UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

[simula . research laboratory]

Autors:
Roberto Peñaranda Cebrián
Ernst Gunnar Gran
Tor Skeie
María Engracia Gómez Requena
Pedro Juan López Rodríguez

A New Fault-Tolerant Routing Methodology for KNS Topologies

Universitat Politècnica de València

March 2016

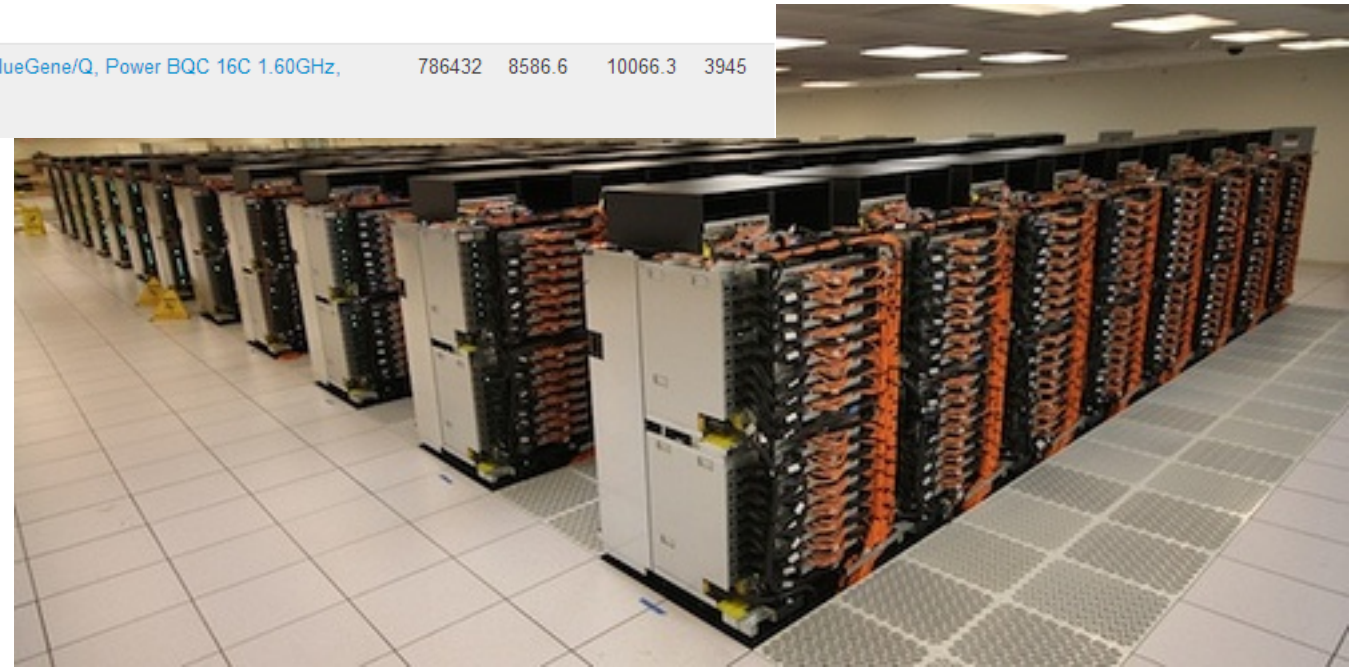
Index

Index

1. Introduction

- The size of large supercomputers has been growing.

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	National University of Defense Technology China	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3120000	33862.7	54902.4	17808
2	DOE/SC/Oak Ridge National Laboratory United States	**Titan** - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560640	17590.0	27112.5	8209
3	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1572864	17173.2	20132.7	7890
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705024	10510.0	11280.4	12660
5	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786432	8586.6	10066.3	3945



1. Introduction

-The high number of elements in an interconnection network heavily impacts the probability of having a failure in the system.

-To solve this:

1. To replicate all network elements, using them as spare components.
2. To modify the routing algorithm to be able to reach the destination nodes.

Index

2. Related Work

- Fault-tolerant techniques based on routing configuration:
  - Network reconfiguration.
  - Fault-tolerant routing algorithms.

