



## **Ultra-Low Latency, High Density 48 port Switch and Adapter Testing**

**Testing conducted by Solarflare and Force10 shows that ultra low latency application level performance can be achieved with commercially available 10G Ethernet Switch and Server Adapter Products**

**Bruce Tolley, Solarflare**

### **Abstract**

Solarflare, the pioneer in high-performance, low-latency 10GbE server networking and application acceleration middleware solutions and Force10 Networks, the leader in high-performance open data center networking recently completed switch to server latency and message rate testing. The test found that the Solarflare SFN5122F server adapter and Solarflare OpenOnload® application acceleration middleware in combination with the Force10 10GbE Switch achieved 5.0 microsecond mean latency over 10GbE in a TCP latency test with 64 byte packet size typical of market data applications. The Force10 switch is one of the lowest latency switches on the market, demonstrating mean latency in this test of 920 nanoseconds. With back to back servers, the server adapter achieved impressive minimum latencies as low as 4.8 microseconds. The latency of the overall system was also very deterministic with 99% of the messages being delivered with a latency of less than 5.5 microseconds. Solarflare and Force10 measured the performance of TCP and UDP messaging using Solarflare developed benchmarks with commercially available products: the Solarflare® SFN5122F 10 Gigabit server adapters and Force10 S4810 10 Gigabit switch. The test platform used servers and processors typically found in use by financial firms today.

### **The Need for Low Latency in Automated, Real-Time Trading**

The rapid expansion of automated and algorithmic trading has increased the critical role of network and server technology in market trading, first in the requirement for low latency and second in the need for high throughput in order to process the high volume of transactions. Given the critical demand for information technology, private and public companies that are active in electronic markets continue to invest in their LAN and WAN networks and in the server infrastructure that carries market data and trading information.

In some trading markets, firms can profit from less than one millisecond of advantage over competitors, which drives them to search for sub-millisecond optimizations in their trading systems. The spread of automated trading across geographies and asset classes, and the resulting imperative to exploit arbitrage opportunities based on latency, has increased the focus on if not created an obsession with latency.



With this combination of forces, technologists, IT and data center managers in the financial services sector are constantly evaluating new technologies that can optimize performance. One layer of the technology stack that receives continuous scrutiny is messaging, i.e., the transmission of information from one process to another, over networks with specialized home-grown or commercial messaging middleware.

The ability to handle predictably the rapid growth of data traffic in the capital markets continues to be a major concern. As markets become more volatile, large volumes of traffic can overwhelm systems, increase latency unpredictably, and throw off application algorithms. Within limits, some algorithmic trading applications are more sensitive to the predictability of latency than they are to the mean latency. Therefore it is very important for the network solution stack to perform not just with low latency but with bounded, predictable latency. Solarflare and Force10 demonstrate in this paper that because of its low and predictable latency, a UDP multicast network built with 10 Gigabit Ethernet (10GigE) can become the foundation of messaging systems used in the financial markets.

### **Financial services applications and other applications that can take advantage of low-latency UDP multicast**

Messaging middleware applications were named above as one key financial services application that produce and consume large amounts of multicast data which can take advantage of low-latency UDP multicast. Other applications in the financial services industry that can take advantage of low-latency UDP multicast data include:

- Market data feed handler software that takes as input multicast data feeds and uses multicasting as the distribution mechanism
- Caching/data distribution applications that use multicast for cache creation or to maintain data state
- Any application that makes use of multicast and requires high packets per second (pps) rates, low data distribution latency, low CPU utilization, and increased application scalability

### **Cloud Networking and the Broader Market Implications of Low Latency to Support Real-time Applications**

As stated above, the low-latency UDP multicast solution provided by Force10 switches and Solarflare server adapters can provide compelling benefit to any application that depends on multicast traffic where additional requirements exist for high throughput, low-latency data distribution, low CPU utilization, and increased application scalability. Typical applications that benefit from lower latency include medical imaging, radar and other data acquisition systems, and seismic image processing in oil and gas exploration. Yet moving forward, cloud networking is a market segment where requirements for throughput, low latency and real time application performance will also develop. The increasing deployment and build out of both public and private clouds will drive the increased adoption of social networking and Web 2.0 applications. These cloud applications will incorporate real-time media and video distribution and will need lower latency applications for both business to consumer (B2C) and business to business (B2B) needs. Perhaps more fundamentally, the need for real-time, high-speed analytics and search of large and often unstructured data sets will create demand for low latency and real time application response.

