Fault Tolerance for PetaScale Systems: Current Knowledge, Challenges and Opportunities

Franck Cappello

INRIA
fci@lri.fr

Abstract. The emergence of PetaScale systems reinvigorates the community interest about how to manage failures in such systems and ensure that large applications successfully complete. Existing results for several key mechanisms associated with fault tolerance in HPC platforms will be presented during this talk. Most of these key mechanisms come from the distributed system theory. Over the last decade, they have received a lot of attention from the community and there is probably little to gain by trying to optimize them again. We will describe some of the latest findings in this domain. Unfortunately, despite their high degree of optimization, existing approaches do not fit well with the challenging evolutions of large scale systems. There is room and even a need for new approaches. Opportunities may come from different origins like adding hardware dedicated to fault tolerance or relaxing some of the constraints inherited from the pure distributed system theory. We will sketch some of these opportunities and their associated limitations.