2. Related Work

Fault-tolerant Routing algorithms

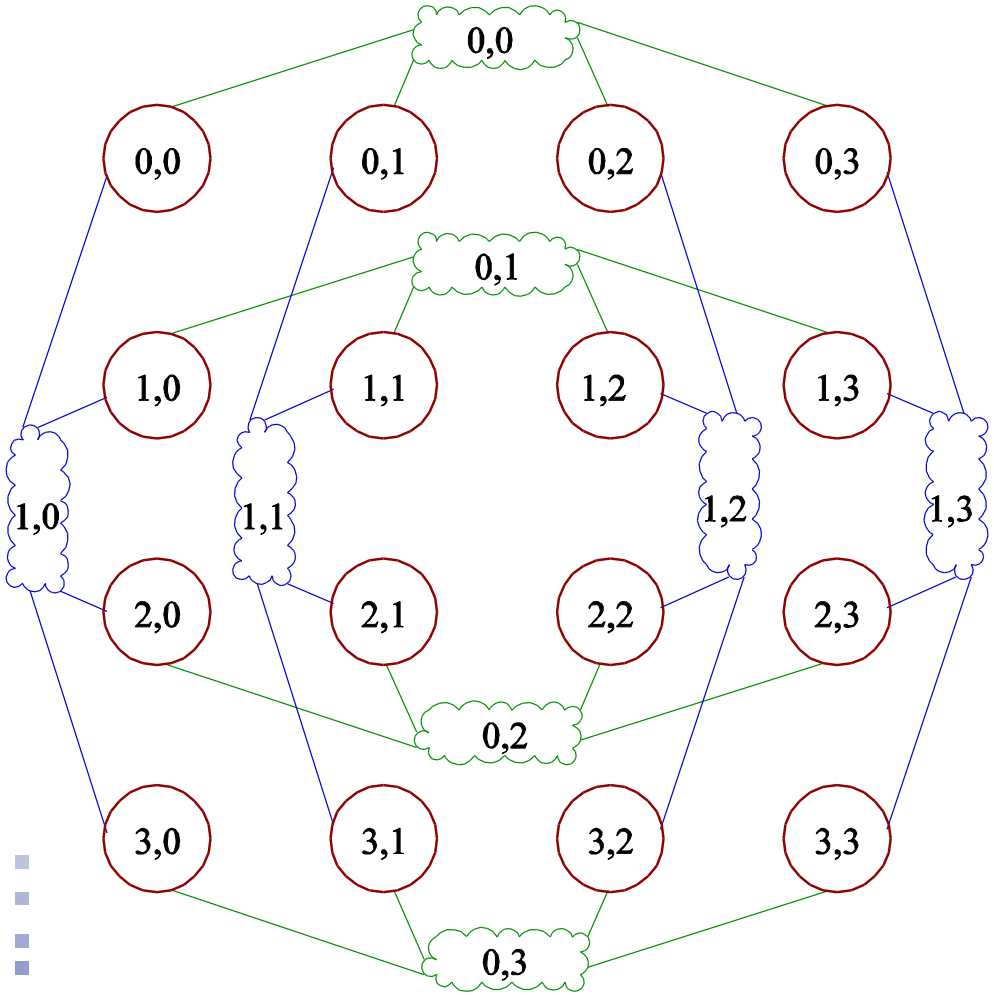
Different solutions for direct topologies:

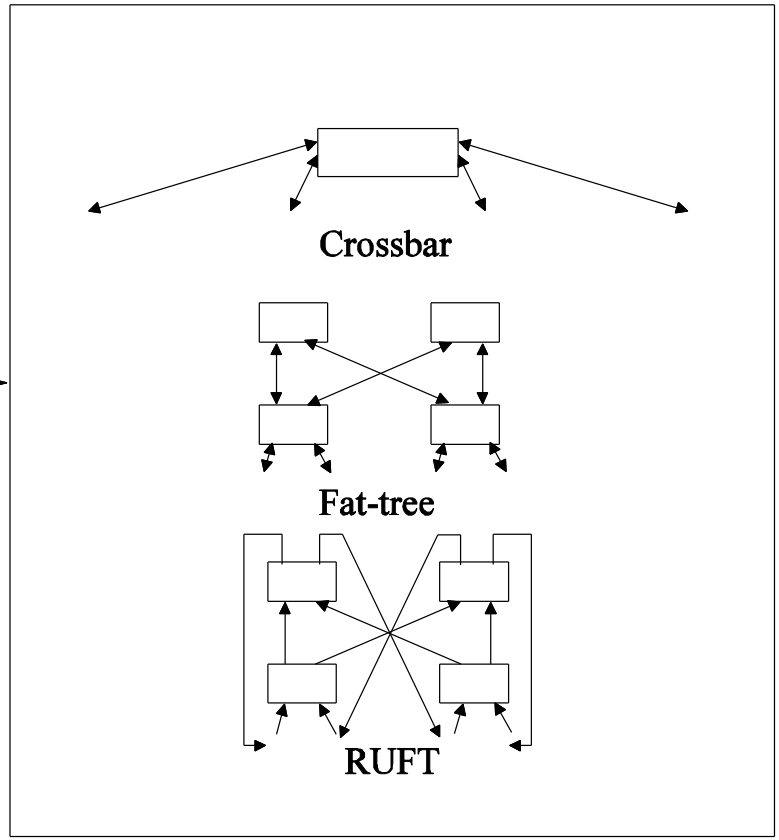
- Requiring extra resources (virtual channels) depending on the number of tolerated faults or the number of dimensions of the topology.
- Disabling fault regions (with healthy nodes).
- Using Valiant routing, avoiding faults through the intermediate nodes.
- Without requiring extra resources (bad traffic balance).

