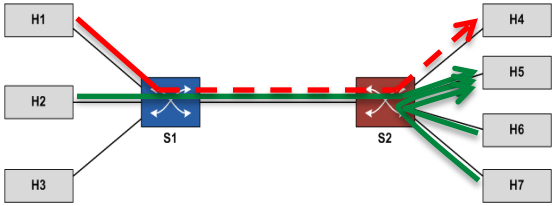
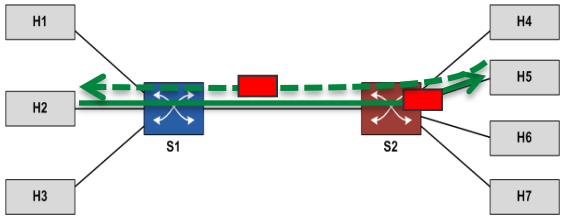
First Experiences with Congestion Control in InfiniBand Hardware

Ernst Gunnar Gran
Simula Research Laboratory

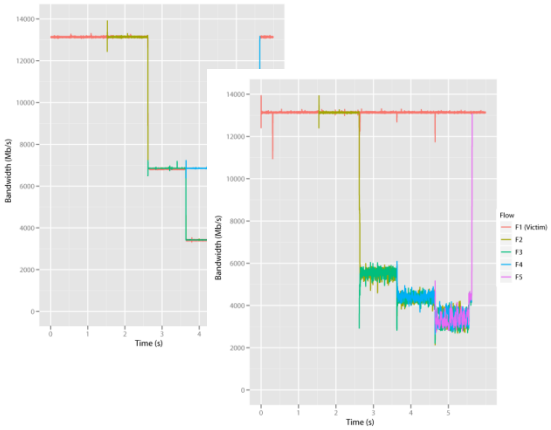
Presentation Outline



