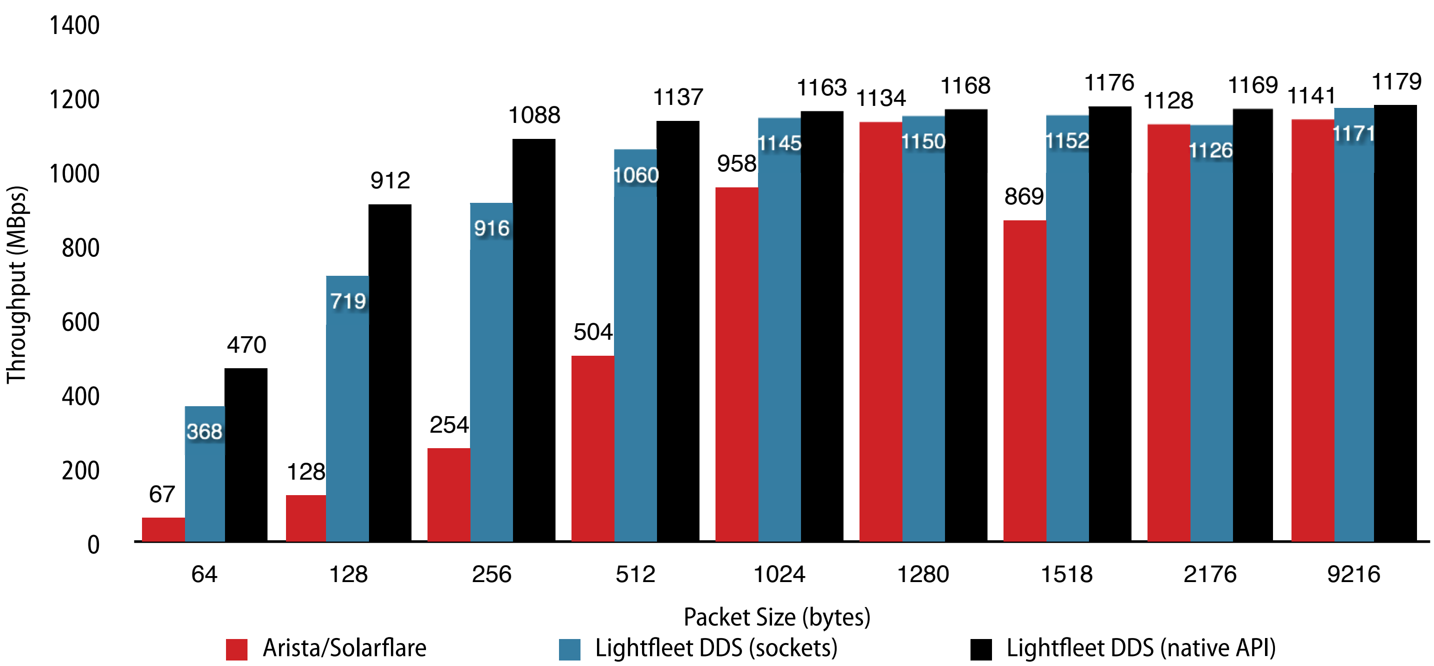
Tolly Certified

Tested August 2016

At the smallest, 64-byte packet size, Lightfleet's latency of 3.7µsec was 31% better than the Arista/Solarflare 5.4µsec. At 1518-bytes, the largest standard Ethernet frame size, Lightfleet's latency of 5.5µsec was, 37% better than the Arista/Solarflare 8.8µsec. At the largest, 9216-byte packet

Throughput Comparison (Two Ports)

Lightfleet DDS vs Arista 7150 10GbE Switch & Solarflare SFN7122F 10GbE Adapters



Notes: Notes: Arista 7150/Solarflare 7122 NIC throughput measured with sockperf network benchmark tool (throughput) using UDP protocol 3) Lightfleet DDS throughput with sockets measured with sockperf and libsocklf, 4) Lightfleet DDS throughput with native API measured using ping-pong in throughput mode. Packet sizes/methodology per RFC 2544 5) Latencies of optical transceivers and fiber cables are included in latency results.

Source: Tolly, August 2016

Figure 2

size, Lightfleet's latency of 7.4µsec was 60% better than the Arista/Solarflare's 18.6µsec.

For all packet sizes through 2176-bytes, the latency for the Lightfleet native API was lower than or equivalent to the latency for the standard TCP stack with Lightfleet. See Figure 1 for all latency results.