Index

3. Preliminaries

KNS Topology





3. Preliminaries

KNS Topology provides:

- High performance (low latency and high throughput).
- Reduced hardware cost and easy implementation.

Index

4. Fault-Tolerant Routing Methodology

New routing algorithm:

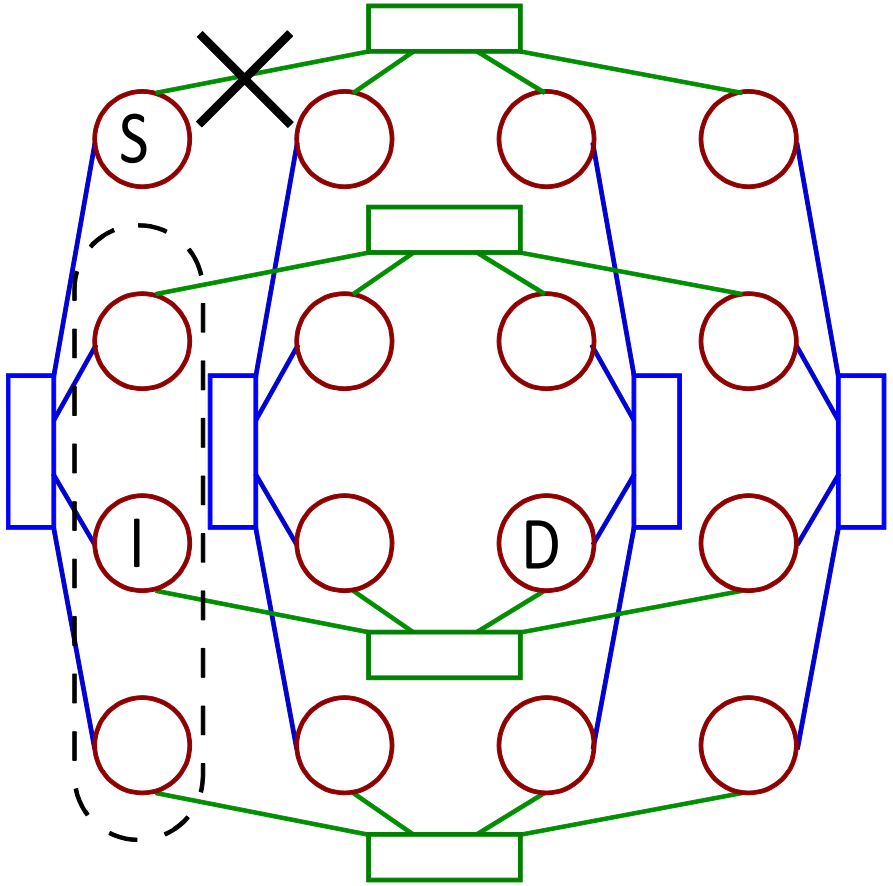
- For each source-destination pair without failures, we use Hybrid-DOR with minimal paths.
- If there is any fault, we use intermediate nodes (Valiant) using Hybrid-DOR in all subpaths .
- We need one extra virtual channel per each intermediate node to break deadlocks.

4. Fault-Tolerant Routing Methodology

One Intermediate node

The intermediate node has to satisfy:

- ***I*** is reachable from ***S*** using Hybrid-DOR.
- ***D*** is reachable from ***I*** using Hybrid-DOR.



4. Fault-Tolerant Routing Methodology

One Intermediate node

How many faults can it tolerate?

(Considering only faults that doesn't physically disconnect the network).

4. Fault-Tolerant Routing Methodology

One Intermediate node

How many faults can it tolerate?

(Considering only faults that doesn't physically disconnect the network).

Being:

T_{RS} the set of nodes that are reachable from ***S*** using Hybrid-DOR.