Congestion

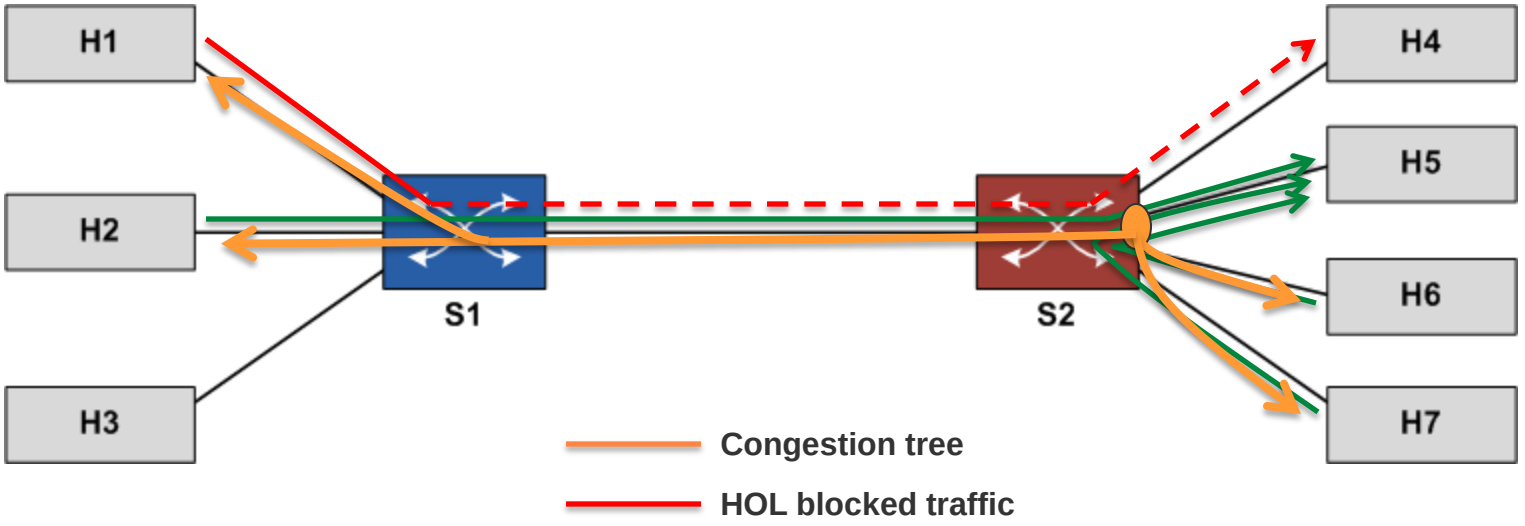


Congestion Control in InfiniBand



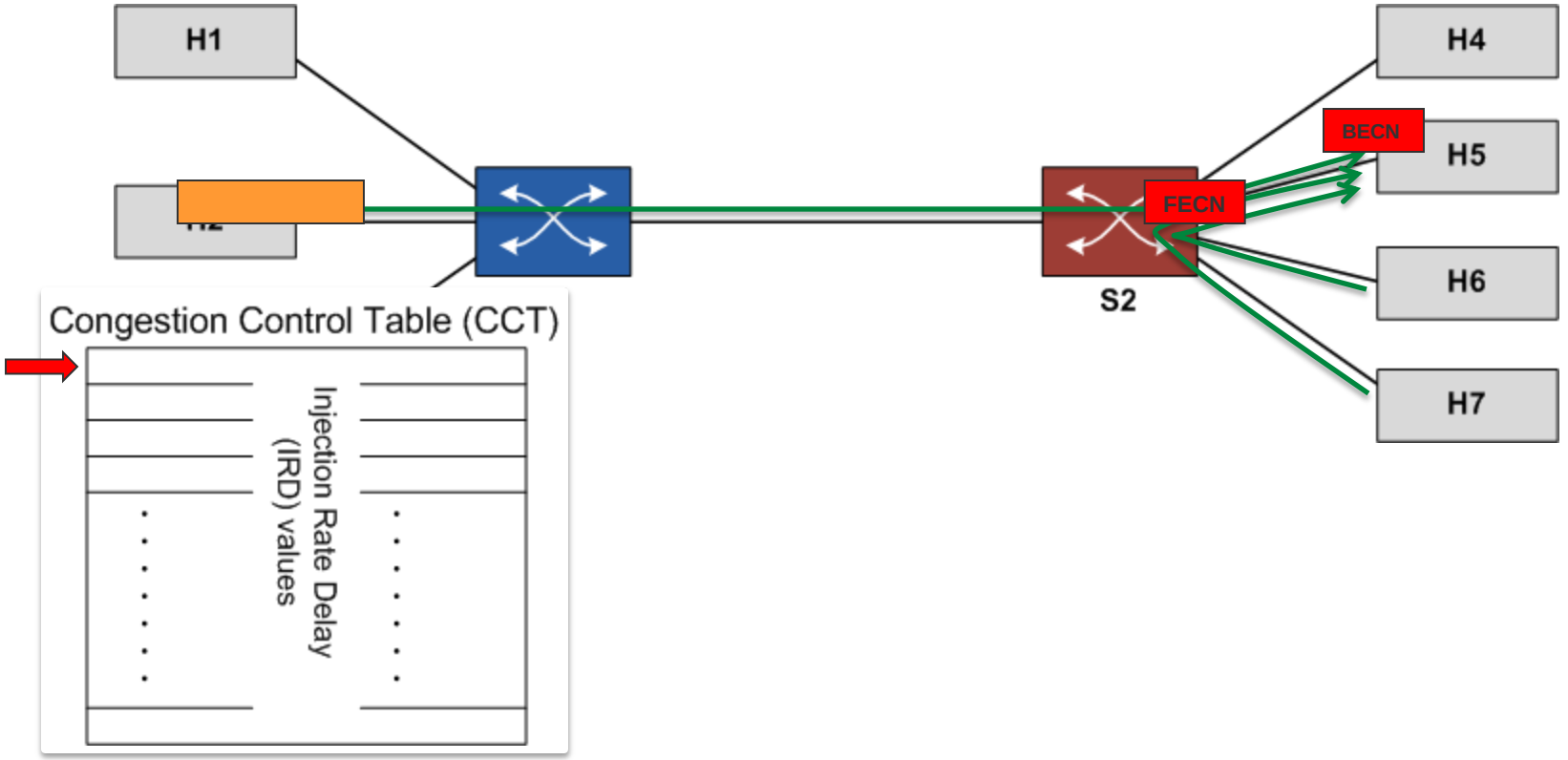
Experiment results

Shared network resources could lead to network congestion and head-of-line (HOL) blocking.