Solarflare and Force10 measured the latency performance of messaging using Solarflare-developed benchmarks with commercially available products: the Solarflare SFN5122F SFP+

10 Gigabit server adapters, Solarflare OpenOnload application acceleration middleware, and the Force10 10 Gigabit switch. A list of the hardware configurations and the benchmarks used is attached as an Appendix. The test platform used servers and processors typically found in use by financial firms today. The tests described below were run both switch to server adapter and server adapter to server adapter. The adapters were run in kernel mode and in OpenOnload mode.

OpenOnload is an open-source high-performance application acceleration middleware product. By improving the CPU efficiency of the servers, OpenOnload enables applications to leverage more server resources, resulting in dramatically accelerated application performance without changing the existing IT infrastructure. Using standard Ethernet, the solution combines state-of-the-art Ethernet switching and server technologies that dramatically accelerate applications. OpenOnload performs network processing at user-level and is binary-compatible with existing applications that use TCP/UDP with BSD sockets. It comprises a user-level shared library that implements the protocol stack, and a supporting kernel module.

## Fundamental Findings

Exhibit 1: Half-Round Trip Latency in Nano Seconds

	<i>size</i>	<i>mean</i>	<i>min</i>	<i>median</i>	<i>max</i>	<i>99%ile</i>	<i>stddev</i>
<b>tcp latency ool switch</b>	<b>64</b>	5042	4896	5006	11577	5546	107
<b>tcp latency b2b ool</b>	<b>64</b>	4122	3963	4092	10957	4616	105

Exhibit 1 summarizes results of TCP latency testing. The Force10 switch is a very low-latency switch contributing a mean latency of 920 nanoseconds to the system latency. In the testing for the 64 byte message sizes typical of market data messaging systems, very low latency was observed. The server adapter in combination with the Force10 switch achieved mean latency of 5.0 microseconds. The Solarflare adapters back to back achieved an amazingly low latency of 4.1 microseconds. This latency was also very deterministic with 99% of the messages achieving a mean latency of less than 5.5 microseconds in the switch to server adapter configuration.

