Exploring Randomized Multipath Routing On 5D Torus Networks

Prajakt Shastry[1], Daniel Parker[2], Sanjiv Kapoor[1], Ioan Raicu[1,3]

[1] Illinois Institute of Technology, [2] University of Chicago, [3] Argonne National Laboratory

Abstract

Network performance is a critical aspect of high-performance computers, and improving its performance is a major goal in the design of future systems; this work proposes to improve network performance through new routing algorithms, leveraging the rich multi-path topologies of multi-dimensional torus networks commonly found in supercomputers built in the past fifteen years. Virtually all torus networks in production today utilize the dimension order routing algorithm, which is essentially a static and deterministic routing strategy to allow internode communication. We propose a new Random Distance Routing algorithm, which randomly distributes packets to different neighboring nodes that are closer to the destination, leading to a global load balanced network. Through the CODES/ROSS simulator, we show that the proposed randomized multi-path routing algorithm can increases throughput of a 5D-Torus network by 1.6X, as well as reduce latency by 40%.

Motivation

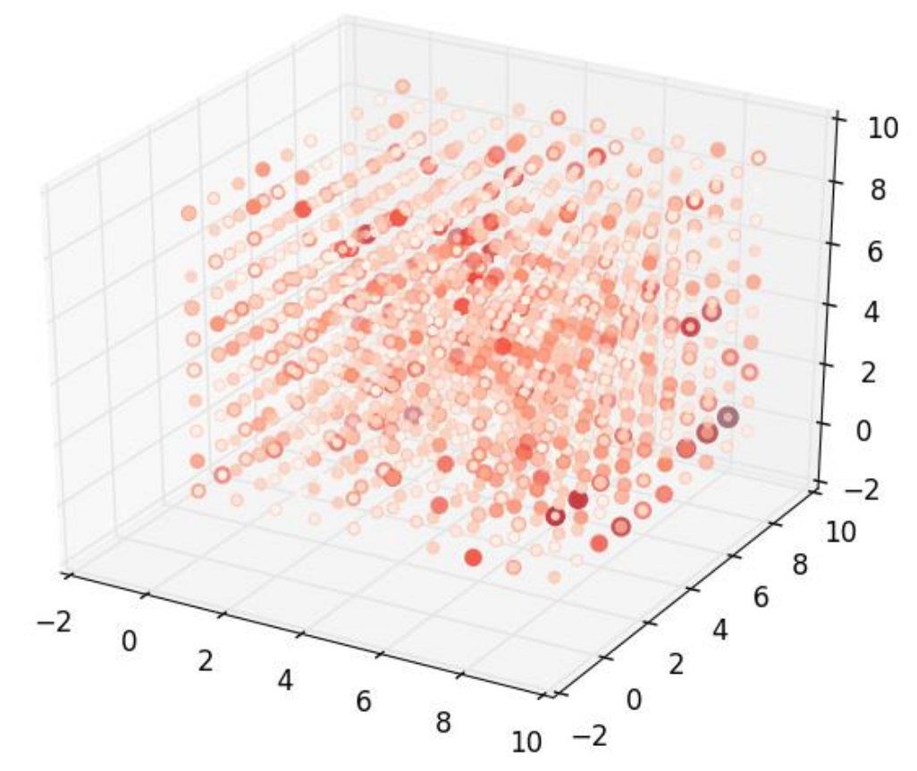
- Provide increase in performance without overhead.
- Provide better scalability than Dimensional Order Routing, and other existing algorithms, by load balancing traffic, using Randomization