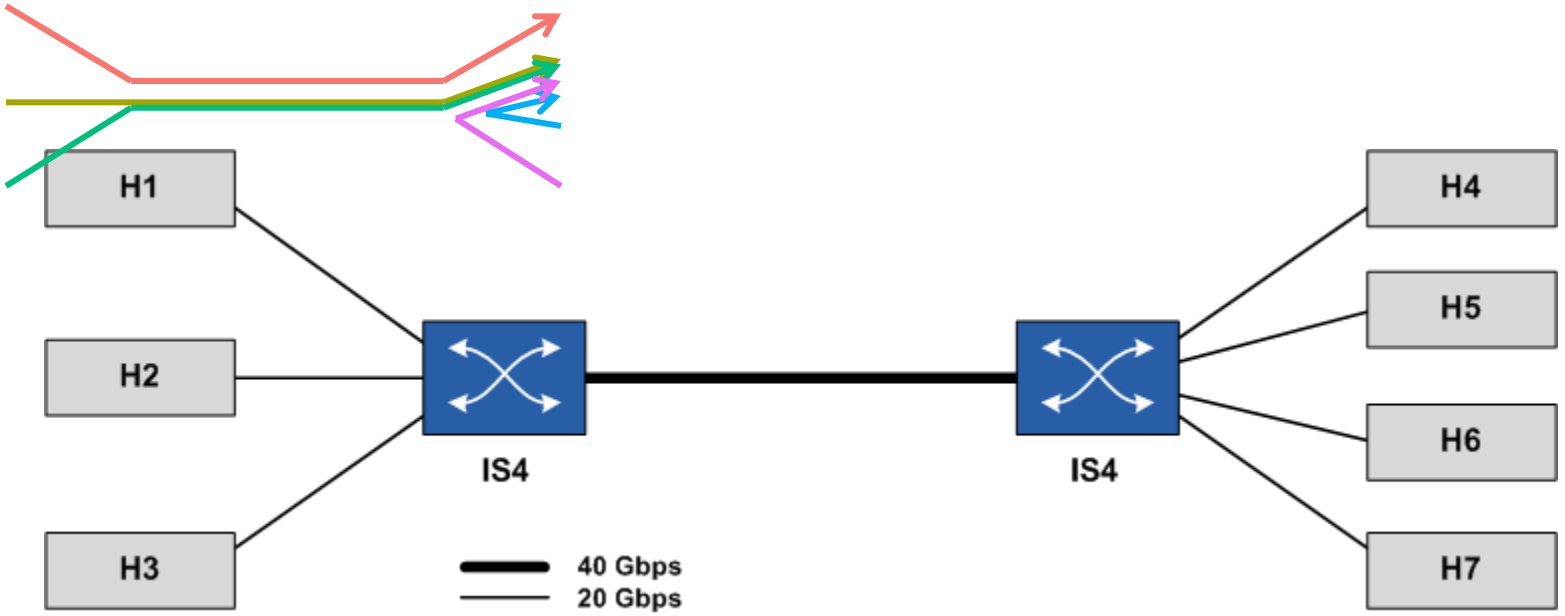


To avoid performance degradation, the HOL blocking must be removed.

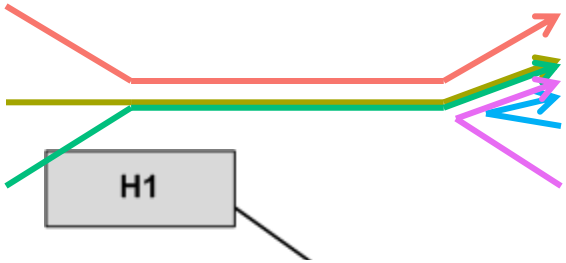
The InfiniBand CC mechanism relies on a closed loop feedback control systems to remove the congestion tree.

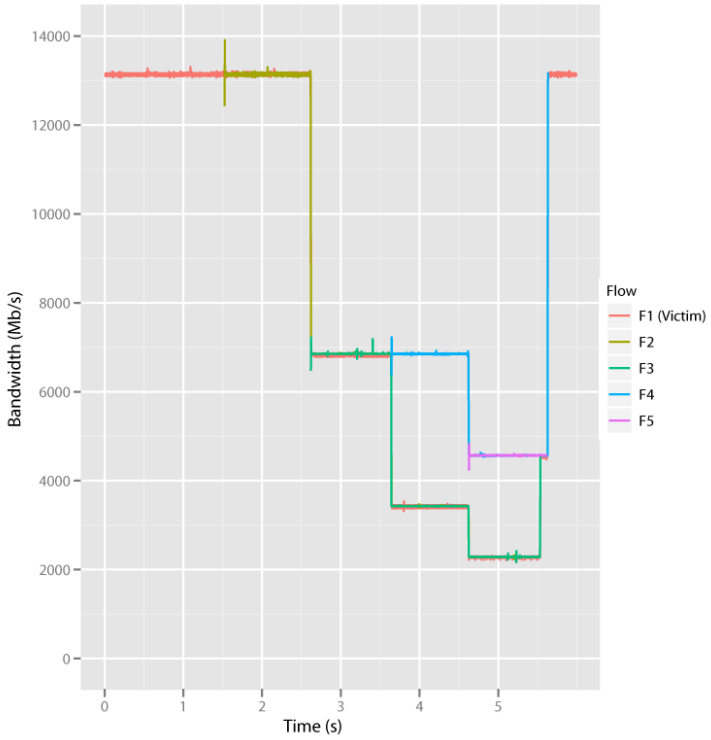


Experiments show that the HOL blocking leads to performance degradation when CC is not activated.