Exhibit 2: TCP latency performance vs. message size

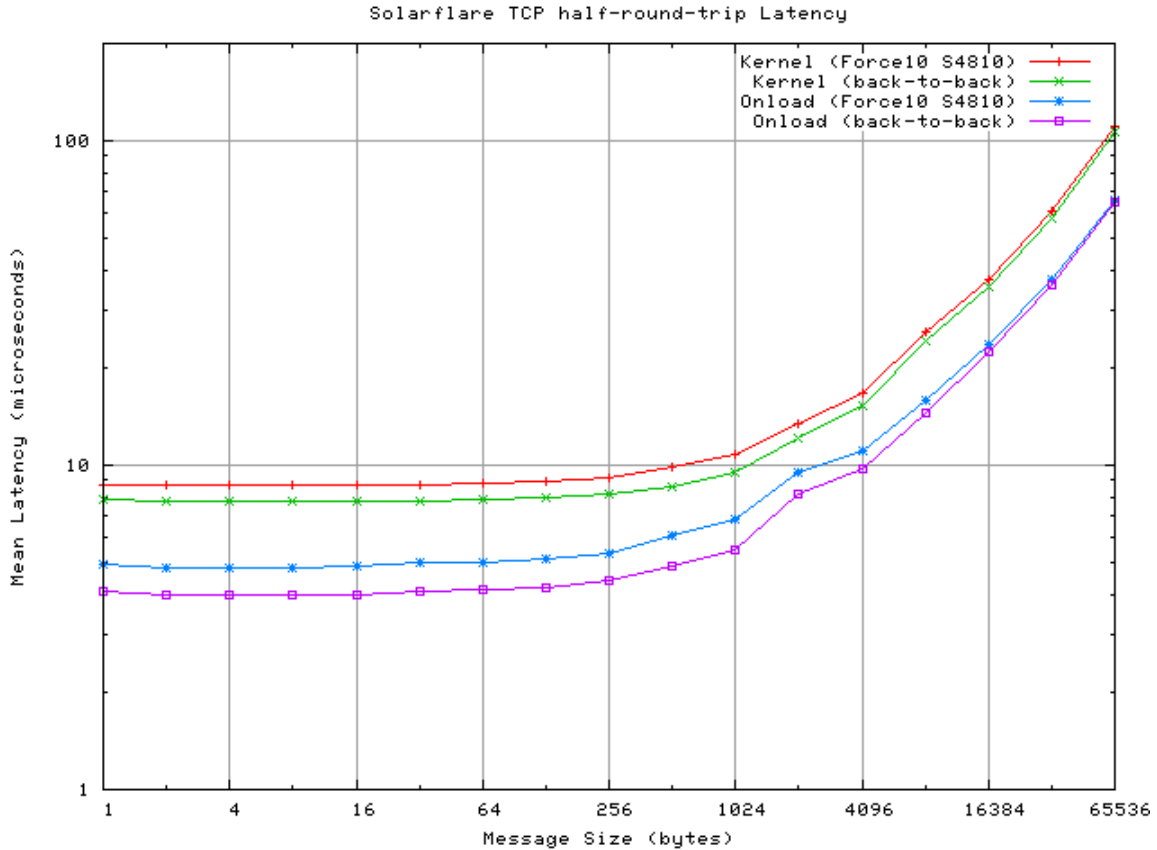


Exhibit 2 above represents the same data and Exhibit 1 but plots TDP half round trip latency where the x axis represents message size in bytes and the y axis represents latency in microseconds. The data shows that the system demonstrated very low and deterministic latency from small up to very large message sizes up to 1500 bytes. The data plot also shows very low latency in both kernel and OpenOnload mode. The Solarflare adapters back to back, achieved an amazingly minimum latency of 3.9 microseconds. This latency was also very deterministic as demonstrated by the flatness of the curve as the message size approaches 1500 bytes.

Exhibit 3 below plots UDP half round trip latency where the x axis represents message size in bytes and the y axis represents latency in microseconds. The data shows that the system demonstrated very low and deterministic latency from small up to very large message sizes of 1024 bytes. The data plot also shows very low latency in both kernel and OpenOnload mode. In OpenOnload mode with the switch and server adapter, minimum latencies go as low as 4.8 microseconds for 64 byte packet messages, and with the server adapters back to back, as low as 3.9 microseconds.

