ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

A Survey on Overcoming Far-end Congestion in Large-Scale Networks

Shalini Vyas, Richa Chauhan

M.Tech Scholar, Dept. of ECE, Oriental Institute of Science and Technology, Bhopal, M.P, India

Professor, Dept. of ECE, Oriental Institute of Science and Technology, Bhopal, M.P, India

ABSTRACT: Accurately estimating congestion for correct global adaptive routing choices (i.e., verify whether or not a packet should be routed minimally or non-minimally) includes a significant impact on overall performance for high-radix topologies, such as the dragonfly topology. Previous work have targeted on understanding near-end congestion – i.e., congestion that occurs at the current router – or downstream congestion – i.e., congestion that occurs in downstream routers. However, most previous work does not evaluate the impact of far-end congestion or the congestion from the high channel latency between the routers. In this work, we have a tendency to refer to far-end congestion as phantom congestion as the congestion isn't "real" congestion. Due to the long inter-router latency, the in-flight packets (and credits) result in inaccurate congestion info and may cause inaccurate adaptive routing choices. In addition, we have a tendency to show how transient congestion happens as the occupancy of network queues fluctuate due to random traffic variation, even in steady-state conditions. This also results in inaccurate adaptive routing decisions that degrade network performance with lower throughput and better latency. To overcome these limitations, we propose a history window based approach to remove the impact of phantom congestion. We have a tendency to also show how using the average of local queue occupancies and adding an offset considerably remove the impact of transient congestion. Our evaluations of the adaptive routing in a large-scale darning needle network show that the combination of these techniques lead to an reconciling routing that just about matches the performance of a perfect adaptive routing algorithmic rule.

KEYWORDS : Overcoming far- end congestion, adaptive routing, large scale networks.

I. INTRODUCTION

Interconnection networks are a crucial element of contemporary computer systems. From large scale systems to multi core architectures [17, 33], the interconnection network that connects processors and memory modules considerably impacts the general performance and cost of the system. As processor and memory performance continues to extend, multicomputer interconnection networks are getting even more crucial as they mostly determine the bandwidth and latency of remote access. A good interconnection network is intended round the capabilities and constraints of accessible technology. Increasing material router pin bandwidth, for instance, has motivated the use of high-radix routers [15] in which the increased bandwidth is used to increase the number of ports per router, rather than maintaining a small number of ports and increasing the bandwidth per port. The Cray BlackWidow system, one of the first systems to use a high-radix network, uses a variant of the folded-Clos topology and radix-64 routers a significant departure from previous low-radix 3D torus networks. Recently, the advent of economical optical signaling enables topologies with long channels. However, these long optical channels are significantly more expensive than short electrical channels. in this paper, we have a tendency to introduce the dragonfly one topology that exploits emerging optical signaling technology by grouping routers to additional increase the effective radix of the network. The topology of an interconnection network for the most part determines both the performance and therefore the cost of the network [8]. Network cost is dominated by the cost of channels, and in particular the cost of the long, global, inter-cabinet channels. Thus, reducing the number of global channels can significantly reduce the value of the network. To reduce global channels without reducing performance, the quantity of global channels traversed by the typical packet must be reduced. The dragonfly topology introduced in this paper reduces the amount of global channels traversed per packet using minimal routing to one. To achieve this unity global diameter, terribly high-radix routers are required.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

