21st IEEE Real-Time Conference 2018, Williamsburg, VA, USA

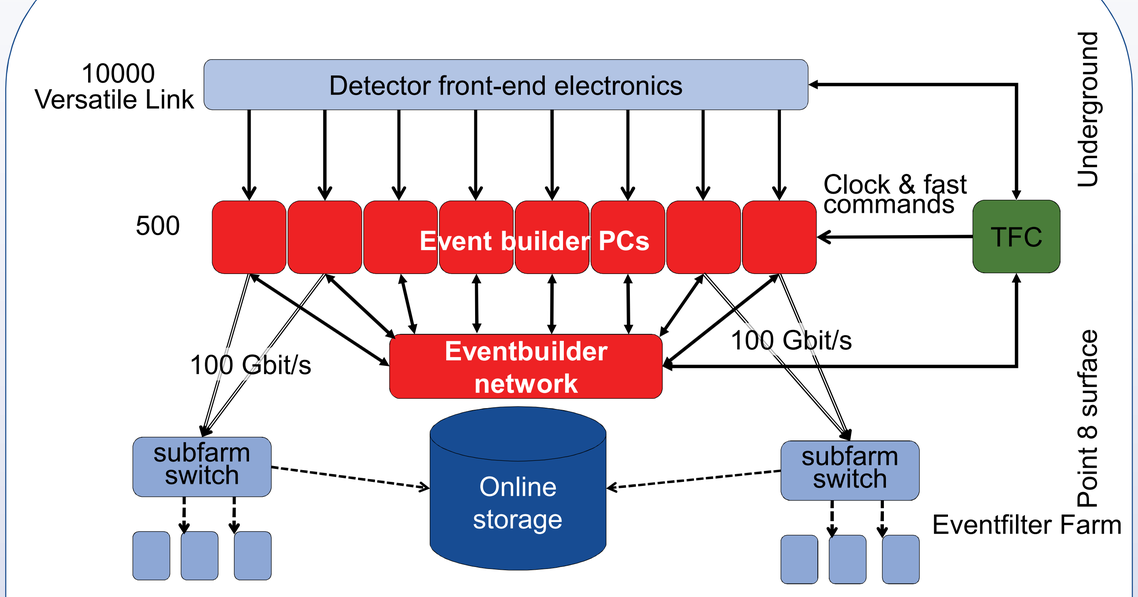
FLIT-level Infiniband network simulations of the DAQ system of the LHCb experiment for Run-3

T. Colombo[1], P. Durante[1], D. Galli[2], M. Manzali[2], U. Marconi[2], N. Neufeld[1], F. Pisani[1,2], R. Schwemmer[1], S. Valat[1]

[1] CERN, Geneva, Switzerland

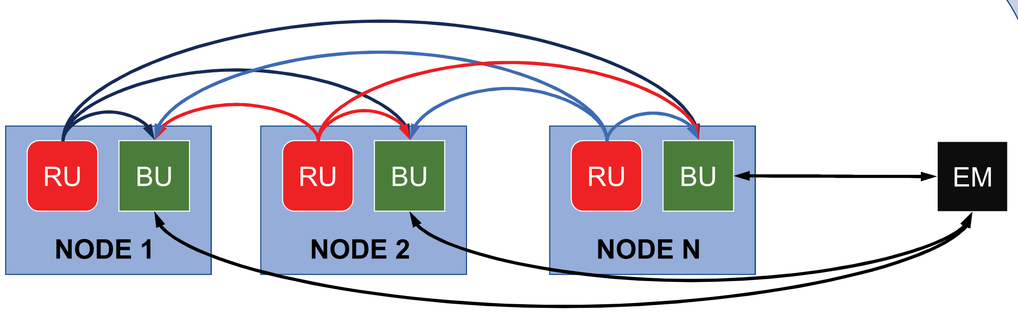
[2] INFN & Università di Bologna, Bologna, Italy

LHCb DAQ and trigger system for RUN3



- 10000 Versatile Links terminated into custom FPGAs
- 500 Event builder nodes interconnected by a non blocking 100 Gb/s fabric (e.g. Infiniband or Omnipath)
- 40 MHz event building rate

Event building: architecture



- Readout Unit (RU): collect the fragments from the DAQ board and sends them to the BUs
- Builder Unit (BU): receive and aggregate the fragments into full events
- Event Manager (EM): assign events to BUs