Exhibit 3: UDP Half Round Trip Latency

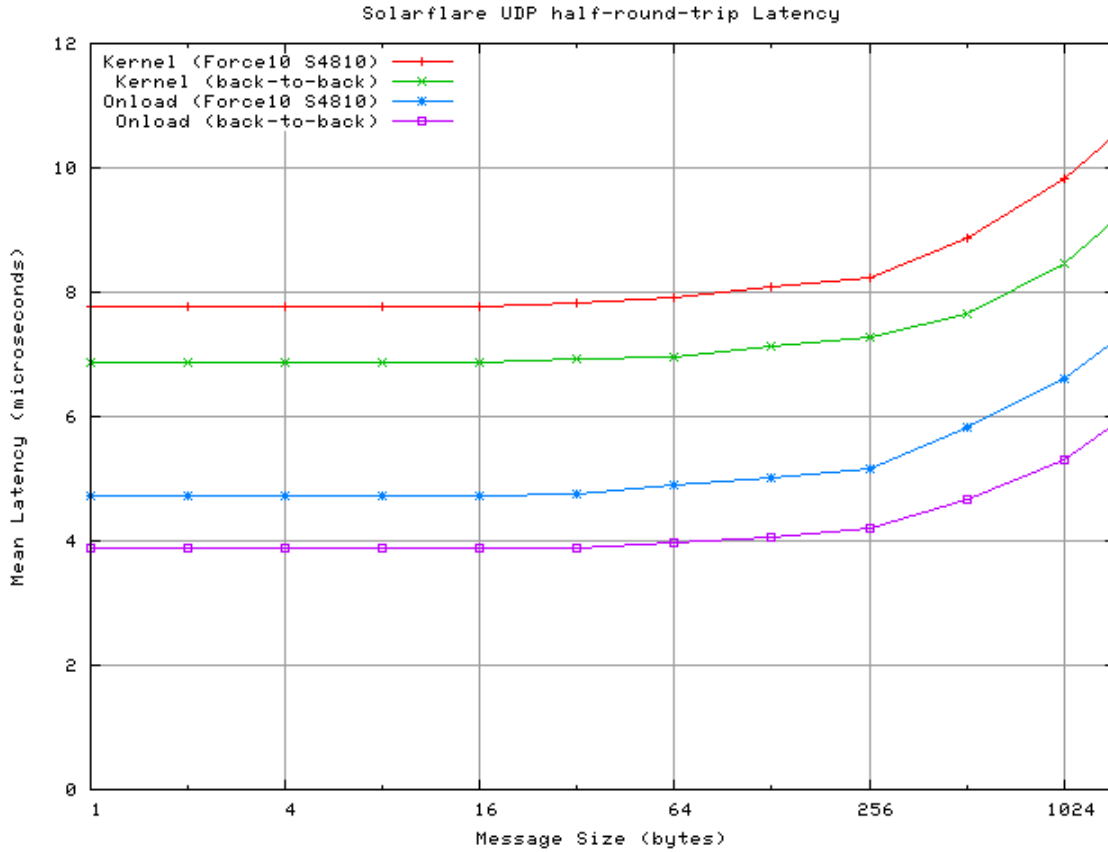
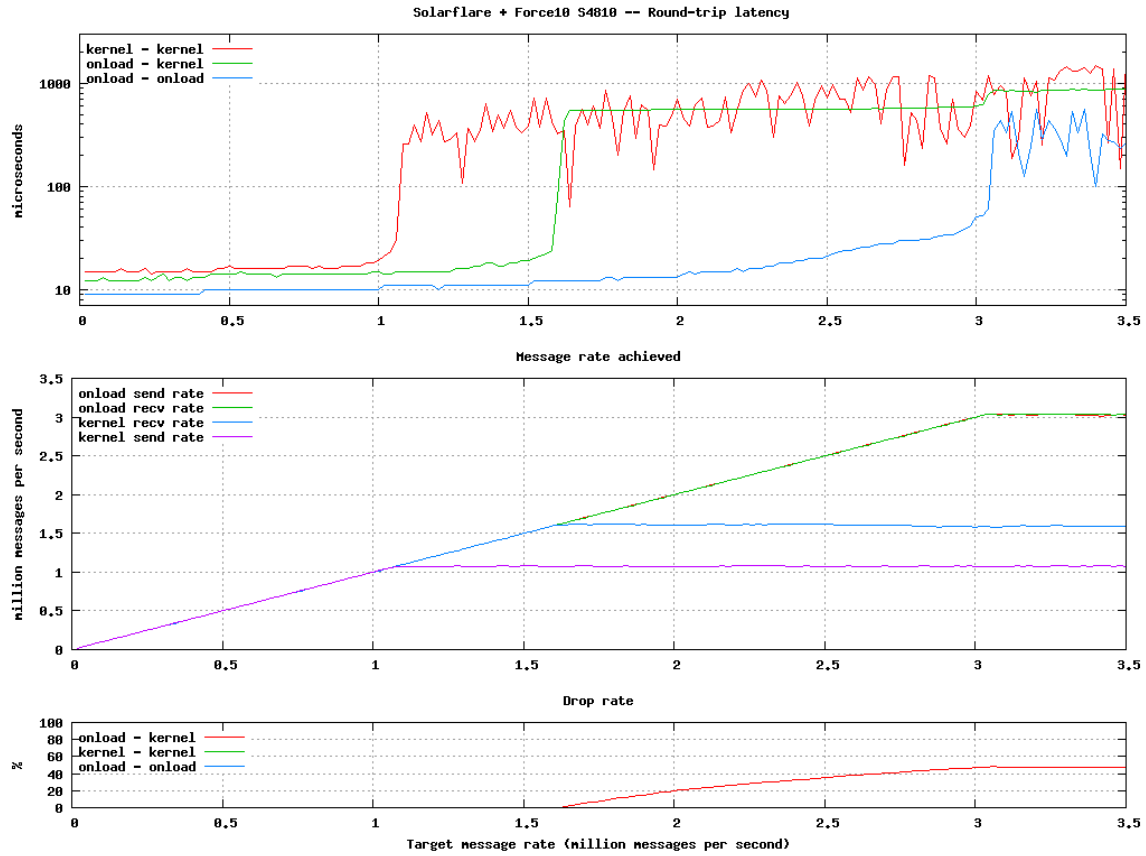