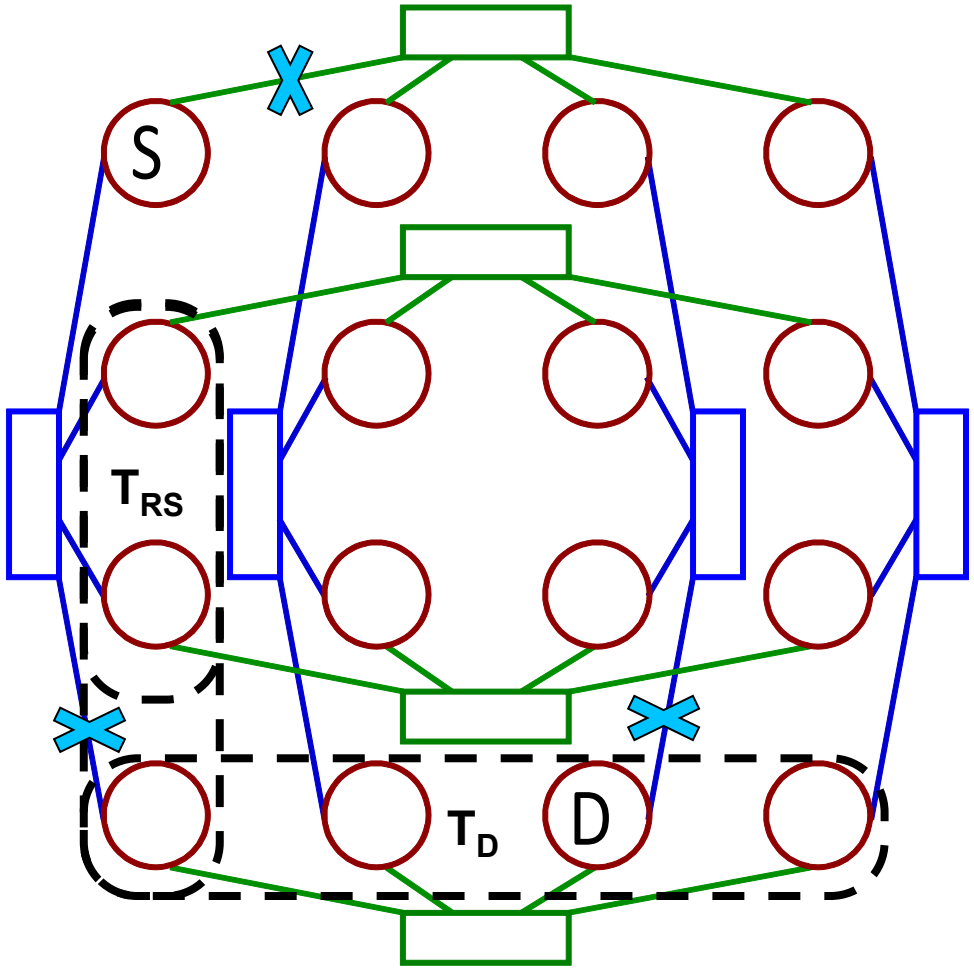
T_D the set of nodes from which ***D*** is reachable using Hybrid-DOR.

4. Fault-Tolerant Routing Methodology

One Intermediate node

How many faults can it tolerate?

(Considering only faults that doesn't physically disconnect the network).