Throughput

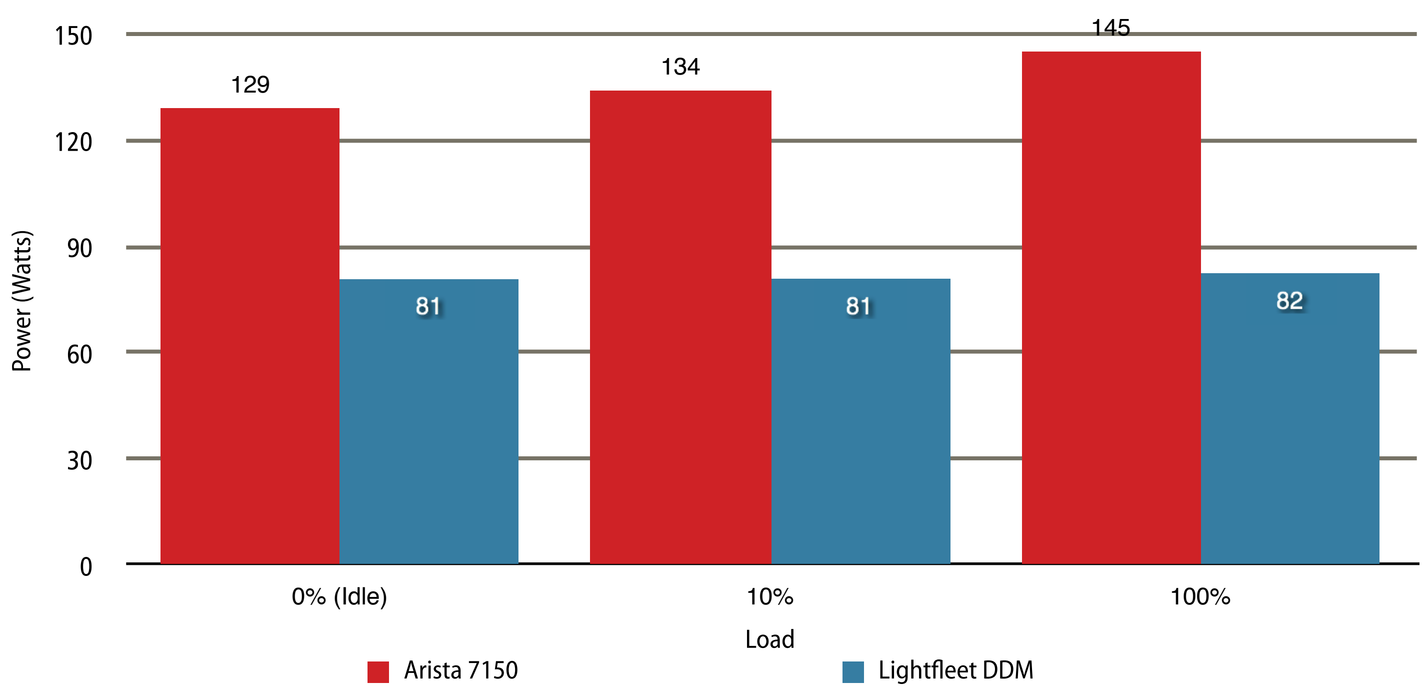
Lightfleet delivered higher throughput than the Arista/Solarflare solution at all packet sizes tested in tests of two ports. This was the case with the sockets test and with the Lightfleet native API. The Lightfleet throughput benefits were particularly noticeable in packet sizes up through 512-bytes. For example, with 64-byte packets, Lightfleet's TCP results of 368MBps was 5.5X that of Arista/Solarflare's 67MBps. Using the Lightfleet native API, the throughput was 470MBps or 7X that of Arista/Solarflare. At 512-byte packets, the Lightfleet TCP and native API results remained over 2X that of Arista/Solarflare. See Figure 2 for all throughput results.

Power Consumption

Lightfleet's DDM consumed less power than the Arista Networks 7150S in all three load scenarios: 1) Idle (0% load), 10% load, and 100% load. Tests were run using 24 ports on each system.

With the system idle, Lightfleet's power consumption was 37% better (lower) than Arista Networks with the Lightfleet DDM consuming 81W compared to 129W for Arista Networks. With a 10% load, the Lightfleet power consumption remained at 81W while the Arista switch consumed 134W. At 100% load, the Lightfleet power draw remained essentially the same at 82W. This was 43% better than the Arista switch which required 145W at this load.