Exhibit 4 below shows two plots of performance versus desired data rate of UDP multicast performance with and without OpenOnload. This test simulates a traffic pattern that is common in financial services applications. In the test, the system streams small messages from a sender to a receiver. The receiver reflects a small proportion of the messages back to the sender which the sender uses to calculate the round-trip latency. The x axis shows the target message rate that the sender is trying to achieve. The y axis shows one-way latency (including a switch) and the achieved message rate. The kernel results are measured with Solarflare server adapters without OpenOnload. The plot combines results from three runs: kernel to kernel, OpenOnload to kernel, and OpenOnload to OpenOnload. The OpenOnload to kernel test is needed in order to fully stress the kernel receive performance.

Exhibit 4: Message Rates Achieved with Upstream UDP



The top plot labeled Round Trip Latency shows the improved, deterministic low latency achieved with the Solarflare adapter, OpenOnload, and the Force10 switch. The y axis shows the round trip latency while the x axis shows the desired message rate in millions of messages per second at the receiver. With OpenOnload, not only is the system performing at much lower latency, but the latency is predictable and deterministic over the range of expected message rates. This is precisely the attribute desired in trading systems or any other application demanding real time performance.

The second plot in Exhibit 4, Message Rate Achieved shows the Solarflare OpenOnload system's ability to scale and perform as the message rate is increased. This is in contrast to the kernel stack where the greater CPU processing overheads of the stack limit performance as higher levels of load are put on the system.

## The Solarflare Solution

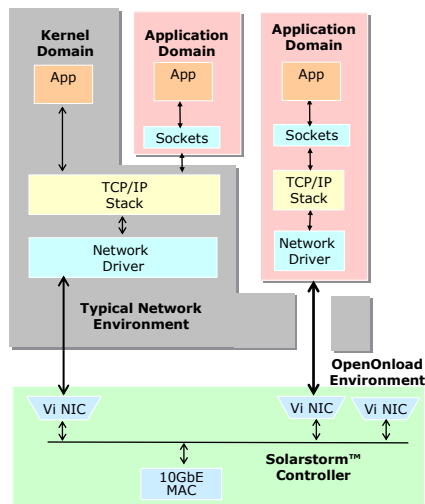
The SFN5122F 10GbE SFP+ server adapter is the most frequently recommended 10Gb Ethernet server adapter for trading networks in New York, London and Chicago. In both kernel and OpenOnload modes, the adapter supports financial services and HPC applications which demand very low latency and very high throughput. Tests were performed using the standard kernel TCP/IP stack as well as Solarflare OpenOnload mode.