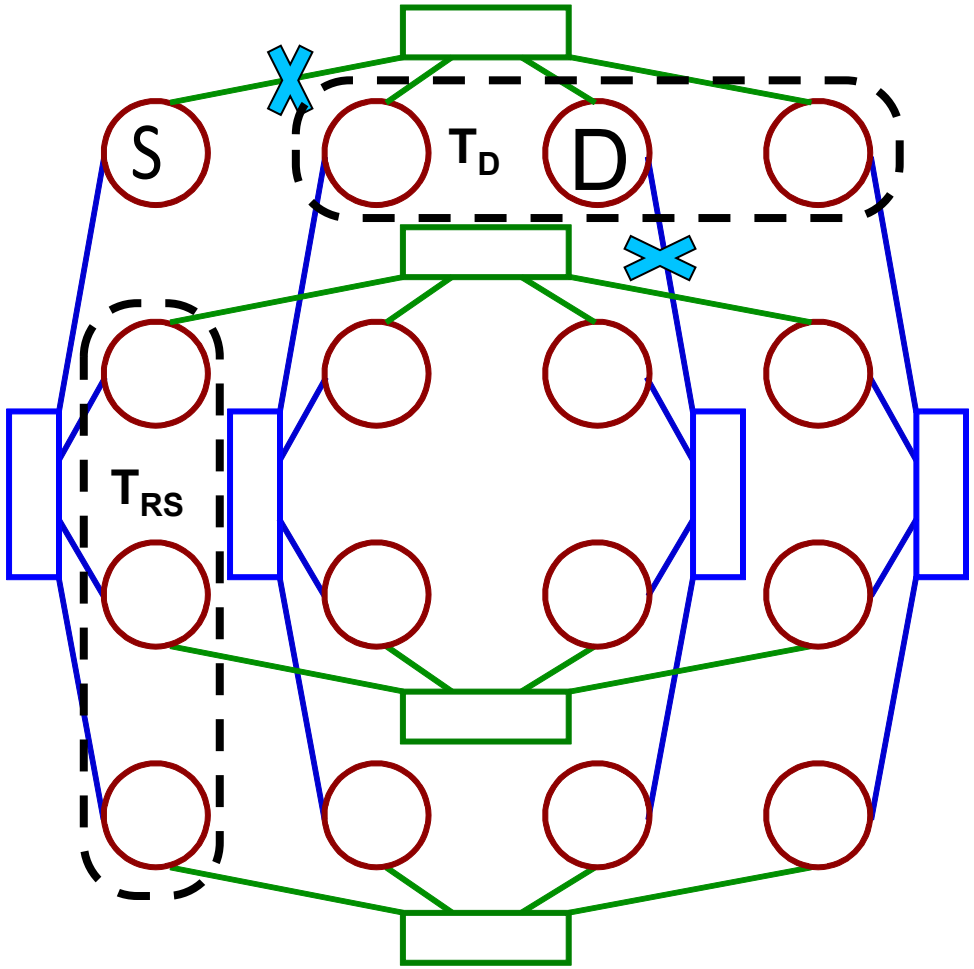


4. Fault-Tolerant Routing Methodology

One Intermediate node

How many faults can it tolerate?

(Considering only faults that doesn't physically disconnect the network).



4. Fault-Tolerant Routing Methodology

One Intermediate node

How many faults can it tolerate?
(Considering only faults that doesn't physically disconnect the network).

In general, for *n*-dimensional KNS networks, the routing algorithm with only one intermediate node is able to tolerate *n-1* failures.

4. Fault-Tolerant Routing Methodology

Multiple Intermediate nodes

In this case, assuming that we have x intermediate nodes ($I_1, I_2, \ldots I_x$), the intermediate nodes have to satisfy:

- $\boldsymbol{I_1}$ is reachable from $\boldsymbol{S}$.
- $\boldsymbol{I_{i+1}}$ is reachable from $\boldsymbol{I_i}$, for $0 < i < x$.
- $\boldsymbol{D}$ is reachable from $\boldsymbol{I_x}$.

4. Fault-Tolerant Routing Methodology

Multiple Intermediate nodes

In this case, assuming that we have x intermediate nodes ($I_1, I_2, \ldots I_x$), the intermediate nodes have to satisfy:

- $\boldsymbol{I_1}$ is reachable from $\boldsymbol{S}$.
- $\boldsymbol{I_{i+1}}$ is reachable from $\boldsymbol{I_i}$, for $0 < i < x$.
- $\boldsymbol{D}$ is reachable from $\boldsymbol{I_x}$.

For instance, for 2 intermediate nodes, $2 \times (n-1) + k - 3$ faults are tolerated for 3 or more dimensions, and $2 \times k - 1$ for 2 dimensions.

4. Fault-Tolerant Routing Methodology

Other KNS configurations

This methodology can be extended to KNS configurations that use other indirect subnetworks like, for example, fat-trees. To do this:

- The intermediate nodes are used globally.
- Each subnetwork has its local methodology.

Direct routers will work normally as long as the subnets can avoid faults.