Power Consumption Comparison (24 Ports)
Lightfleet DDM 1.2 vs Arista 7150S-52 10GbE Switch
(Lower is better)

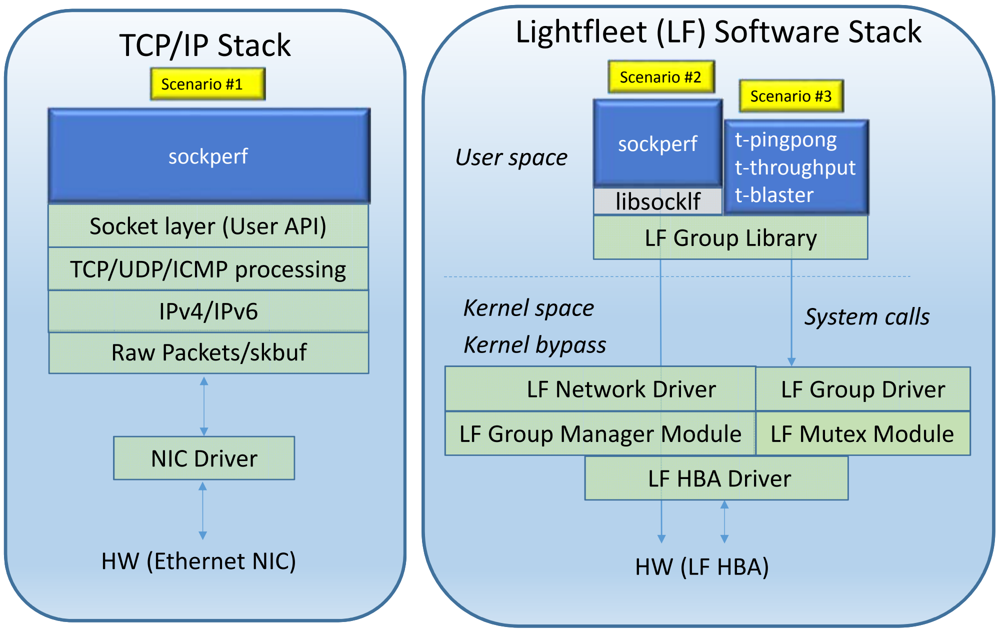


Note: Arista tested with 24 of its possible 52 ports installed. SPF+ transceivers were used for ports under test.

Source: Tolly, August 2016

Figure 3

Test Stack Configurations



Source: Lightfleet, August 2016

Figure 4

Solutions Under Test

	Switch/Data Fabric	Networking Adapter
Arista Networks	Arista DCS-7150S-52-CL-F (Hardware: 2.01, Software: 4.17.0F) Tested using 24 10GbE ports (SFP-10G-SR)	Solarflare SFN7122F 10GbE Adapter (Drivers: 4.8.2.1004)
Lightfleet	Lightfleet Data Distribution Module (Firmware v2016/8/12)	Lightfleet HBA (Drivers lf_hbadrv.ko 0.0.1-test)

Source: Tolly, August 2016

Table 1

Understanding Lightfleet's Switch-Free Approach

The Lightfleet throughput advantage is of key importance because the performance of HPC and Big Data systems is ultimately limited by the throughput of the networks that connect the processors that comprise them. With switch-based networks, one can't just add more processors and expect overall performance to increase commensurately. This is because switching delays compound latencies and increase the effects of congestion, including dropped packets, as more processors attempt to communicate. Since the Lightfleet DDS architecture is data-directed, the effects of congestion are less significant and clear much more quickly than with switched architectures.

The potential throughput advantage of the Lightfleet fabric is actually much greater than that exhibited by the current Tolly test data which was produced using unicast transfers. This is because, with the DDS, multicast is the native transfer mode. One data source can send packets to an unlimited number of destinations simultaneously. Switched architectures must generally implement multicast transfers as a sequence of unicast ones.

A further advantage for the Lightfleet architecture comes from the fact that the DDS fabric is continuously adaptive according to the needs of the data. The data flow control function (the protocol stack), which other network architectures implement in software, is now built into the hardware of the fabric connections, and different data-flow types are handled automatically. This means that the DDS architecture can adapt much more easily to meet the needs of future HPC and Big Data applications.

The lower power consumption of the Lightfleet configuration is also significant, as it will contribute to customers saving energy and cooling system costs versus other networking systems.

Source: Lightfleet

Test Setup & Methodology

Test Environment

Testing was "end-to-end" and thus benchmarked not only the switch/data fabric but also the end-station network adapter and drivers.