OpenOnload is an open-source application acceleration middleware product from Solarflare. As Exhibit 5 shows, the OpenOnload software provides an optimized TCP/UDP stack into the application domain which can communicate directly with the Solarflare server adapter. With OpenOnload, the adapter provides the application with protected, direct access to the network, bypassing the OS kernel, and hence reducing networking overheads and latency.

The typical TCP/UDP/IP stack resides as part of the kernel environment and suffers performance penalties due to context switching between the kernel and application layers, the copying of data between kernel and application buffers, and high levels of interrupt handling.

Exhibit 5: The Solarflare Architecture for OpenOnload

### Solarflare Architecture for OpenOnload



- Binary compatible with industry standard APIs
- Leverages existing network infrastructure
- Requires no new protocols
- Single ended acceleration
- Scales easily to support Multi-core CPU Servers.
- Self balances to optimize cache locality

The kernel TCP/UDP/IP and OpenOnload stacks can co-exist in the same environment. This co-existence allows applications that require a kernel based stack to run simultaneously with OpenOnload. This coexistence feature was leveraged as part of the testing where the benchmarks were run through both the kernel and OpenOnload stacks in back to back fashion using the same build and without having to reboot the systems.



## **Force10 S4810 ultra-low-latency switch: Industry Leading Layer 3 Performance, Scalability, and High Availability**

The Force10 S4810 is recognized through independent industry testing to be the best-in-class 48 port switch with the lowest packet latencies available today. The Force10 S series Datacenter Ethernet switches feature the industry's highest density, lowest power, lowest latency and 40GbE/10GbE/GbE interface connectivity.

The Force10 S-Series S4810 is an ultra low-latency 10/40 GbE Top-of-Rack (ToR) switch purpose-built for applications in high-performance data center and computing environments. Leveraging a non-blocking, cut-through switching architecture, the S4810 delivers line-rate L2 and L3 forwarding capacity with ultra low latency to maximize network performance. The compact S4810 design provides industry leading density of 48 dual-speed 1/10 GbE (SFP+) ports as well as four 40 GbE QSFP+ uplinks to conserve valuable rack space and simplify the migration to 40 GbE in the data center core.

(Each 40 GbE QSFP+ uplink can support four 10 GbE ports with a breakout cable). Powerful QoS features coupled with Data Center Bridging (DCB) support via a future software enhancement, make the S4810 ideally suited for iSCSI storage environments. In addition, the S4810 incorporates multiple architectural features that optimize data center network flexibility, efficiency, and availability, including Force10's VirtualScale stacking technology, reversible front-to-back or back-to-front airflow for hot/cold aisle environments, and redundant, hot-swappable power supplies and fans.

The S4810 also supports Force10's Open Automation Framework, which provides advanced network automation and virtualization capabilities for virtual data center environments. The Open Automation Framework is comprised of a suite of inter-related network management tools that can be used together or independently to provide a network that is more flexible, available and manageable while reducing operational expenses.

### **Conclusions**

- The findings analyzed in this white paper represent the results of testing of transmit latency of a configuration with the Solarflare server adapter with OpenOnload and the Force10 switch at transmission rates up to 3 million messages/second (mps). For the 64 byte message sizes typical of market data messaging systems, very low latency was observed:
- TCP latency Mean did not exceed 5.0 microseconds with switch
- TCP latency Mean did not exceed 4.1 microseconds without switch
- For TCP messaging traffic, 99th percentile did not exceed 5.5 microseconds with switch
- For UDP latency server to server, mean latency was as low at 3.8 microseconds

The system also demonstrated very low jitter for both TCP and UDP traffic which delivers a very predictable messaging system. With Solarflare's 10GbE server adapter and OpenOnload application acceleration middleware, and Force10's 10GbE switch, off-the-shelf 10GbE hardware can be used as the foundation of messaging systems for electronic trading with no need to re-write applications or use proprietary, specialized hardware.