II. LITERATURE SURVEY

Ted Jiang et. al. [1] "Overcoming Far-end Congestion in Large-Scale Networks" Accurately estimating congestion for correct global adaptive routing selections (i.e., verify whether or not a packet should be routed minimally or non-minimally) contains a important impact on overall performance for high-radix topologies, like the dragonfly topology. Previous work have targeted on understanding near-end congestion – i.e., congestion that happens at the present router or downstream congestion – i.e., congestion that happens in downstream routers. However, most previous work don't measure the impact of far-end congestion or the congestion from the high channel latency between the routers. in this work, we have a tendency to refer to far-end congestion as phantom congestion because the congestion isn't "real" congestion. we tend to determine the impact of far-end congestion that occurs in large-scale networks due to long latency between neighboring routers and therefore the completely different length channels within the topology. The congestion at the far-end of the channel isn't accurately represented at the near-end since in-flight packets (or credits) that are being transmitted don't represent true congestion. W. J. Dally et. al. [2] "Technology-Driven, Highly-Scalable Dragonfly Topology" Developing technology and increasing pin-bandwidth encourage the utilization of high-radix routers to reduce the diameter, latency, and expenditure of inter-connection system. High-radix networks, however, require longer cables than their low-radix corresponding item. Because cables control network expenditure, the number of cables, and particularly the amount of long, global cables must be decrease to realize a capable network. In this paper, we have a tendency to introduce the dragonfly topology that uses a group of high-radix routers as a virtual router to increase the effective radix of the network. The dragonfly topology which uses a group of routers as a virtual router to increase the effective radix of the network, and therefore decrease network diameter, cost, and latency. Since it reduces the number global cables in a network, while at an equivalent time increasing their length, the dragonfly topology is especially well matched for implementations by means of emerging dynamic optical cables—which have a high fixed price however a low cost for each unit length compared to electrical cables. Using dynamic optical cables for the global channels, a dragonfly network reduces price by two hundredth compared to a two-dimensional butterfly and by fifty two compared to a folded Clos network of a similar bandwidth. Baba Arimilli et. al. [3] "The PERCS High-Performance Interconnect" The PERCS arrangement was intended by IBM in response to a DARPA challenge that desirable a high-productivity high-performance divides arrangement. A major modernization in the PERCS design is that the network that's designed using Hub chips that are integrated into the calculate nodes. Each Hub chip is about 580 mm2 in size, has over 3700 signal I/Os, and is put together in a component that also include LGA-attached optical electronic procedure. The Hub module implements 5 styles of high-bandwidth interconnect with multiple links that are fully-connected with a high-performance internal crossbar switch. These links give over 9 Tbits/second of raw bandwidth and are used to construct a two-level direct-connect topology on both sides of up to tens of thousands of POWER7 chip by means of high bisection bandwidth and low latency. The Blue Waters System that is being made at NCSA is an example large-scale PERCS installation. Blue Waters is predicted to bring persistent Petascale performance over a wide range of applications. John Kim et. al. [4] "The BlackWidow High-Radix Clos Network" In this work describe the radix-64 fold over Clos arrangement of the Cray Black-Widow scalable vector multiprocessor. We describe the Black-Widow network which extent to 32K progression with a worst case diameter of seven hops, and therefore the underlying high-radix router microarchitecture and its implementation. By using a high-radix router with several narrow channels which may be capable to take benefit of the higher pin concreteness and faster signaling rates available in modern ASIC technology. The BlackWidow router is an 800 MHz ASIC with sixty four 18.75 GB/s bidirectional ports for an aggregate off chip bandwidth of 2.4Tb/s. every port consists of three 6.25 GB/s differential signals in each direction. The router supports deterministic and adaptive packet routing with separate buffering for request and reply virtual channels. YARC may be a high-radix router utilized in the network of the Cray black widow multiprocessor. using YARC routers, every with 64 3-bit wide ports, the BlackWidow level up to 32K procedure uses a folded-Clos topology with a worst-case diameter of seven hops. Every YARC router has an aggregate bandwidth of 2.4Tb/s and a 32K-processor BlackWidow system includes a division bandwidth of 2.5Tb/s. Paul Gratz et. al. [5] "Regional Congestion Awareness for Load Balance in Networks-on-Chip" Interconnection networks-on-chip (NOCs) are quickly replacement other types of interconnect in chip multiprocessors and arrangement on-chip design. Reachable interconnection system use either oblivious or adaptive routing algorithms to observe the way in use by a packet to its goal. Regardless of somewhat higher implementation complexity, adaptive routing benefit from better fault easiness characteristics, will enlarge network throughput, and reduces latency compared to insensible strategy when appearance with non-uniform or bursty traffic. Effective routing

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

algorithms create best use of the link bandwidth and extend transfer as required to sense of balance the stack. Ideal adaptive routing algorithms would accurately predict future congestion and route every message to minimize the contention. Since such an approach is impractical, most adaptive routing algorithm rule employ easy local congestion metrics in each router to see where to next send any given message.