RDR

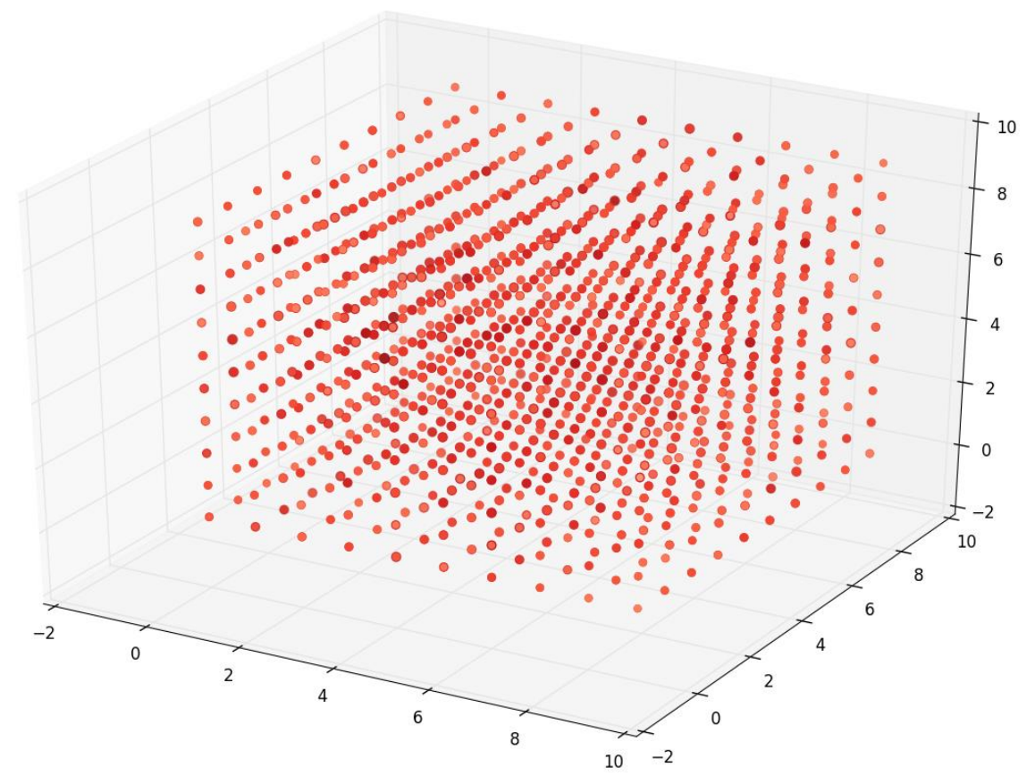
```
RDR(V):
1. If V = destination: - Stop
2. Mark neighboring node W as viable if distance from W to destination is less.
3. Randomly select one of viable options. Send packet to selected option, to process using RDR (W)
```

Limitations Of some current methods – Load Balance and Locality

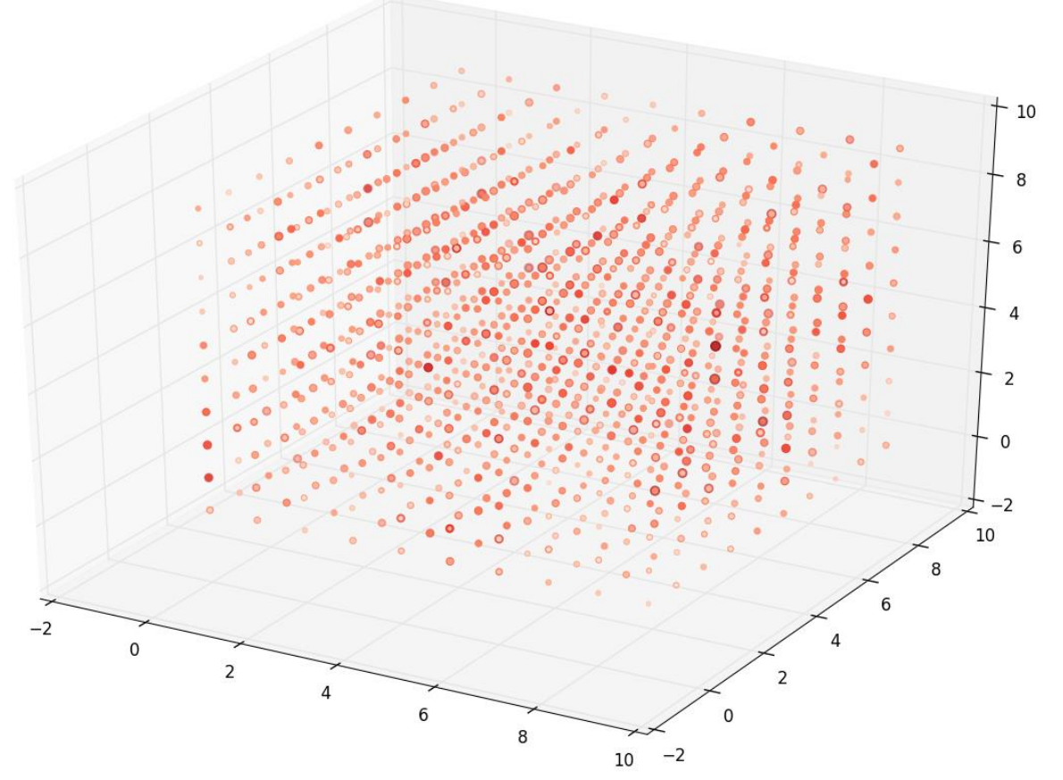
- **Dimensional Order Routing(DOR)**
  - Good average case performance doesn't load balance well.