All end stations used in the testing were Dell PowerEdge R220 Rack Servers outfitted with a single network adapter. The end-stations ran CentOS/Red Hat Enterprise Linux Version 7.2.

The Arista Networks 10GbE switch was paired with 10GbE network adapters from Solarflare. Solarflare was chosen for the network adapter as the company is recognized as producing high-performance, low-latency adapters that are in broad use throughout the industry.

Lightfleet's Data Distribution System (DDS) was tested. This consisted of several components.

Lightfleet implements its switch-replacement data fabric in its Lightfleet Data Distribution Module - referred to as the Lightfleet DDM.

The Lightfleet HBA provides the network adapter function for the end station along with various standards-based and proprietary protocol stack modules. See Table 1 for details and version levels for the systems under test.

The Arista switch was outfitted with 24 10GbE ports (out of a possible 52) using SFP+ interfaces. The switch had two power supplies. The switch was configured to run cut-through switching to reduce packet latency.

The Lightfleet DDM was outfitted with 24 ports and two power supplies.

Test Methodology

Tests were focused on measuring the end-to-end latency and throughput of the systems under test and, additionally, on power consumption. For Lightfleet, latency and throughput tests were run using both standard TCP drivers in the end stations and again using Lightfleet's proprietary application user interface (API). Additionally, tests were run to illustrate the power consumption of the two networking

Lightfleet's Data Distribution Module (DDM)

switch/data fabric devices. For details of the network stacks tested, please see Figure 4.

Benchmarking Programs

TCP system latency and performance was measured using Sockperf[1]. The version of sockperf that was used was modified by Mellanox to add the VMA interface and is publicly available via GitHub. Sockperf "is a network benchmarking utility over socket API that was designed for testing performance (latency and throughput) of high-performance systems..." To run this utility on the Lightfleet HBA, Lightfleet has developed a library identified as libsocklf. Libsocklf is a library designed to the VMA library interface on sockperf version 2.7-54.git4e9e71bf405b.

For testing the Lightfleet API, engineers used Lightfleet-developed utilities that provide similar latency and throughput measurement capabilities as sockperf. These programs are known as t-pingpong and t-throughput.

For the power consumption tests, engineers used the iperf3 utility to drive traffic. A Kill A Watt meter was used to measure the power consumption which was transcribed by engineers.

Latency & Throughput

All latency and throughput tests were run using two ports. The Sockperf utility was used for benchmarking the TCP stack on both systems. SockPerf was configured to run all the RFC 2544 packet sizes along with several jumbo sizes (2176- and 9216-bytes).

Sockperf "ping-pong" mode was used for latency tests and Sockperf "throughput" mode was used for the throughput tests.

For tests using the Lightfleet API, Lightfleet's t-pingpong utility was used. For latency tests it was used in its default "ping-pong" mode. For throughput tests it was run using its "throughput" parameter.

Tests were run three times for 60 seconds each run and the average result was used.

Power Consumption

The iPerf3 utility was used to generate traffic to measure the power consumption of the switch/data fabric under test.

Traffic was bidirectional and ran station-to-station using 24 stations. iPerf3 command line run at each station was: a) iperf3 -s -p 7002 -D, and b) iperf3 -t -p 4 -u -p 7002 -b 1.25G -i 60 - c 192-168.n.n. Note that the bandwidth value varied depending upon the test. Tests were run continually until measurement was taken.

The power consumption was measured first with no traffic running. Then, iPerf3 was run to generate loads of 10% and then 100%. Engineers measured the power consumption at each of these levels.

Relative Performance Calculation

To calculate how much better one solution is than another, the formula used is 1- (T1/T2) where T1 is the better result and T2 is the slower (worse) result. This is multiplied by 100 to give the percentage benefit. For "X" (times) better calculations, the better result is simply divided by the worse result.

[1] https://github.com/mellanox/sockperf/wiki

About Tolly

The Tolly Group companies have been delivering world-class IT services for more than 25 years. Tolly is a leading global provider of third-party validation services for vendors of IT products, components and services.

You can reach the company by E-mail at sales@tolly.com, or by telephone at +1 561.391.5610.

Visit Tolly on the Internet at: http://www.tolly.com

Interaction with Competitors

In accordance with Tolly's Fair Testing Charter, Tolly personnel invited representatives from Arista Networks to participate in the testing. Arista Networks did not respond to this invitation.

For more information on the Tolly Fair Testing Charter, visit:

http://www.tolly.com/FTC.aspx

Terms of Usage

216144 mjo-1 wt-2016-09-12-VerH