Index

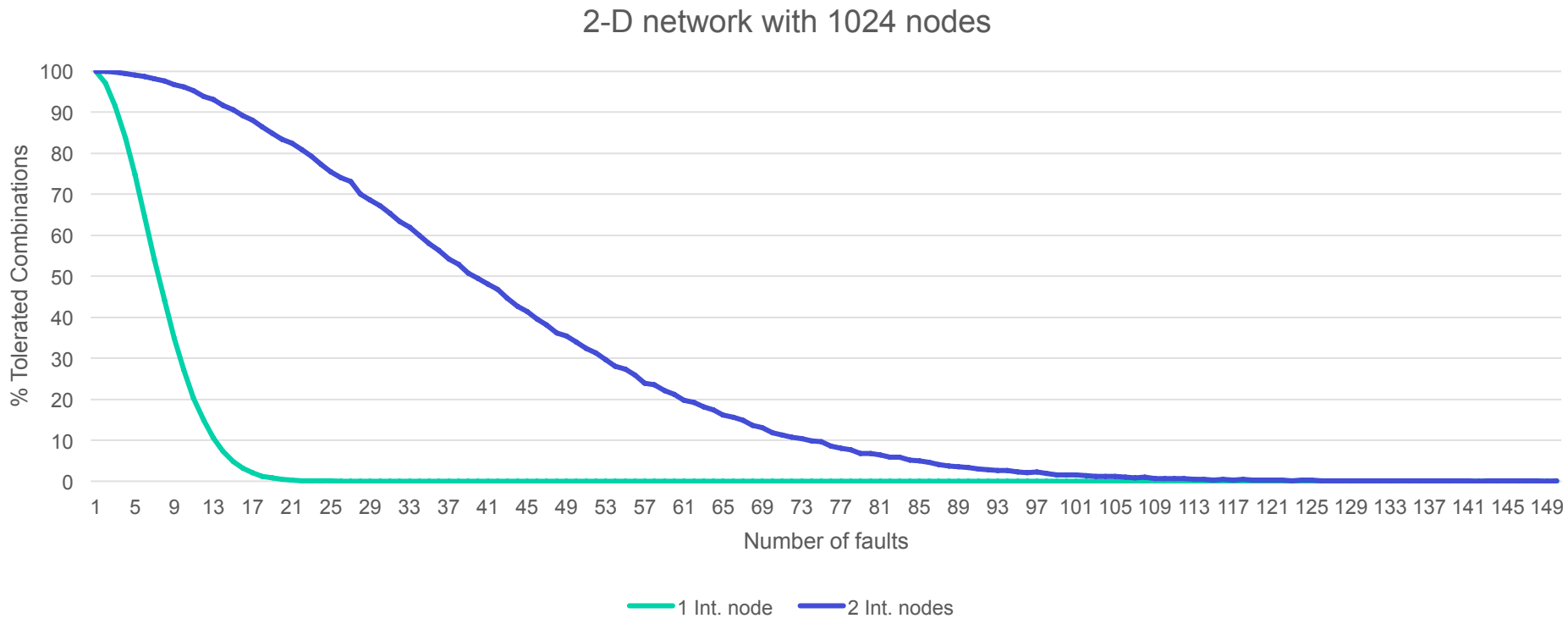
4. Experimental Evaluation

Divided into two parts:

- We analyze the number of failures which can be tolerated.
- We evaluate the performance of the methodology with different number of faults.