III. METHOD

- **Dragonfly Topology**

The following symbols are utilized in our description of the dragonfly topology in this section and therefore the routing algorithms
N = Number of network terminals
p = Number of terminals connected to every router
a = Number of routers in every group
k= radix of the routers
k'= Effective radix of the group (or the virtual router).
h= Number of channels inside each router used to connect with different groups
g =Number of groups within the system
q =Queue depth of an output port
q_{vc} =Queue depth of an individual output VC
H = Hop count
Out_i =Router output port i

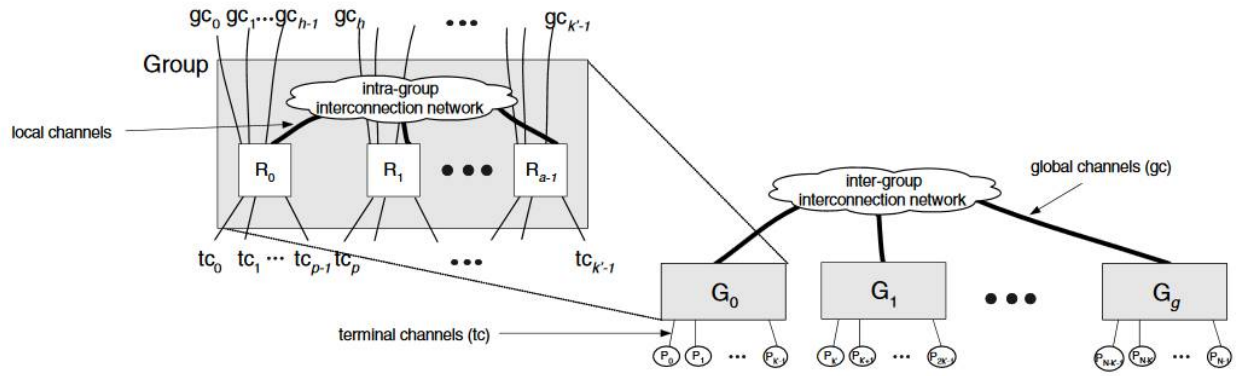

Figure (a) : Group of virtual routers Figure (b) : High level Dragonfly topology

The dragonfly may be a hierarchical network with 3 levels: router, group, and system as shown in Figure. At the bottom level, each router has connections to p terminals, a − one local channels—to different routers within the same group—and h global channels—to routers in different groups. Hence the radix (or degree) of every router is k = p + a + h − 1. A group consists of a routers connected via an intra-group interconnection network formed from local channels (Figure (a)). Every group has ap connections to terminals and ah connections to global channels, and all of the routers during a group collectively act as a virtual router with radix k' = a(p + h). This very high radix, k '>> k allows the system level network (Figure (b)) to be realized with very low global diameter (the maximum number of expensive global channels on the minimum path between any 2 nodes). Up to g = ah + 1 groups (N = ap(ah + 1) terminals) will be connected with a global diameter of 1. In contrast, a system-level network built directly with radix k routers would require a larger global diameter. In a maximum-size (N = ap(ah + 1)) dragonfly, there's precisely one connection between each pair of groups. In smaller dragonflies, there are more global connections out of each group than there are different groups. These excess global connections are distributed over the groups with every pair of groups connected by a minimum of $[\frac{ah+1}{g}]$ channels.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

The dragonfly parameters a, p, and h will have any values. However to balance channel load on load-balanced traffic, the network should have a = 2p = 2h. Because every packet traverses 2 local channels along its route (one at every end of global channel) for one global channel and one terminal channel, this ratio maintains balance. Additional details of routing and load-balancing are going to be discussed. because global channels are expensive , deviations from this 2:1 ratio should be done in a manner that over provisions local and terminal channels, so the expensive global channels remain fully utilized. That is, the network should be balanced so a ≥2h, 2p ≥2h.
By increasing the effective radix, the dragonfly topology is extremely scalable – with radix-64 routers, the topology scales to over 256k nodes with a network diameter of only 3 hops. Arbitrary networks will be used for the intra-group and inter-group networks in above Figures (a) and (b).