The InfiniBand CC mechanism is able to remove both the HOL blocking and the parking lot problem.





H4

Parameter Values

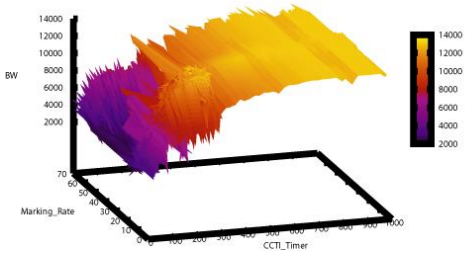
Threshold	15
Marking Rate	1
Packet Size	8
CCTI Increase	1
CCTI Limit	127
CCTI Min	0
CCTI Timer	150

The experiments repeated with the HOL blocked victim flow replaced by the HPCC benchmark.

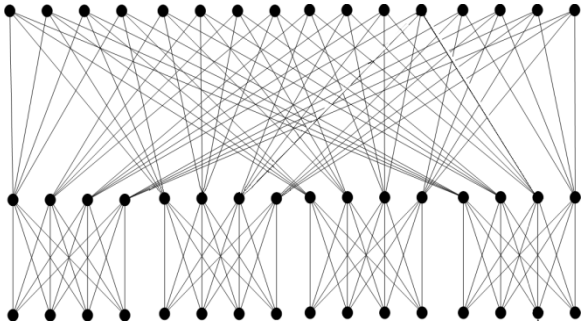
Network Lat. And BW	a) No cong.	b) Cong, CC off	c) Cong, CC on	Impr.
Min Ping Pong Lat. (ms)	0.001132	0.001192	0.001172	1.7%
Avg Ping Pong Lat. (ms)	0.001678	0.012385	0.001729	86.0%
Max Ping Pong Lat. (ms)	0.001957	0.018001	0.002056	88.6%
Naturally Ordered Ring Lat. (ms)	0.002193	0.011396	0.002098	81.6%
Randomly Ordered Ring Lat. (ms)	0.002036	0.011088	0.002073	81.3%
Min Ping Pong BW (MB/s)	880.463	663.235927	876.049	32.1%
Avg Ping Pong BW (MB/s)	1354.021	733.159	1360.26	85.5%
Max Ping Pong BW (MB/s)	1590.559	879.125	1611.025	83.3%
Naturally Ordered Ring BW (MB/s)	742.469675	213.687109	743.769828	248.1%
Randomly Ordered Ring BW (MB/s)	684.66655	350.356751	683.451954	95.1%

Other HPCC Benchmarks	a) No cong.	b) Cong, CC off	c) Cong, CC on	Impr.
PTRANS GB/s	0.755254	0.347585	0.611816	76.0%
HPLinpack 2.0 Gflops	1.819	1.79	1.827	2.1%
MPIRandomAccess Updates GUP/s	0.015118991	0.01195898	0.014409549	20.5%
MPIFFT Gflops/s	1.3768	0.982365	1.36891	39.3%

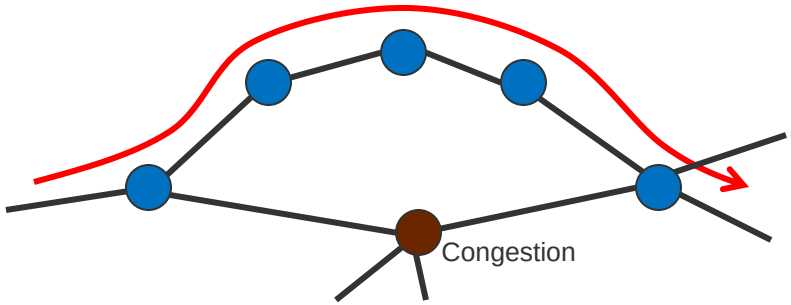
Ongoing research includes both further hardware experiments and simulation studies to:



...explore the CC parameter space



...study CC in larger topologies



...look at adaptive routing as a supplementary mechanism to CC