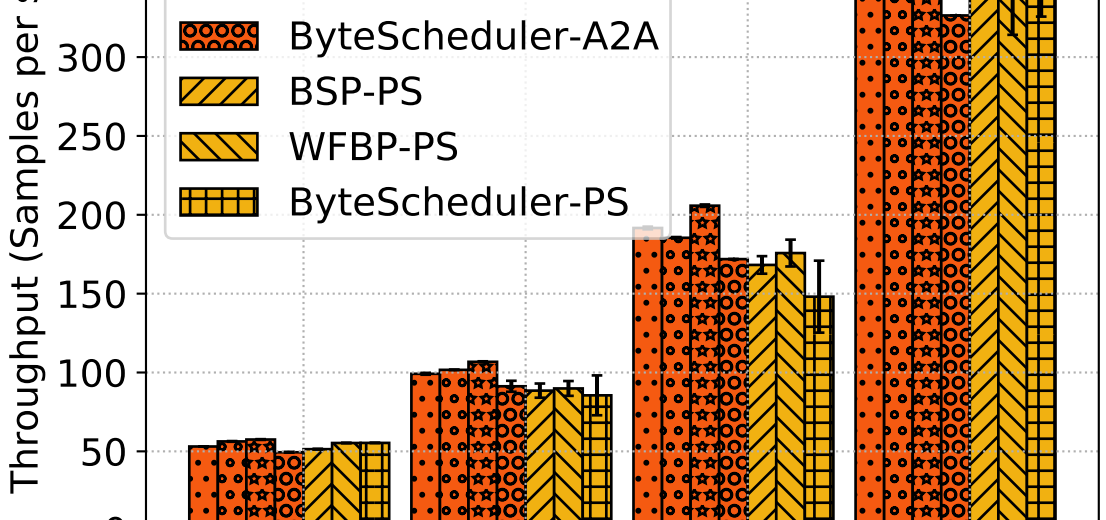


BERT-Large (BS=8) on 100GbIB with RDMA



Throughput (Samples per second)

of GPUs

4.0x

8.0x

15.9x

31.6x