SFIO a striped file I/O library for MPI

Global Computing Techniques et Applications
Ecole d'ingénieurs de Genève (EIG)
March 19, 2001

Emin Gabrielyan

EPFL, Computer Science Dept.
Peripheral Systems Lab.
Emin.Gabrielyan@epfl.ch

- SFIO as a parallel I/O solution for distributed computing.
- General requirements to the Parallel I/O system and the parallel file striping paradigm
- SFIO library architecture
- SFIO on top of MPICH and on top of MPIFCI, performance on T1
- Swiss-T1's topology. Possible influence to the SFIO performance
- Conclusion and future work

- SFIO can be an I/O solution for Global Computing
- SFIO uses for its transport layer the MPI
- The MPICH version of MPI allows to run SFIO on Internet

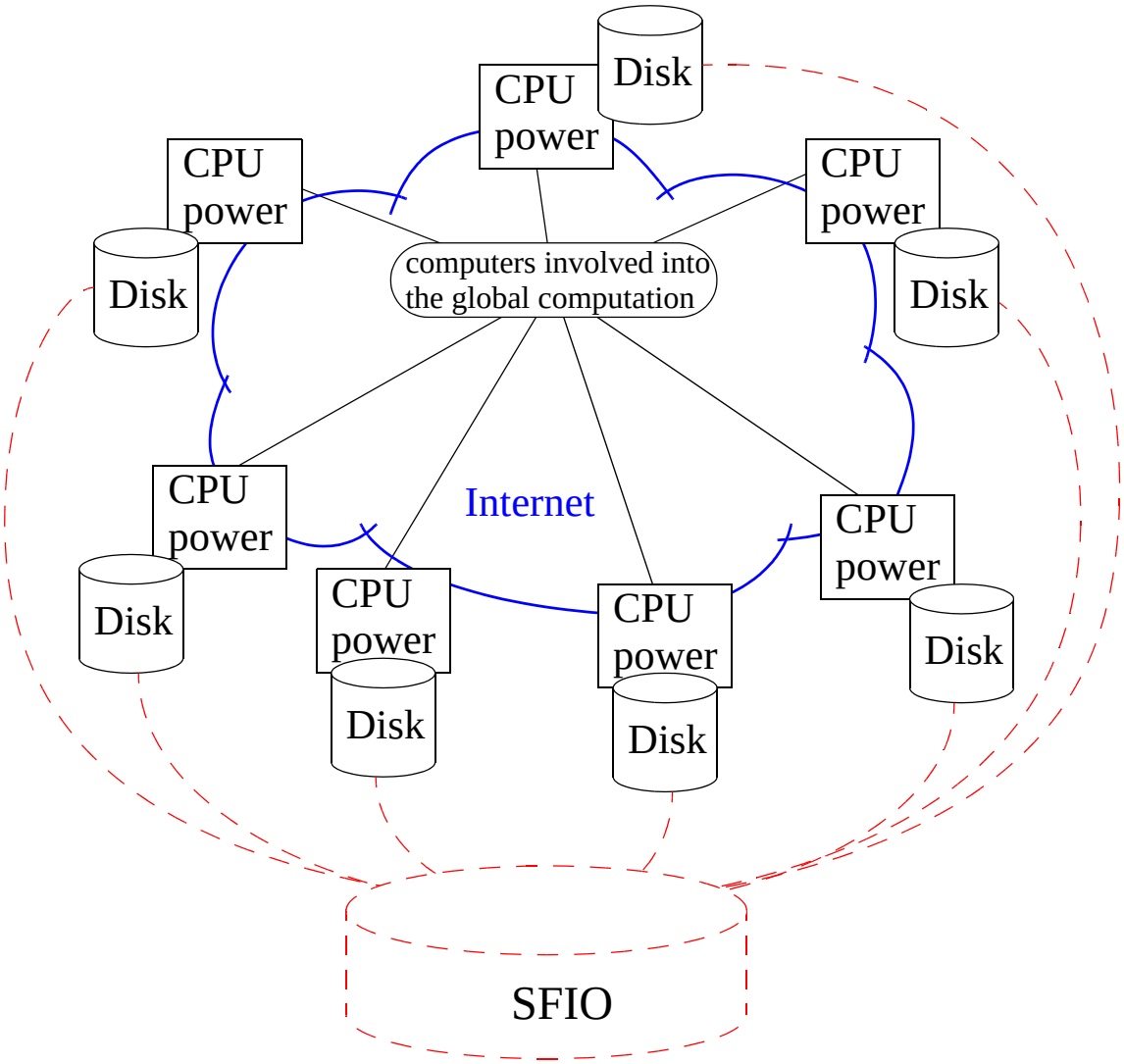


Fig. 1

- For an I/O bound global application a single physical storage in the network is not a scalable solution.
- Parallel I/O systems must exhibit a scalable performance as a function of the number of Compute and I/O nodes.
- Parallel I/O systems should offer highly concurrent access capabilities to a common data file by all global application processes.

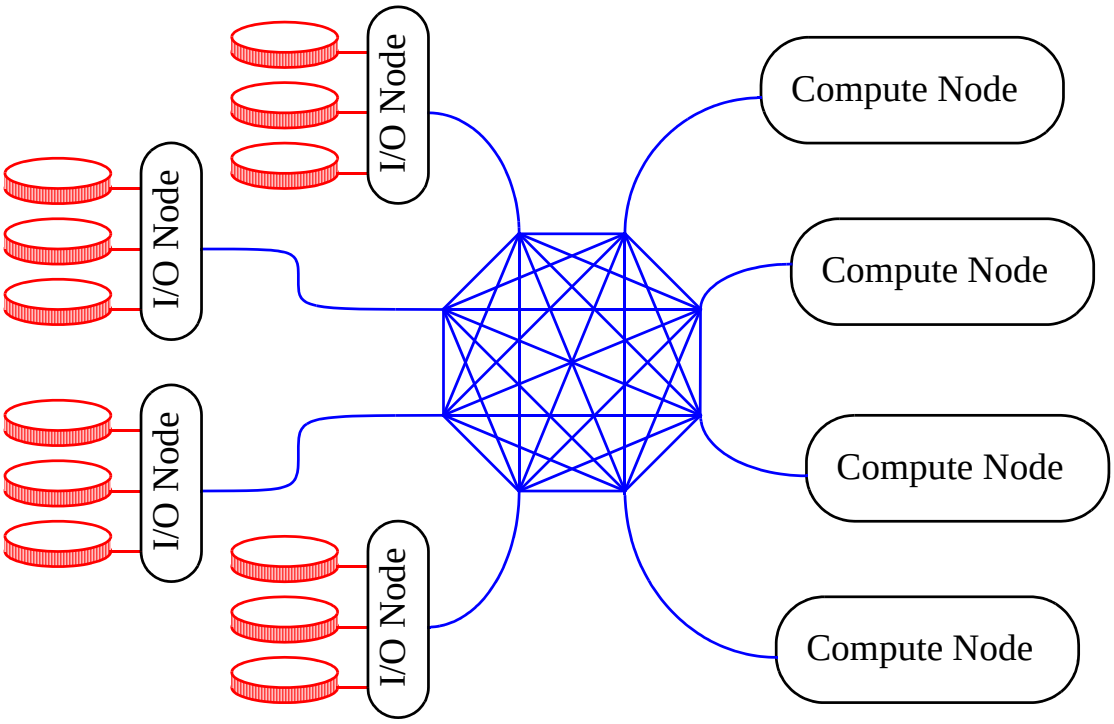
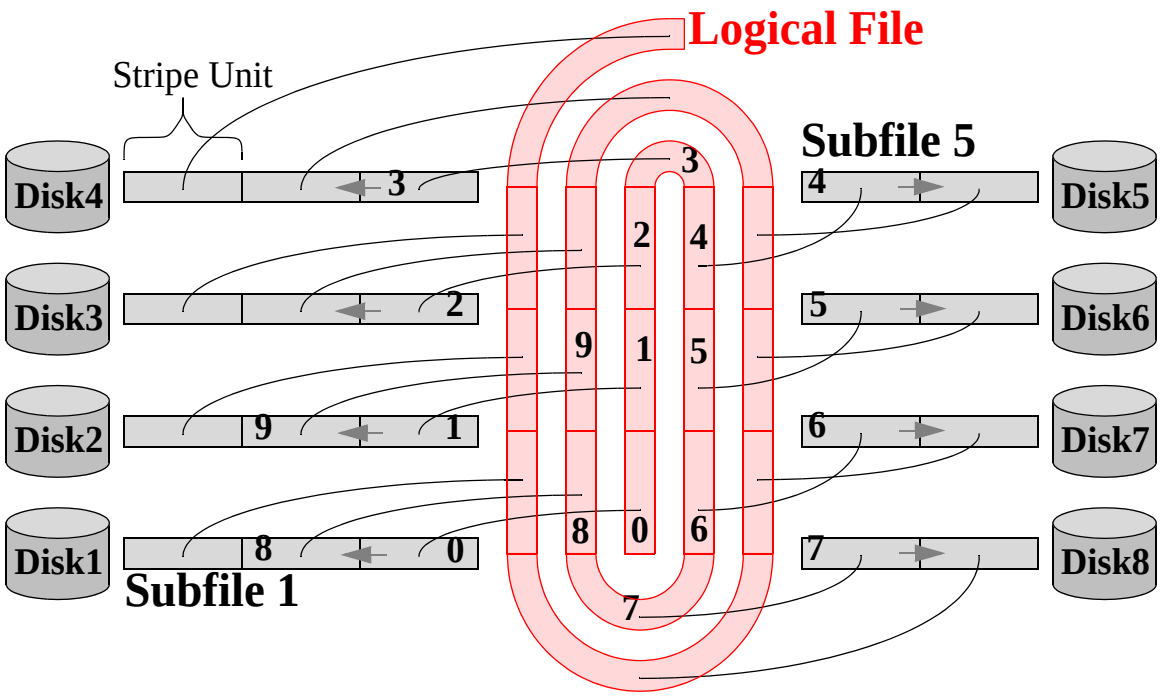


