UNA NUEVA FAMILIA DE TOPOLOGÍAS INDIRECTAS, EFICIENTES Y TOLERANTES A FALLOS

-ABSTRACT-

Large parallel computers are currently adopting the cluster architecture as the basis for their construction. These machines are being built with thousands of processing nodes that are interconnected through high-speed interconnection networks.

Performance, fault tolerance and network cost are key factors in the overall design of these systems. The levels of computing power required can only be reached by increasing the number of network nodes. As systems grow, however, so does the amount of network components and with it, the probability of network faults. Since availability is important with these computer systems, fault-tolerance mechanisms are often implemented that are based on increasing network size and duplicating components, which directly affects cost.

In the field of interconnection networks, indirect topologies are often the design of choice for HPC systems. The most commonly used indirect topology is the fat-tree, which is a multi-stage bidirectional-link topology providing good network performance and high fault-tolerance levels, but at a high cost. To reduce cost, RUFT has been proposed, a multi-stage unidirectional topology providing network performance similar to that of the fat-tree but using fewer hardware resources (approximately half). RUFTs weak point, however, is that it has zero fault tolerance.

This work focuses on designing a simple indirect topology that offers high performance and fault tolerance while keeping hardware cost as low as possible. In particular, we propose a set of new topologies with different properties in terms of cost, performance and fault tolerance. All of them are capable of achieving performances similar to or better than that of the fat-tree, while also providing good fault tolerance levels and tolerating faults in the links connecting to end nodes, which most available topologies cannot do.

Our first contribution is RUFT-PL, a topology that duplicates the number of injection, network and ejection links, while using the RUFT connection pattern to interconnect all network elements. This topology provides high network performance and a slight level of fault tolerance, using the same hardware resources as a fat-tree.

Our second contribution is the FT-RUFT-212 topology, which provides better network performance than the fat-tree, as well as good fault tolerance for a low design cost, thanks to the proposed injection/ejection system implemented by the processing nodes.

The third contribution, FT-RUFT-222, is a topology combining the best properties of the previous two proposals. In particular, this topology implements the injection/ejection used by FT-RUFT-212 and the double network links used by RUFT-PL to interconnect the switches. It provides high performance and fault-tolerance levels while using the same hardware resources required by a fat-tree.

Our fourth and last contribution is FT-RUFT-XL, a topology in which both the injection/ejection and the connection between the switches have been redesigned. It offers a significant improvement on the other proposals' fault-tolerance levels, and also provides high network performance. Furthermore, unlike many unidirectional topologies, it allows

packets to take different routes at every network stage, always bringing them closer to their destination with each hop.