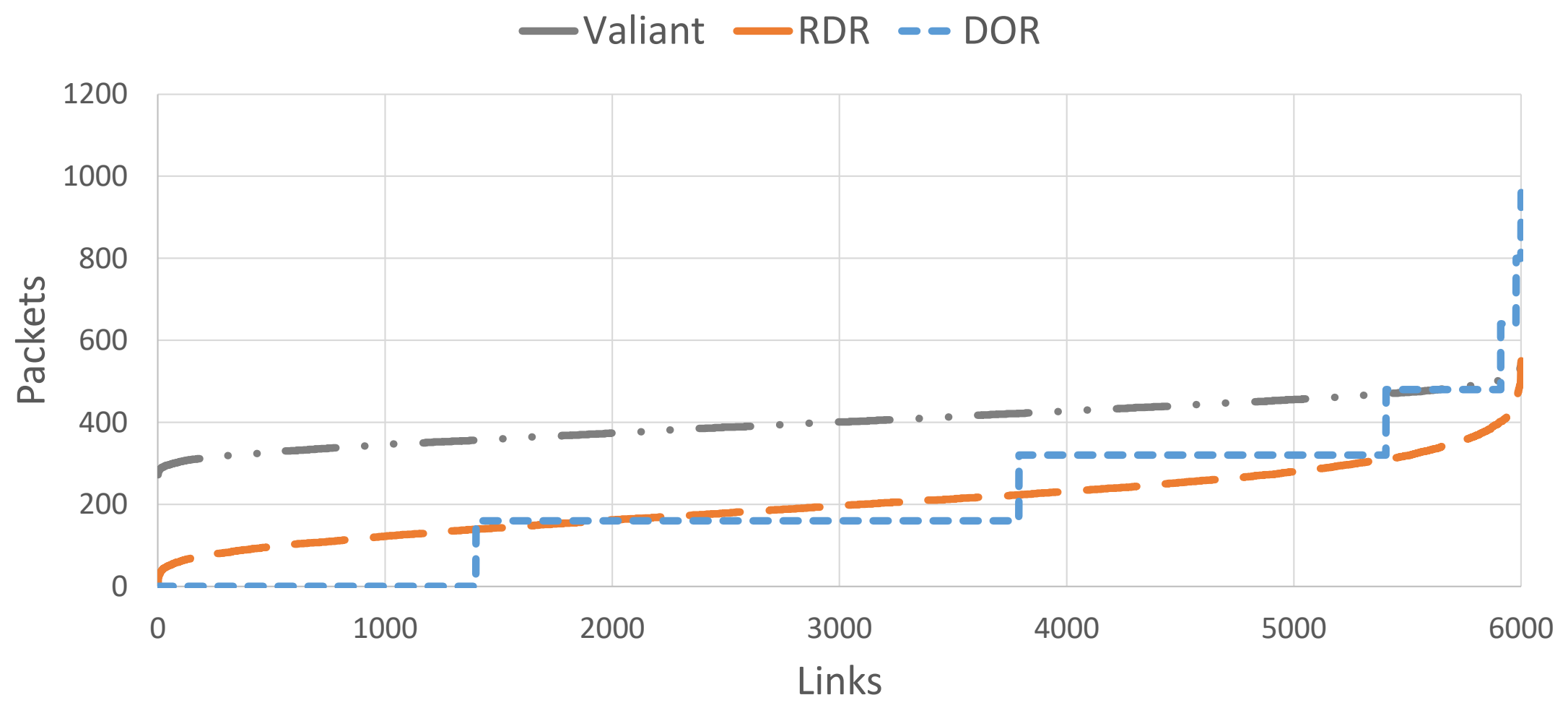
- **Valiant**
  - Good load balancing, but loses locality