Fig. 2

- Parallelism for input/output operations can be achieved by striping the data across multiple disks so that read and write operations occur in parallel.

- To be able to provide the highest level of parallelization of access requests as well as a good load balance, small striping units are required.



Parallel file striping paradigm

Fig. 3

SFIO library architecture

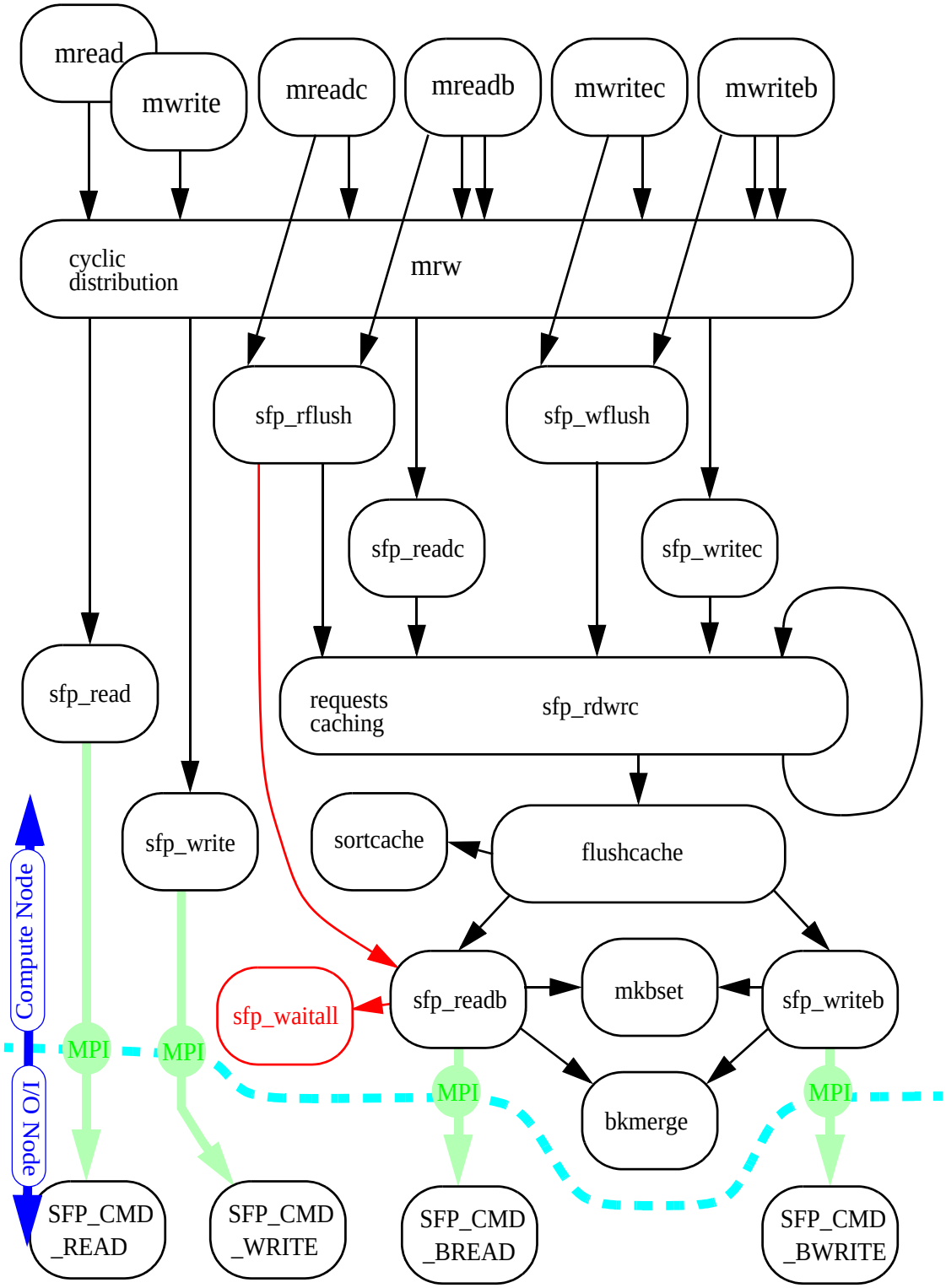
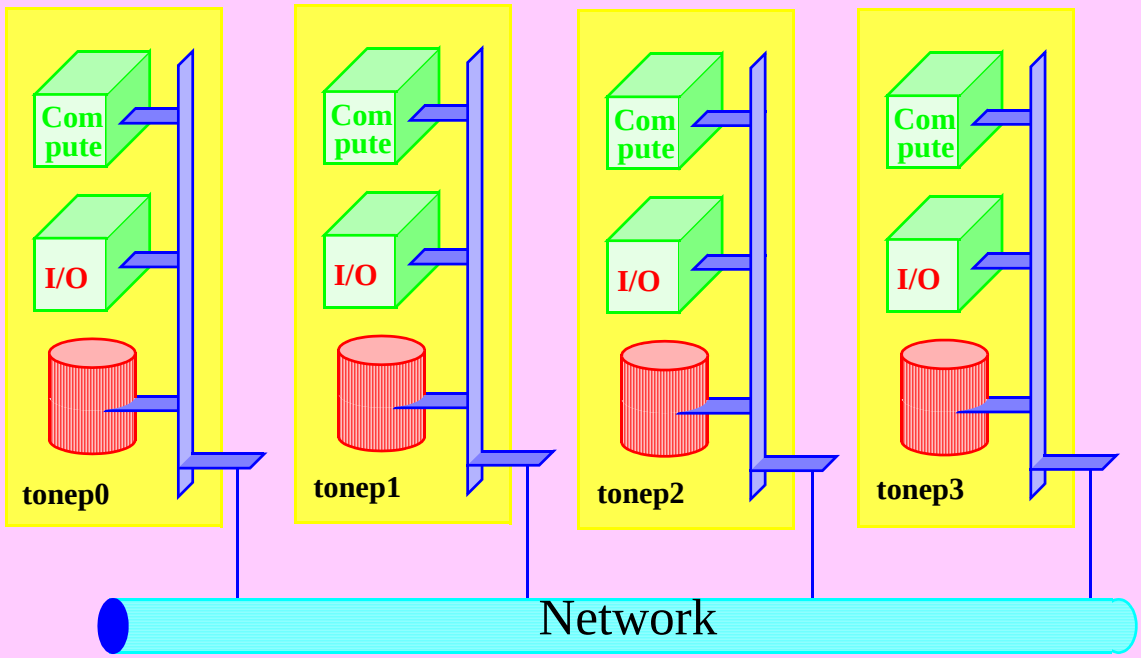


