†

The SHIELD: Self-Healing Interconnect

Background

High performance computing (HPC) has always been used to solve complex problems. However, with the current trend of exponential data growth, new magnitudes of cluster computing scale are now necessary to tackle the computational challenges of today and tomorrow. Scientific computing can analyze ever larger and higher-resolution models, while machine learning and artificial intelligence are increasingly becoming pervasive techniques on their own. Hybrid software techniques, which combine these methodologies, are creating new challenges for HPC applications.

In HPC, clusters have depended on a high-speed and reliable interconnect. Efficient inter-process communications depend on an interconnect fabric that is capable of high bandwidth and low latency while supporting a massive number of endpoints (compute and storage servers). MPI, SHMEM/PGAS, and UPC, fast access to large scale, shared storage, machine learning frameworks, and even new heterogeneous computing architectures all share common characteristics and requirements for a robust and resilient network in order to achieve maximum scalable performance.

InfiniBand – the Network of Choice for High Performance Computing

For many years, InfiniBand has been the network of choice for cluster computing, delivering services at a very high degree of reliability and performance. As InfiniBand networks continue to expand to accommodate higher scales of computation and storage capacity, the increase in adapters and switches, and in particular, the cables accompanying that growth, will occasionally sustain physical or electrical damage.

Traditional software mechanisms addressing this issue include job checkpointing, which creates a point-in-time snapshot of the computation. If the computation fails at a later point, the job will resumes from the last successful state and point in time. Of course, network protocols have data integrity checks and retransmission mechanisms, but these methods all have a negative impact on performance and are impractical at very large scales.

In the case of today's networks, a traditional subnet manager will recognize failed links and recalculate routes to avoid the problem, but this can take up to 5 seconds for 1,000 nodes and 30 seconds for clusters with 10,000 or more endpoints – certainly not fast enough to ensure the seamless integrity of a running computation. In fact, no software mechanism can be responsive enough at very large scales to detect and fix fabrics that suffer from a link failure.

To address this problem, Mellanox designed a new and innovative solution called SHIELD™—(Self-Healing Interconnect Enhancement for Intelligent Datacenters), which takes advantage of the intelligence already built into the latest generation of InfiniBand switches. By making the fabric capable with self-healing autonomy, the speed with which communications can be corrected in the face of a link failure can be sped up by 5000x, fast enough to save communications from expensive retransmissions or absolute failure. Having introduced SHIELD in its EDR 100Gb/s Switch-IB® 2 switch devices, SHIELD now also appears in Mellanox's HDR 200Gb/s generation of switch devices.

† For illustration only. Actual products may vary.

SHIELD supports two mechanisms for communications recovery. The first and simplest case is one in which a switch has more than one forward route to the desired destination. In this case, the switch can make an independent decision to forward the packet out an alternate port that sets it on the new viable route.

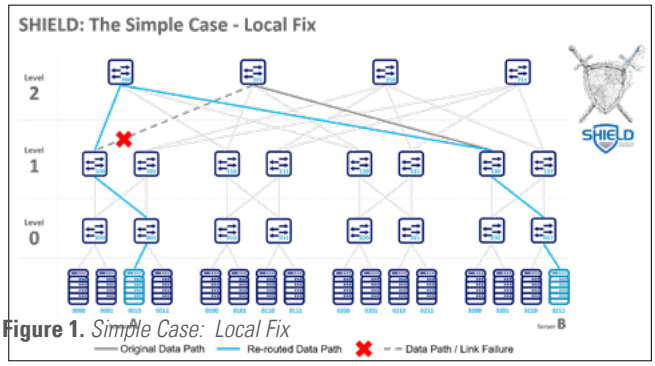

Figure 1. *Simple Case: Local Fix*

However, there are cases in which no alternative routes to the desired endpoint are available from the switch that experiences a failed link, such as a downstream switch in a Fat-Tree network topology. In this case, the switch can pass in-band information to another switch in the network, a switch that can select the most efficient alternative route so it can then take responsibility for rerouting the traffic. The total time required to perform this action is on the order of 1 microsecond, quick enough to allow communications that depend on those connections to continue successfully.

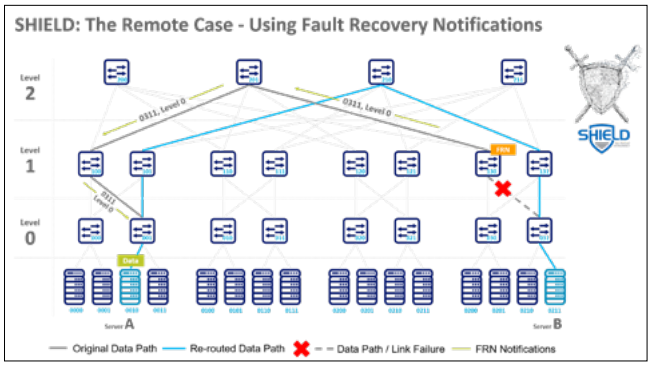

Figure 2. *Remote Case: Using Fault Recovery Notifications*

Conclusion

Networks have reached a greater degree of autonomy, detecting and correcting link failures and ensuring reliable, high performance data delivery, even as the industry approaches Exascale computing.

With SHIELD, Mellanox InfiniBand solutions are unleashing the possibilities of next generation high performance computing, today.

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