- **Random Order Routing(RDR)**
  - Balances load without loosing locality



- **Link Utilization stats**

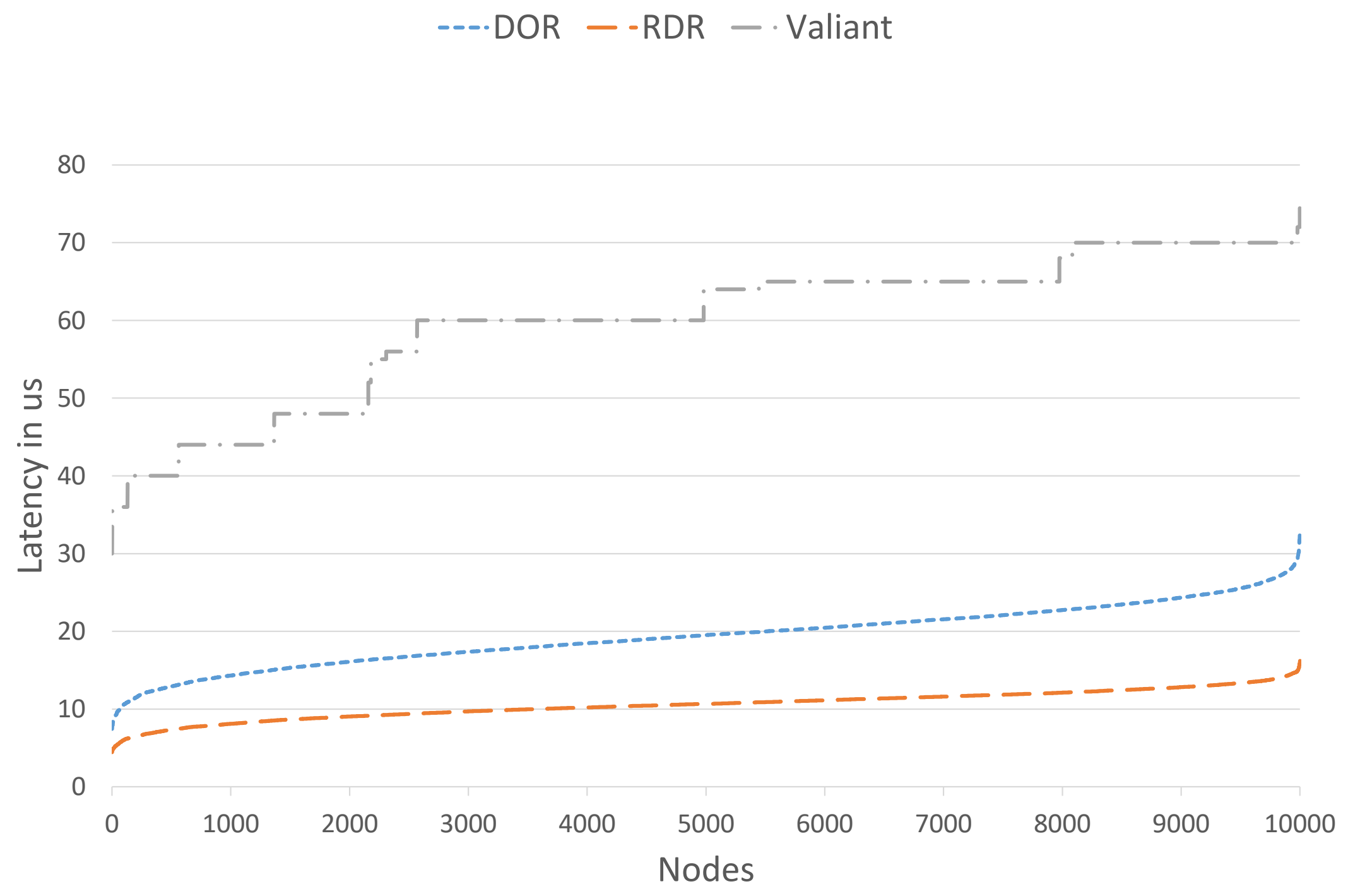


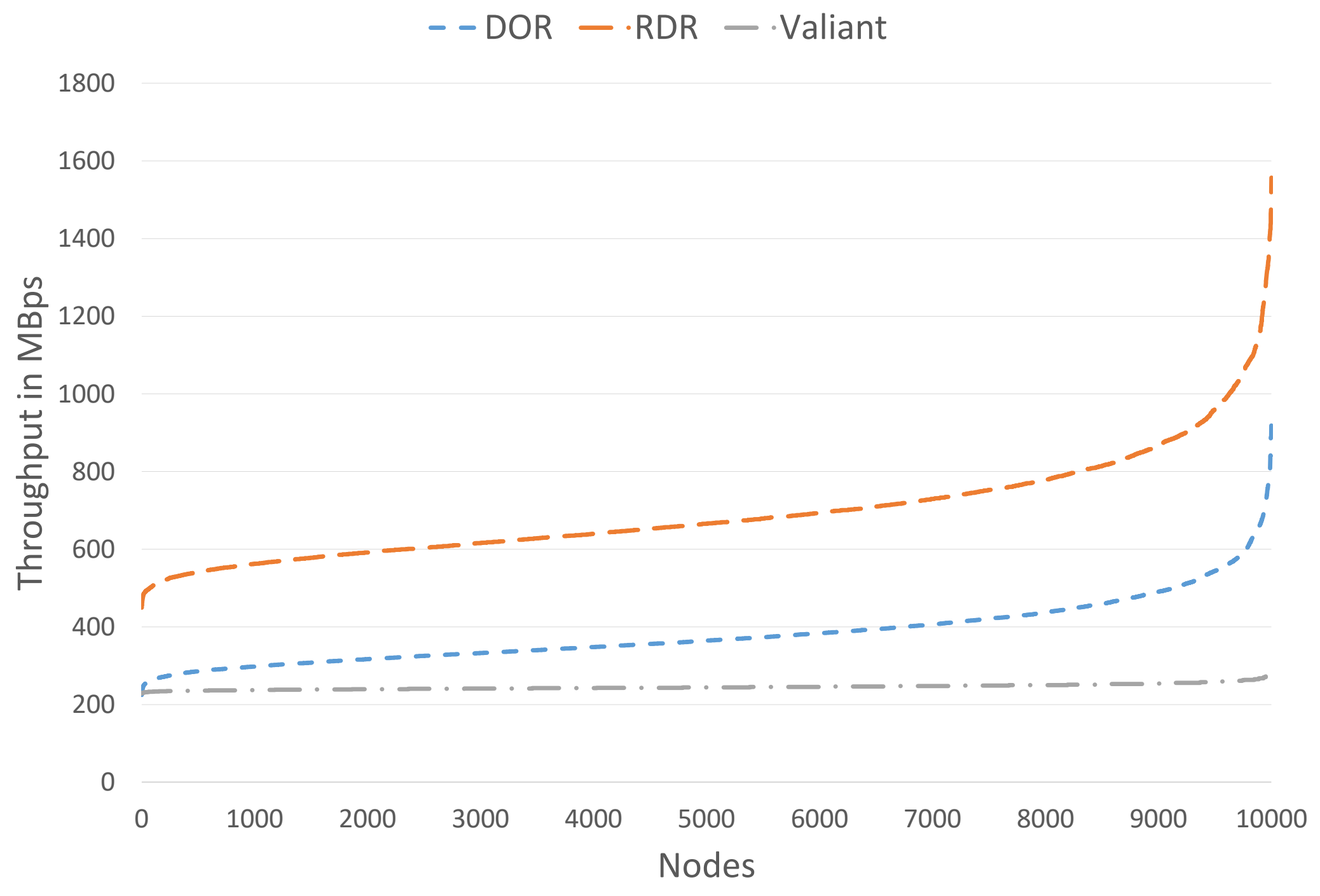
Results

- **Experiment Setup**

- 20000 nodes - Network setup:10x10x10x10x2 - Message Size:8 KB

- Packet Size: 512 Bytes – Link Bandwidth : 2 GB





Contribution

- A simplistic randomized multipath algorithm for multidimensional Torus Networks
- Comparison between RDR, DOR and Valiant routing

Conclusion and Future Work

RDR increases performance of the network by randomly distributing data, and achieves a 1.6X improvement in average case throughput. RDR reduces latency by 50% when compared with DOR in average case. RDR scales well with increasing number of nodes in the system. GOAL [8] seems to reduce latency when compared with DOR by 40% in worst case. RDR will perform much better since it achieves 50% in average case. Extensive comparison between GOAL and RDR will be done. We aim to improve on RDR by using network state for making routing decisions.

References

[1] https://asc.llnl.gov/computing_resources/bluegenel/images/torus.jpg