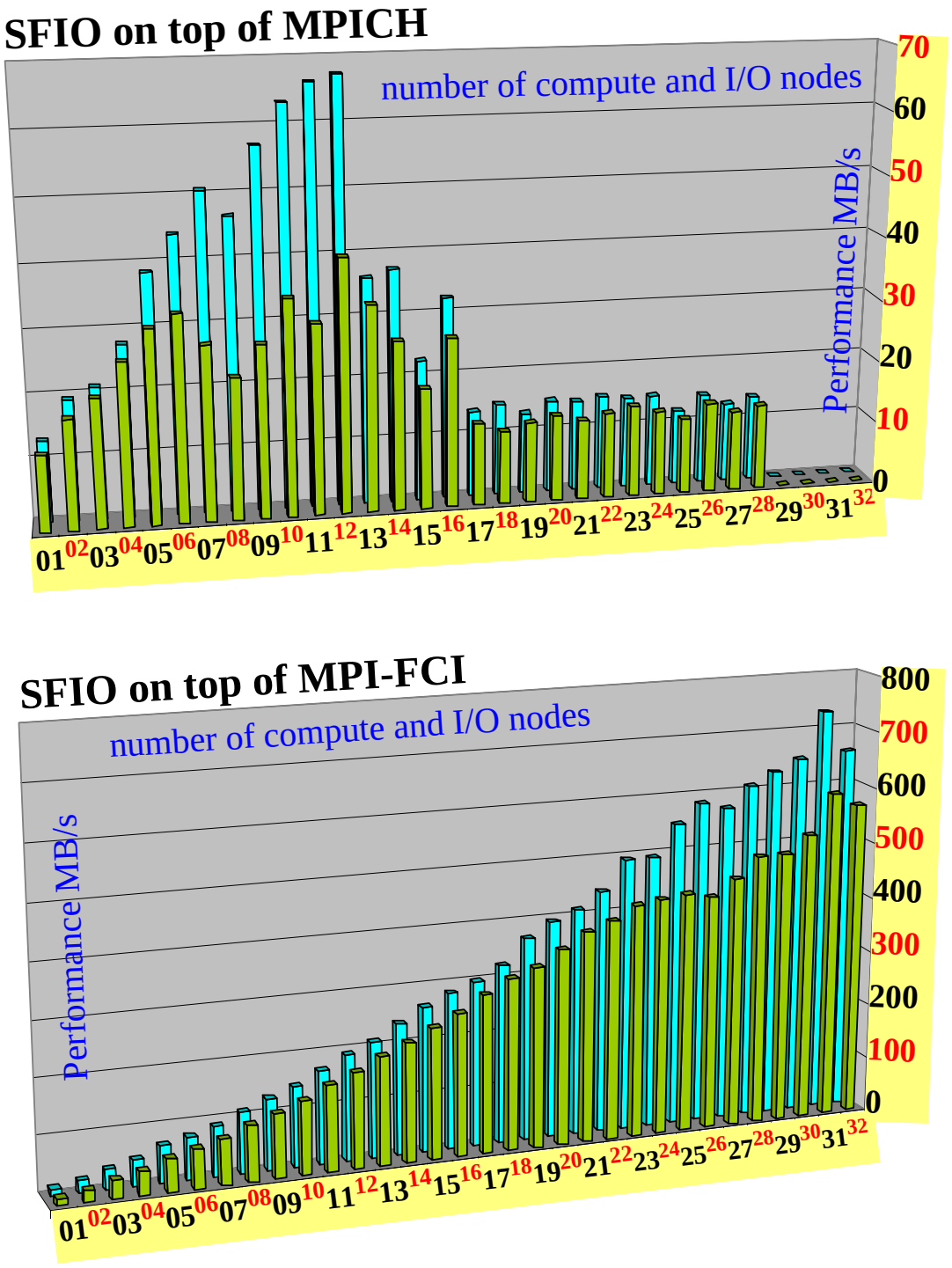
Fig. 4



- SFIO All-to-All concurent write access from all compute nodes to all I/O nodes
- Global File size is 2000MByte for MPICH and MPIFCI
- Stripe unit size is 200Byte only

Fig. 5