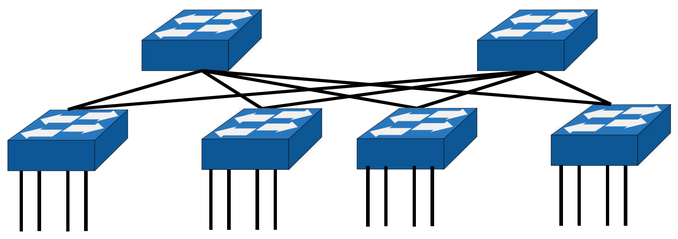
This traffic pattern tends to create congestion

Event building benchmark: DAQPIPE

DAQPIPE is an event building benchmark for the LHCb experiment. It implements all the EB traffic described above in a flexible and configurable way. In this work we will focus in particular on two parameters:

- Credits: number of events processed in parallel by the BU
- Parallel sends: number of fragments on the same event in flight

Event building: fat-tree network topology

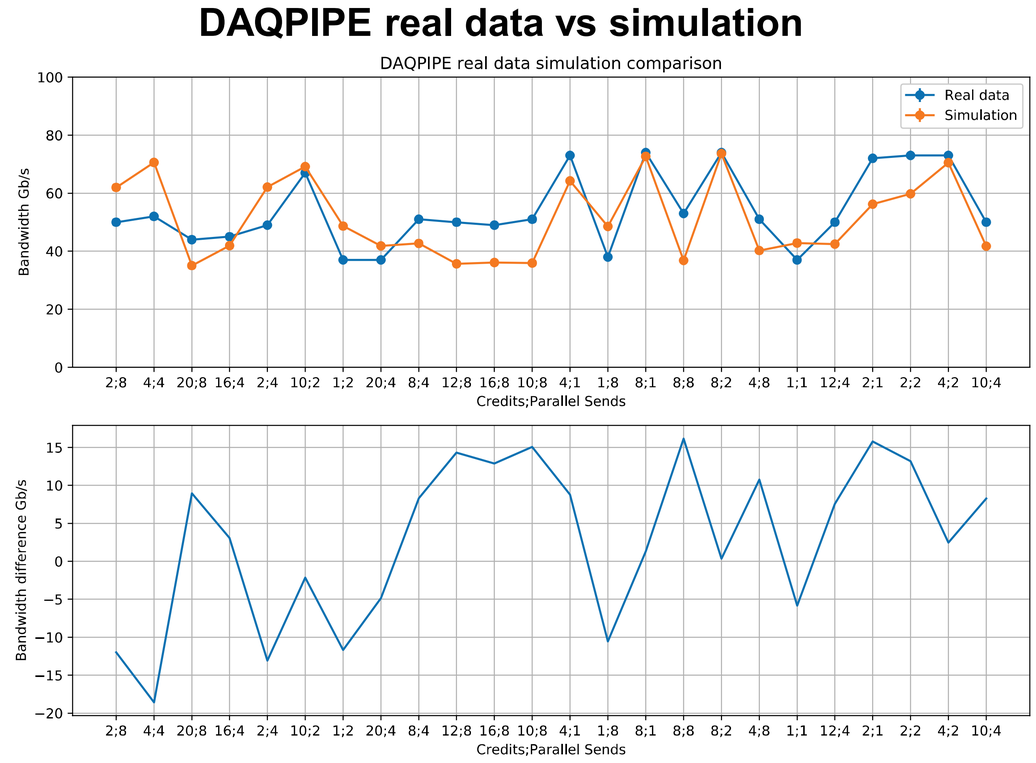


- It provides full bisection bandwidth
- The regular topology helps with traffic shaping
- It is widely used and supported by vendors