[2] J. Breckling, Ed**., Looking Under the Hood of the IBM Blue Gene/Q Network**, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] L. G. Valiant. "**A scheme for fast parallel communication**." SIAM Journal on Computing, 11(2):350–361, 1982.

[4] Cope, Jason, Ning Liu, Sam Lang, Phil Carns, Chris Carothers, and Robert Ross. "**Codes: Enabling co-design of multilayer exascale storage architectures**." In Proceedings of the Workshop on Emerging Supercomputing Technologies, pp. 303-312. 2011.

[5] T. Nesson and S. L. Johnsson. **ROMM routing on mesh and torus networks.** In Proc. 7th Annual ACM Symposium on Parallel Algorithms and Architectures SPAA'95, pages 275– 287, Santa Barbara, California, 1995.

[6] A. Singh,W. J. Dally, B. Towles, and A. K. Gupta. **Locality preserving randomized routing on torus networks**. In *Proc.* 12th Annual ACM Symposium on Parallel Algorithms and *Architectures SPAA'02*, Winnipeg, Canada, 2002.

[7] Bolding, M. L. Fulgham, and L. Snyder. **The case for chaotic adaptive routing.** IEEE Transactions on Computers, 46(12):1281–1291, 1997

[8] Arjun Singh and William J Dally and Amit K Gupta and Brian Towles. GOAL: A load-balanced adaptive routing algorithm for torus networks. International Symposium on Computer Architecture (ISCA) ACM2003