Enabling financial trading customers to implement highly predictable systems, Solarflare's and Force10's 10GbE solutions provide a competitive advantage and offer increased overall speeds, more accurate trading and higher profits. Now, financial firms can use off-the-shelf Ethernet, TCP/IP, UDP and multicast solutions to accelerate market data systems without requiring the implementation of new wire protocols or changing applications. By leveraging the server adapter



with OpenOnload, IT managers are able to build market data delivery systems designed to handle increasing message rates, while reducing message latency and jitter between servers.

## Summary

Solarflare Communications and Force10 have demonstrated performance levels with 10 Gigabit Ethernet that enable Ethernet to serve as the foundation of messaging systems used in the financial markets. Now, financial firms can use off-the-shelf Ethernet, TCP/IP, UDP and multicast

solutions to accelerate market data systems without requiring the implementation of new wire protocols or changing applications. With off the shelf 10GbE gear, Solarflare's server adapter and the Force10 switch can be used as the foundation of messaging systems for electronic trading and the support of low-latency UDP multicast with no need to re-write applications or use proprietary, specialized hardware. IT and data center managers can deploy plain old Ethernet solutions today.

Moving forward, Solarflare Communications and Force10 also expect high performance 10G Ethernet solutions with low-latency UDP multicast to become an important technology component of public and private clouds that rely on real time media distribution for business to consumer and business to business applications.

## About Solarflare

Solarflare is the pioneer in high-performance, low-latency 10GbE server networking solutions. Our architectural approach combines hardware and software to deliver high-performance adapter products and application acceleration middleware for superior performance in a wide range of applications, including financial services, high performance computing (HPC), cloud computing, storage and virtualized data centers. Solarflare's products are used globally by many of the world's largest companies, and are available from leading distributors and value-added resellers, as well as from Dell and HP. Solarflare is headquartered in Irvine, California and has an R&D site in Cambridge, UK. For more information, please visit [www.solarflare.com](http://www.solarflare.com)

## About Force10

Force10 Networks develops high-performance data center solutions powered by the industry's most innovative line of open, standards-based, networking hardware and software. The company's Open Cloud Networking framework grants Web 2.0/portal operators, cloud and hosting providers, enterprise and special-purpose data center customers new levels of flexibility, performance, scale and automation—fundamentally changing the economics of data center networking. Force10 Networks operates globally, providing 24x7 service and support to its customer base in more than 60 countries worldwide. For more information, visit [www.force10networks.com](http://www.force10networks.com).

## About the Author

Bruce Tolley is responsible for solutions marketing at Solarflare including technical, event, and partner marketing activities. Previously, he served as Solarflare's Vice President of Marketing, and earlier Director of Product Management. Prior to joining Solarflare, Bruce was a Senior Product Line Manager at Cisco Systems where he managed the Ethernet transceiver business that included product life cycle management and the launch of Metro Ethernet, 10 Gigabit, and 1000BASE-T switch solutions. Prior to Cisco, he served in various product and marketing



management roles at 3Com Corporation. Formerly Study Group Chair of the IEEE 802.3aq 10GBASE-LRM standards project, Bruce is a frequent contributor to the IEEE 802.3 Ethernet standards projects. He is currently serving as Secretary and Director of the Ethernet Alliance. He is an alumnus of Selwyn College, Cambridge University and Tuebingen University, Germany and holds MA and Ph.D degrees from Stanford University and an MBA from Haas School of Business, UC Berkeley.

#### **Appendix: List of Benchmarks**

Solarflare's test procedures are documented in its "Low Latency Quickstart Guide" available for download from the driver download page at [www.solarflare.com](http://www.solarflare.com)

**For more information on Solarflare products, visit <http://www.solarflare.com> or contact [productinfo@solarflare.com](mailto:productinfo@solarflare.com).**

**Information in this document is provided in connection with Force10 products. For more information, visit Force10 at <http://www.Force10networks.com>, or contact us at [sales@Force10networks.com](mailto:sales@Force10networks.com).**