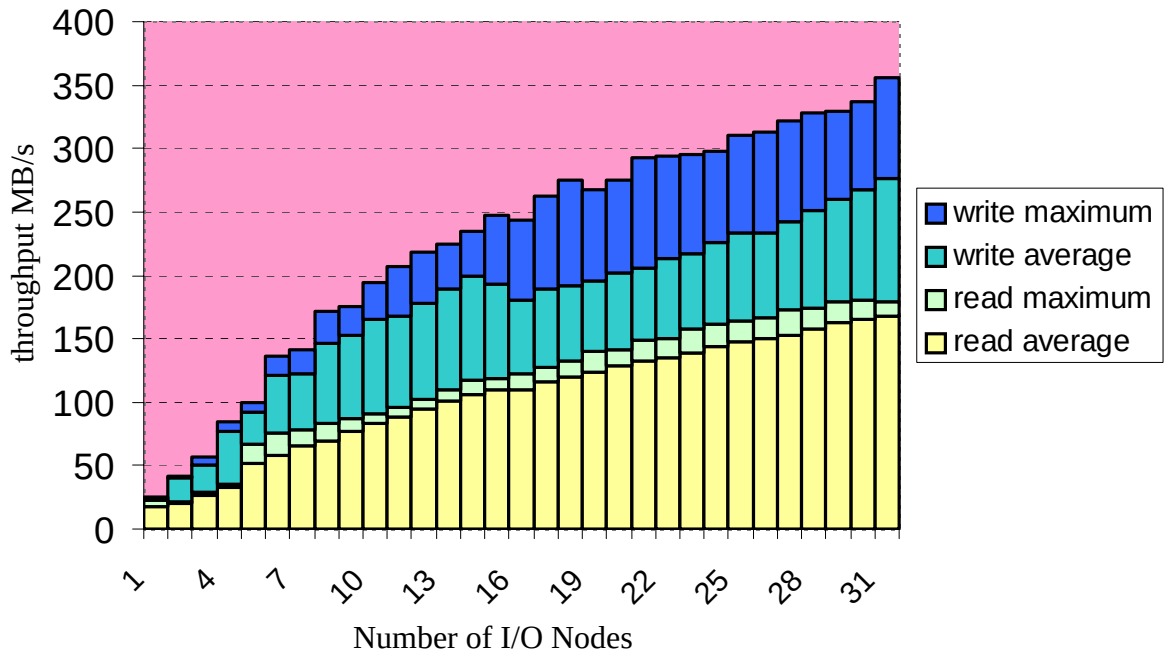
SFIO all-to-all I/O performance on Swiss-T1's Fast Ethernet and Tnet



Superlinear speedup of SFIO/FCI due to augmentation of cache effect when increasing the number of I/O nodes.

