Benchmarking Next-Generation Switch Fabrics

Itamar Elhanany, University of Tennessee at Knoxville
Vahid Tabatabaee, University of Maryland

Switch fabrics are a principal building block in networking and communications platforms, but the growing use of merchant fabric silicon for diverse market segments is making it increasingly challenging to evaluate and compare the various product offerings.

Historically, original equipment manufacturers (OEMs) have been hesitant to adopt merchant fabrics into their strategic platforms because of the difficulty in accurately assessing performance and the lack of clear differentiating features. Next-generation fabrics incorporate many unique capabilities that have made the process even more expensive and time-consuming.

Current fabric selection methodology involves complex comparisons of speeds and feeds using limited data that switch-fabric vendors provide. This data is commonly based on idealistic traffic patterns and environmental parameters suited to a vendor-specific architecture rather than real-world, application-oriented scenarios that stress fabric implementations.

To address this problem, the Network Processing Forum (www.npforum.org) has launched a task group to develop a standard fabric benchmarking framework and suite of performance test benches that provide system OEMs with open, objective, and verifiable results while enabling fabric vendors to leverage their core intellectual property.

The Network Processing Forum's switch fabric benchmarking framework will make it easier to evaluate different commercial offerings.

The task group is focusing on traffic modeling, performance metrics, and actual test benches. Although the emphasis is on switches specifically pertaining to Internet-based platforms, the same framework is applicable to storage area networks and other switching applications.

TRAFFIC MODELS

Evaluating switch fabrics requires a comprehensive set of data-traffic models. In a typical CSIX (common switch interface) fabric-based platform, the traffic manager is responsible for scheduling and inserting datagrams called CFrames. The fabric is transparent to the original packet protocol, which may include asynchronous transfer mode, IPv4, or IPv6; its primary task is to allow CFrames to traverse from the ingress line cards to the egress line cards.

Modeling real-life data traffic is tremendously difficult because it varies greatly with network type, topology, time of day, and many other parameters. From its inception, the NPF task group has sought to develop a set of statistical characteristics for explicitly describing each traffic model.

Several key questions must be addressed when modeling data traffic:

- What is the traffic load's behavior?
- Is the traffic bursty? How correlated are packet arrivals?
- How is the traffic distributed among the available output ports? While a uniform distribution is an easy case study, a pragmatic analysis should consider other more elaborate models.
- How many quality-of-service classes are there, and how is the traffic distributed among them?
- What portion of the traffic is multicast, and what are its characteristics?

By strictly defining the answers to these questions, the NPF task group can characterize data traffic in an unambiguous manner.

In recent years, academic and industry researchers have focused on modeling Internet traffic. Researchers can use several core statistical models to analyze generic traffic patterns. Rather than trying to accurately portray network traffic behavior, the NPF task group has applied a wide range of diverse scenarios that accentuate the differences among various fabrics.

PERFORMANCE METRICS

In today's environment, it is often possible for one fabric vendor to present benchmark results that are measured and calculated in a way that is

very different from those that other vendors present, making it difficult to accurately compare products.

Naturally, the definition of each metric should be expressed in terms of the actual benchmarks executed. To that end, the NPF's goal is to establish an unambiguous specification that defines performance metrics such as latency and jitter as well as to delineate the way vendors should measure them. This will minimize the amount of tweaking by fabric vendors when they present benchmark results intended to comply with the NPF standard.

Although the specifications aim to stringently define all system parameters and environmental conditions under which the vendors carry out the test benches, they also are pragmatic enough to give vendors considerable flexibility in configuring their products to optimize performance while still adhering to an "apples-to-apples" comparison. Fabric providers and equipment manufacturers can thus maximize the gain from these benchmarks.

Because the fabric is an integral part of a switch/router system, isolating it from the rest of the system is quite difficult. For example, when measuring latency, it is necessary to take into account the response of complementing components such as traffic managers to the backpressure the fabric applies. The NPF specifications reflect such considerations.

BENCHMARK TEST SUITES

The benchmark test suites are a set of well-defined experiments that researchers use to study, evaluate, and compare various switch fabrics under different traffic scenarios. Each test bench focuses on one aspect of the switch fabric under a specific scenario. Together, the test benches are intended to be comprehensive enough to provide a good understanding of the general characteristics and behavior of each fabric.

To ensure objectivity, the tests include a standard format for plotting and reporting the results; a standard format for listing all parameters required to specify the traffic models; uniform attributes such as test duration, the data recording procedure, and CFrame contents; and design requirements for the hardware platform used to perform the tests.

The NPF benchmarking standards should have a significant positive impact on future fabric deployment.

Each suite outlines an objective that lets customers determine how much of the test pertains to their own platform and what features of the switch fabric are under scrutiny. Based on these objectives, customers can request the switch-fabric vendor to run a second set of test suites customized for their particular platform.

The test suites are divided into four groups that each focus on a different aspect of the switch fabric:

- *Hardware benchmarks* evaluate memory speed, CFrame processing speed, minimum latency, and other basic hardware aspects when there is no contention in the fabric.
- *Arbitration benchmarks* consider port-level contention between CFrames. Switch-fabric arbitration mechanisms play a key role in queue management and thereby greatly impact overall system performance. A considerable source of queuing occurs when CFrames in several ingress line cards have the same destination or when multiple CFrames in the same ingress line card have different destinations. In either case, the arbitration mechanisms must determine which set of contending CFrames is granted service and forwarded to its destination.
- *Scheduling benchmarks* evaluate switch-fabric mechanisms that support multiple classes of service to provide differentiated service levels. These tests study how effectively a switch fabric controls key performance metrics such as latency and the impact on other classes of traffic.
- *Multicast benchmarks* study fabric performance under multicast traffic, which is naturally dissimilar to unicast traffic and has its own unique challenges.

Dedicated specification documents define the traffic arrival models used in each test suite as well as the performance metrics measured in each test.

By applying the same input stimuli and identically interpreting the performance metrics, switch-fabric customers can accurately compare multiple product offerings. Wide adoption of the NPF benchmarking standards should have a significant positive impact on future fabric deployment, enabling next-generation network equipment to optimize performance in a cost-efficient manner. ■

Itamar Elhanany *is an assistant professor in the Department of Electrical and Computer Engineering at the University of Tennessee at Knoxville. Contact him at itamar@ieee.org.*

Vahid Tabatabaee *is a PhD candidate in the Department of Electrical and Computer Engineering at the University of Maryland, College Park. Contact him at vahid@Glue.umd.edu.*

Editor: Upkar Varshney, Department of CIS, Georgia State University, Atlanta, GA 30002-4015; uvarshney@gsu.edu