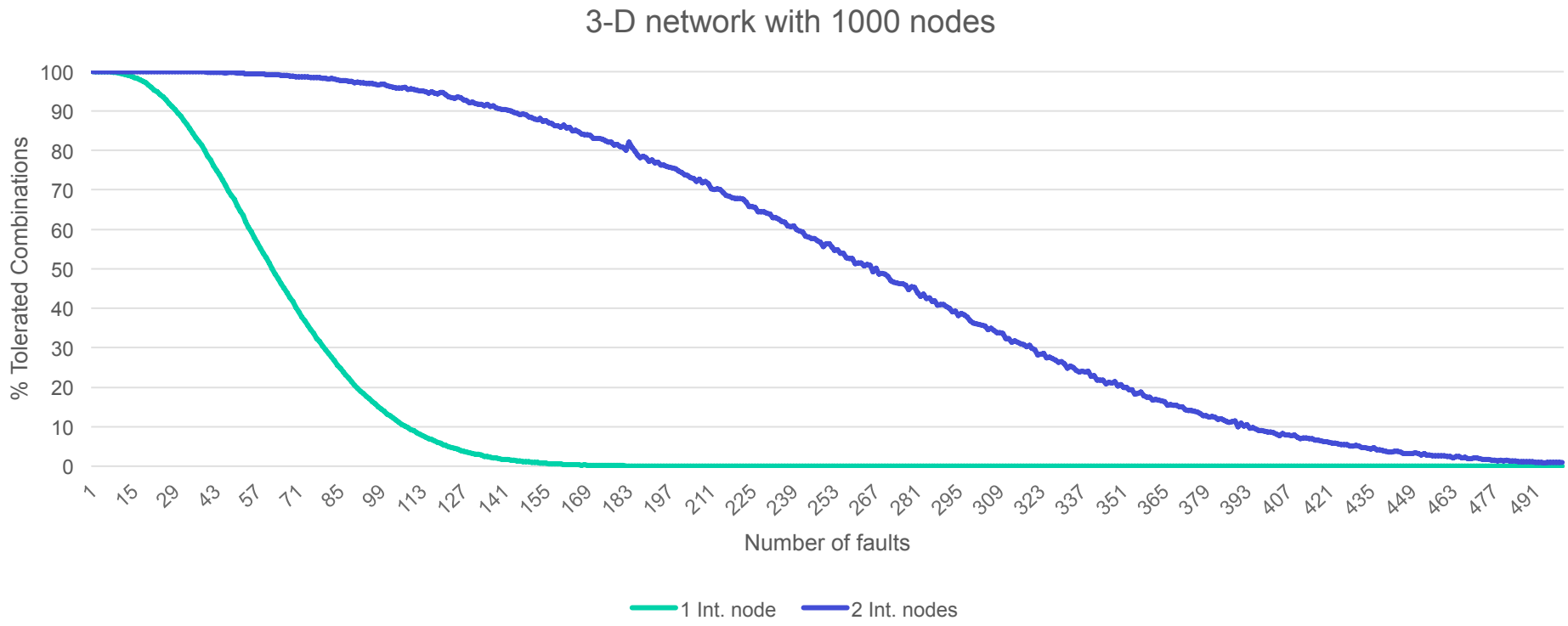
4. Experimental Evaluation

Fault analysis



4. Experimental Evaluation

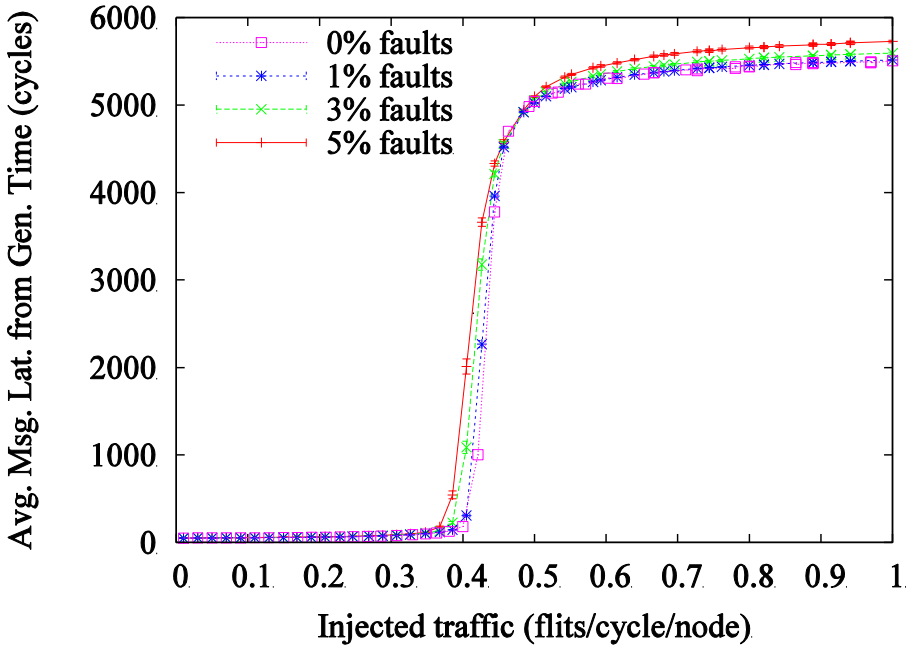
Fault analysis

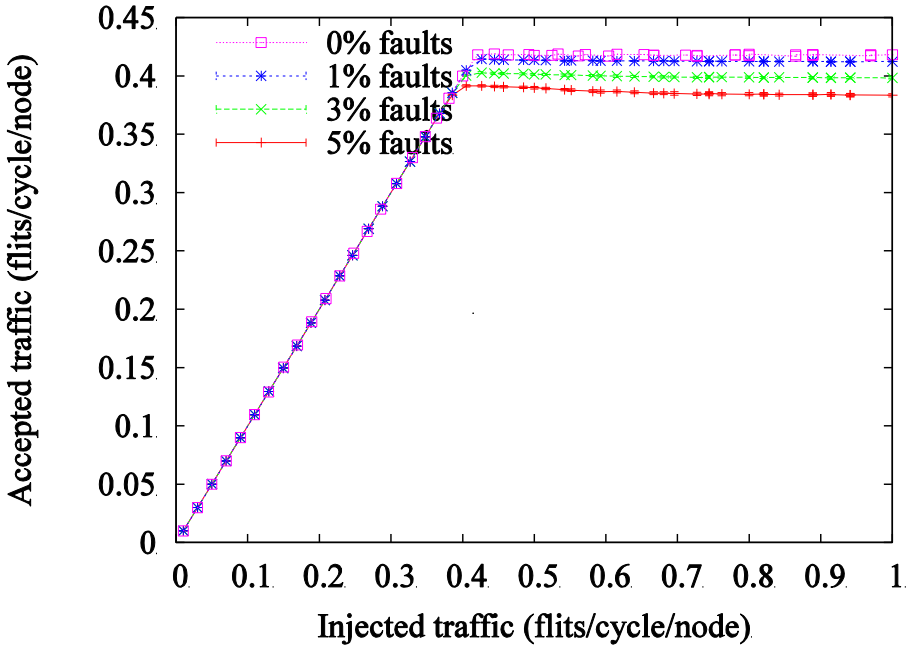


4. Experimental Evaluation

Performance analysis

2-D network with 1024 nodes

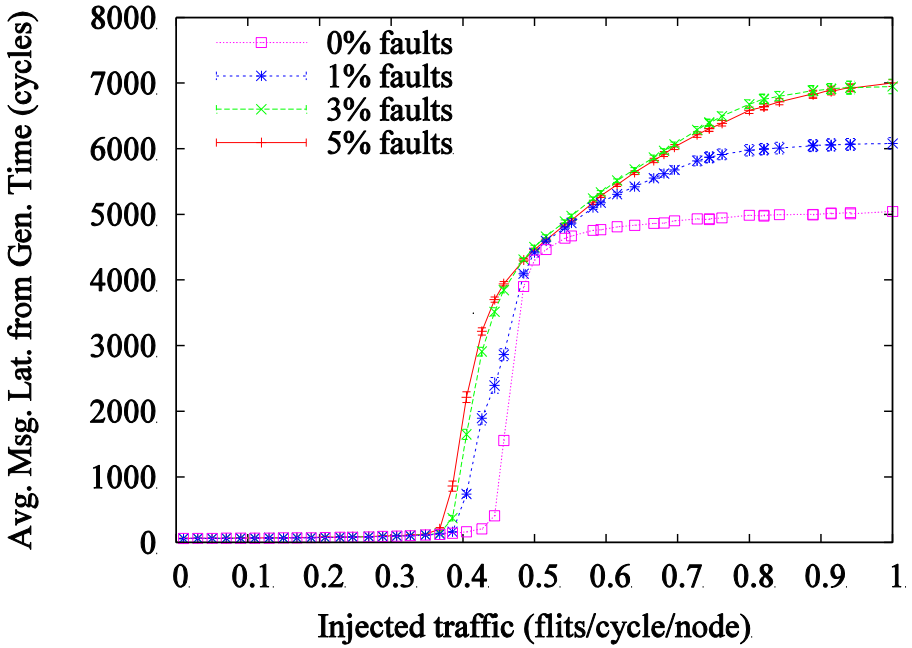


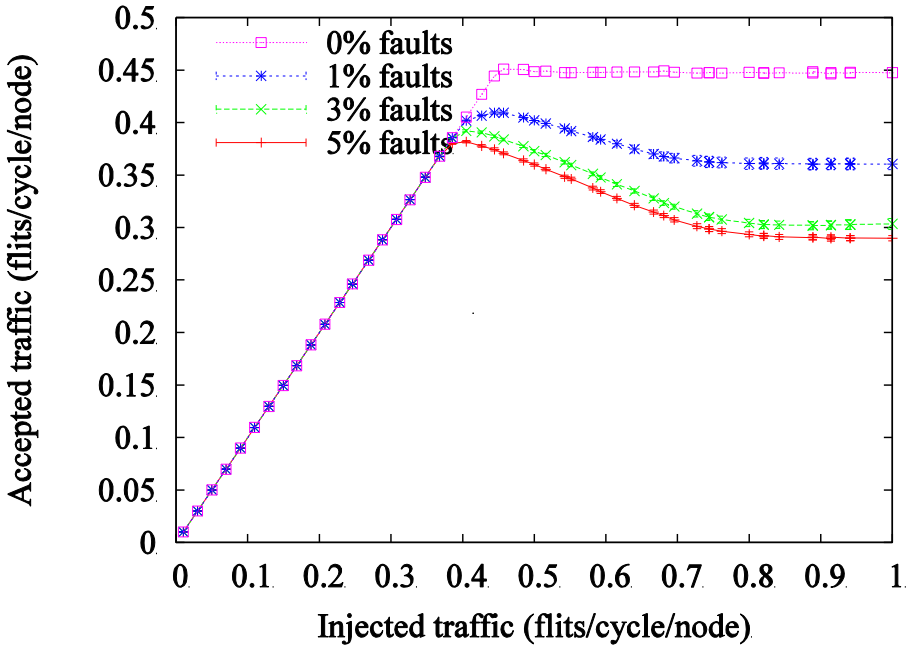


4. Experimental Evaluation

Performance analysis

3-D network with 1000 nodes





Index

5. Conclusions

- We have proposed a new fault-tolerant routing algorithm for KNS topologies.
- It can tolerate a large number of faults without disabling any healthy node and without suffering a great fall in performance, needing only one extra virtual channel per intermediate node.
- For instance, the results show a degradation in performance of 1% for a 2D-network with 1024 nodes and 1% faulty links.

Thank you!

ropeaceb@gap.upv.es