Introduction to the Special Issue on SYSTOR 2017

Two of the articles in this edition of *ToS* were originally presented in shorter form at the 10th ACM International Systems and Storage Conference (SYSTOR'17) in Haifa, Israel in May 2017, and were selected based on feedback from members of the Program Committee and relevance to this journal.

Direct memory access over the PCIe bus has enabled new facilities in both networking and storage, with RDMA allowing direct access to remote memory without CPU intervention, and NVMe providing a multi-queue storage interface optimized for fast devices, rather than the slow disks for which earlier interfaces were designed. The combination of the two offers the potential of extremely high performance access to networked storage, but the overhead of today's storage and networking stacks prevents delivery of much of this performance. In "[FlashNet: Flash/Network Stack Co-Design](#)" a team of researchers (Trivedi et al.) from IBM Research and ETH Zurich present a combined storage and networking stack for NVMe and RDMA, with cross-layer optimizations, delivering substantial improvements in bandwidth, latency, and IOPS relative to custom code over existing stacks; unlike custom code, FlashNet is shown to be easy to use in creating distributed applications.

An alternate approach to making use of RDMA and NVMe is NVMe-over-fabrics, which makes use of the PCIe mapping of NVMe and the RDMA capabilities of many modern NICs to bypass the remote CPU, and associated latency, when accessing storage over the network. Although widely discussed (and even hyped) in the industry press, until recently the only way to get solid information on performance and behavior of these systems was to buy and measure one yourself. In "[Performance Characterization of NVMe-over-Fabrics Storage Disaggregation](#)", Guz, Li, Shayesteh, and Balakrishnan from Samsung provide a solid introduction to NVMe-over-fabrics, as well as comparisons of NVMeoF with both direct access NVMe and remote iSCSI in cases ranging from micro-benchmarks to a full disaggregated key/value store, with measurements being performed on a sufficiently capable system (100Gbit/s, 750K/120K IOPS) to ensure that the results do not become rapidly outdated.

We invite all readers to enjoy these papers that combine a practical bent with deep innovation.

Peter Desnoyers, Eyal de Lara
Program Committee of SYSTOR 2017