
Plan B: Interruption of Ongoing MPI Operations to Support Failure Recovery

Aurélien Bouteiller*¹

¹Innovative Computing Laboratory (ICL) – 1122 Volunteer Blvd 203 Claxton Complex 37996-3450 TN,
USA, United States

Abstract

Advanced failure recovery strategies in HPC system benefit tremendously from in-place failure recovery, in which the MPI infrastructure can survive process crashes and resume communication services. We will present the rationale behind the specification of ULFM fault tolerant MPI, and outline some application results. We will also present a scalable reliable broadcast, with a limited degree, that supports the ‘Revoke’ ULFM MPI operation. Evaluation at scale, on a Cray XC30 supercomputer, demonstrates that the Revoke operation has a small latency, and does not introduce system noise outside of failure recovery periods.

*Speaker