Infiniband and event building simulation model

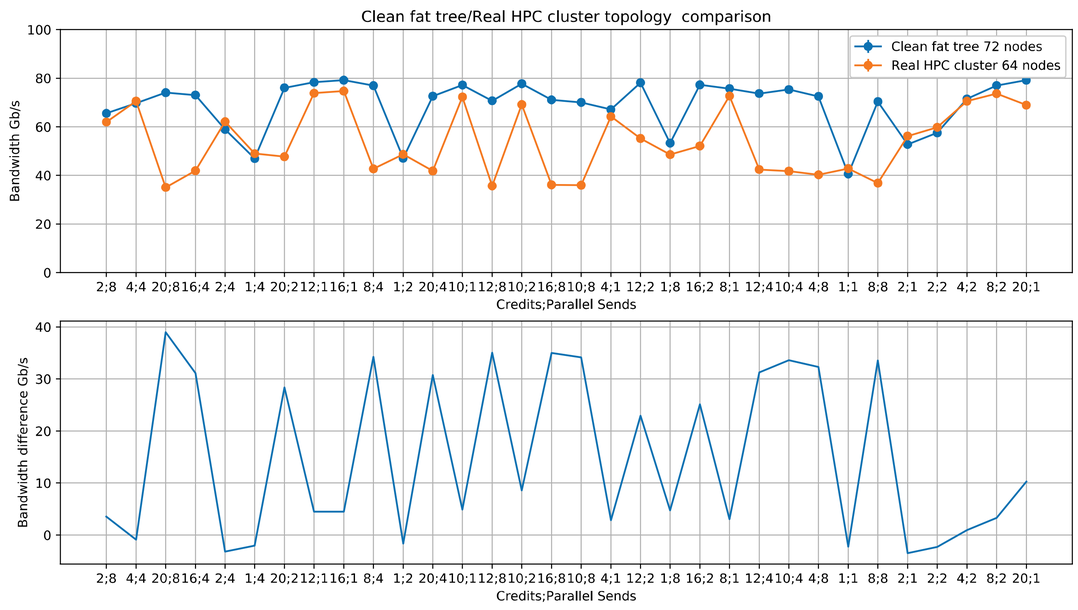
- Low level Infiniband model developed by Mellanox in OMNeT++
- Extensive reverse engineering and tuning of the model to match the behavior of real Infiniband EDR (100 Gb/s) systems
- Implementation of accurate event building traffic injectors
- Network topology building for both parametric ideal ones and topologies extracted from existing HPC clusters
- Model validation against real data
- Simulation of systems larger than the HPC clusters we can access today

DAQPIPE real data vs simulation



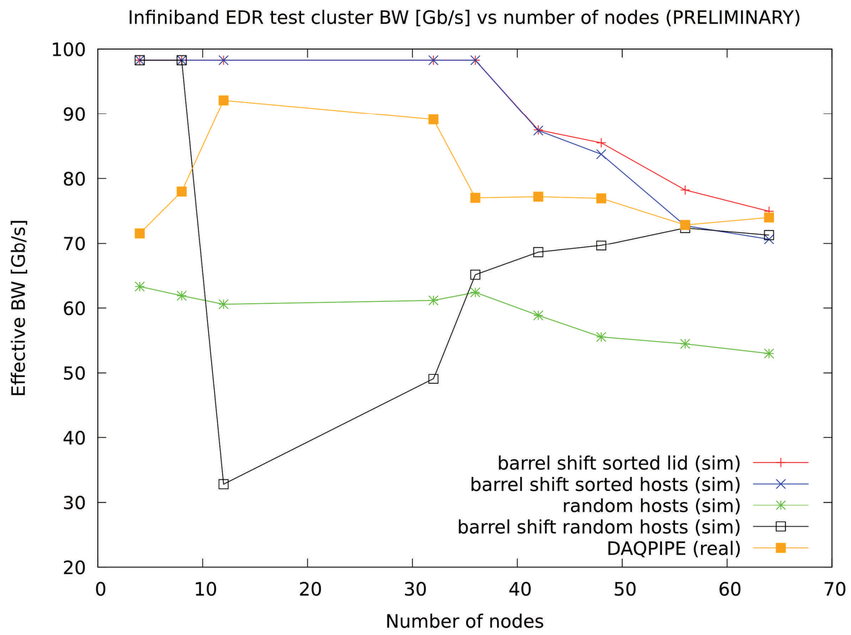
This plot shows the comparison between a real DAQPIPE run and a simulated one using OMNeT++, for different parameters. In both cases the underlying network topology is an imperfect fat-tree of 64 nodes interconnected by Infiniband EDR (100 Gb/s).

DAQPIPE clean vs imperfect fat-tree



This plot shows the comparison between two runs of the OMNeT++ simulated DAQPIPE with different Infiniband EDR based network topologies: a fat-tree of 72 nodes and imperfect fat-tree of 64 nodes.

Synthetic load on imperfect fat-tree



This plot shows the comparison between an ideal barrel shifter simulated with OMNeT++ and a real DAQPIPE run, in both cases the topology is an imperfect fat-tree of 64 Infiniband EDR nodes.