Fig. 6

SFIO All-to-all performance on T1.
(1GB-31GB file size, 200Byte chunk, 53 measurements)



To avoid the cache effect the total size of SFIO files is increasing when the number of I/O nodes grows.

Fig. 7

Swiss-T1 SFIO over TNet topology

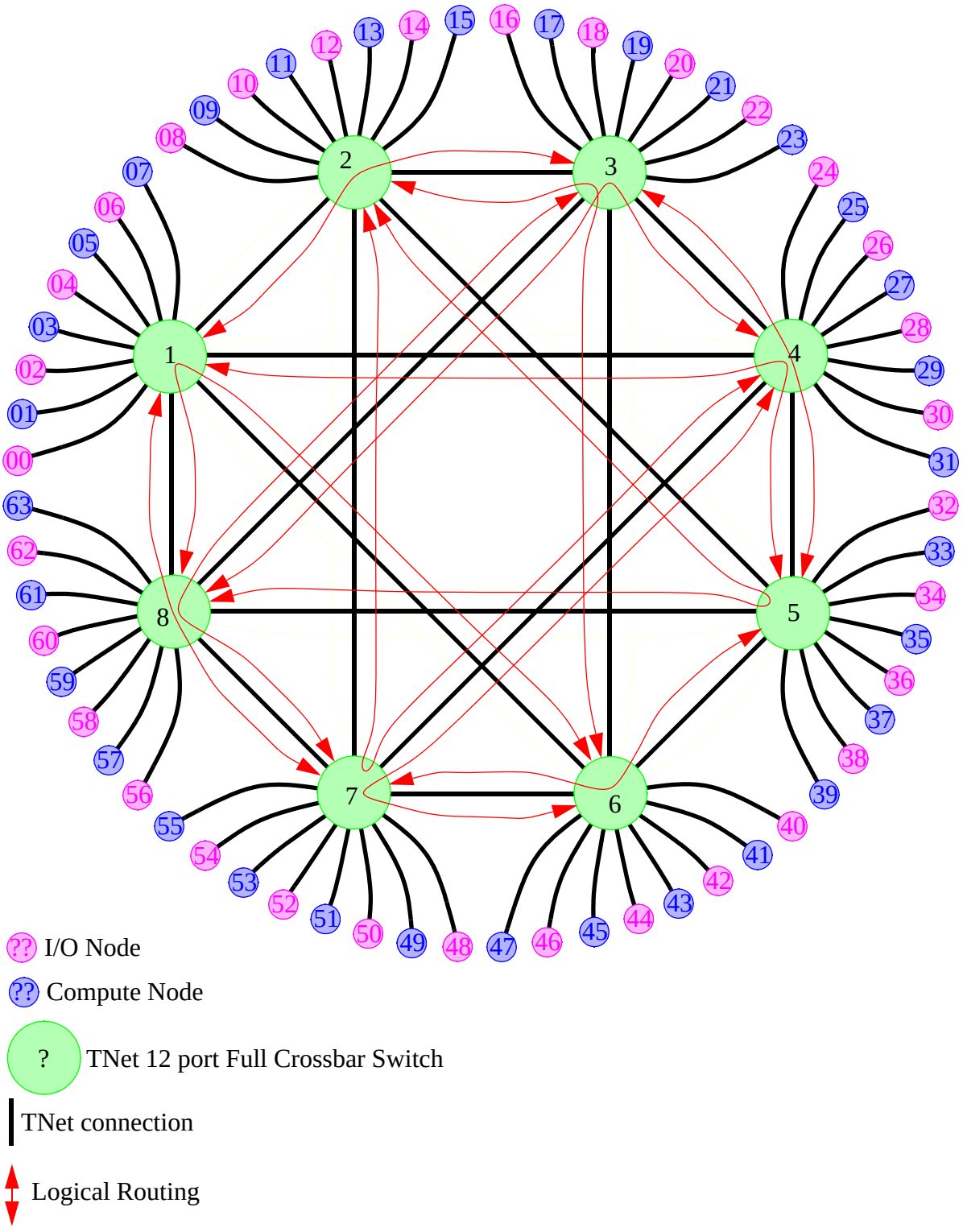


Fig. 8

Theoreticaly predicted Network All-to-All performances for 363 defferent allocation topologies on the Swiss-T1 machine

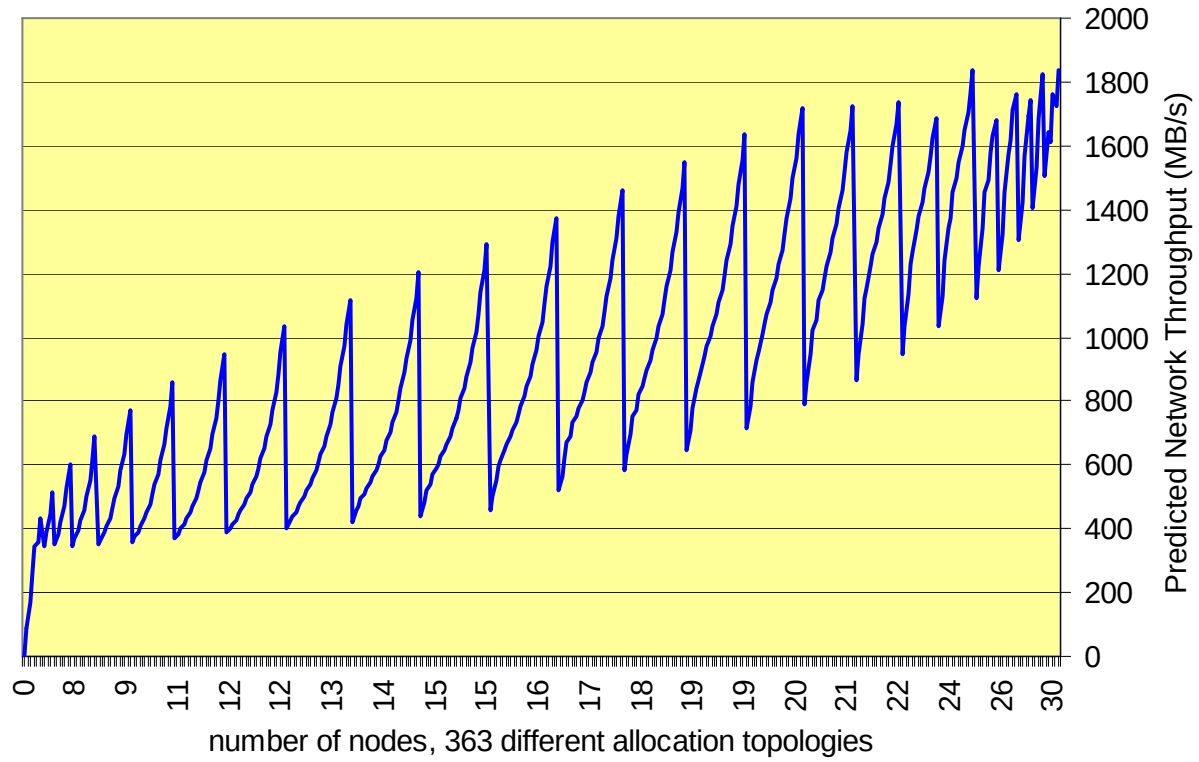


Fig. 9

Conclusion

- SFIO is portable, highly scalable, and ready for the distribution.

Future work

- SFIO performance benchmarking on the large supercomputer of Sandia National Laboratory (USA)
- Creation of a portable MPI-I/O interface library to SFIO
- Integration of techniques using network topolgy knowledge with the SFIO