- **Routing**

We discuss minimal and non-minimal routing algorithms for the dragonfly topology. We show how global adaptive routing using native information leads to limited throughput and extremely high latency at intermediate loads. To overcome these problems, we propose new mechanisms to global adaptive routing, which offer performance that approaches an ideal implementation of global adaptive routing. Adaptive routing on the dragonfly is challenging because it's the global channels, the group outputs, that need to be balanced, not the router outputs. This leads to an indirect routing problem. Each router should pick a global channel to use using only local info that depends only indirectly on the state of the global channels. Previous global adaptive routing ways used local queue information, source queues and output queues, to get accurate estimates of network congestion. In these cases, the local queues were an accurate proxy of global congestion, because they directly indicated congestion on the routes they initiated. With the dragonfly topology, however, local queues only sense congestion on a global channel via backpressure over the local channels. If the local channels are overprovisioned, significant numbers of packets must be enqueued on the overloaded minimal route before the source router will sense the congestion. This results in degradation in throughput and latency.

IV. CONCLUSION

In this paper, we take a survey report to identify the impact of far-end congestion that occurs in large-scale networks because of long latency between neighboring routers and the different length channels in the topology. In which all previous related work congestion at the far-end of the channel is not accurately represented at the near-end since in-flight packets (or credits) that are being transmitted do not represent true congestion. In this survey we see that Transient congestion is the result of fluctuation of network queue occupancy due to random traffic variation and also gives inaccurate adaptive routing decisions.

REFERENCES

1. Jongmin Won, Gwangsun Kim, John Kim, S. K., anTed Jiang, Mike Parker, Steve Scott. “Overcoming Far-end Congestion in Large-Scale Networks” The 21st IEEE International Symposium on High Performance Computer Architecture (HPCA), 2015, page no. 415 – 427, year 2015.
2. J. Kim et al., “Technology-driven, highly-scalable dragonfly topology,” Computer Architecture, 2008. ISCA '08. 35th International Symposium on , Beijing, China, June 2008, page no.. 77–88, year 2008.
3. B. Arimilli, Ravi Arimilli, Vicente Chung Scott Clark, Wolfgang Denzel, Ben Drerup , Torsten Hoef “The percs high-performance interconnect,” in High Performance Interconnects (HOTI), 2010 IEEE 18th Annual Symposium *on*, Mountain View, CA, August 2010, page no. 75–82, year 2010.
4. S. Scott et al., “The blackwidow high-radix clos network,” 33rd International Symposium on Computer Architecture (ISCA'06) Boston, MA, page no. 16–28,year 2006.
5. P. Gratz et al., “Regional congestion awareness for load balance in networks-on-chip,” 2008 IEEE 14th International Symposium on High Performance Computer Architecture, Salt Lake City, UT, pp. 203–214, Feb 2008.
6. N. Jiang *et al.*, “Indirect adaptive routing on large scale interconnection networks,” ISCA 2009 The 36th International Symposium on Computer Architecture, Austin, TX, page no220– 231, June 2009.
7. W. Dally, “Virtual-channel flow control,” IEEE Transactions on parallel and distributed systems, vol. 3, no. 2, page no. 194–205, year 1992.
8. D. Helbing, “Traffic and related self-driven many-particle systems,” Rev. Mod. Phys., vol. 73, page no. 1067–1141, Dec year 2001.
9. J. Bell et al., “Boxlib users guide,” 2013, Center for Computational Sciences and Engineering, Lawrence Berkeley National Laboratory, year 2013.
10. H. Dong et al., “Quasi diffusion accelerated monte carlo,” Los Alamos National Laboratory, year 2011.
11. D. F. Richards et al., “Beyond homogeneous decomposition: Scaling long-range forces on massively parallel systems,” Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis , page no. 1-12, November year 2009.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

12. P. F. Fischer et al., 2008, http://nek5000.mcs.anl.gov. H. Adalsteinsson et al., "A simulator for large-scale parallel computer architectures," International Journal of Distribution System Technology, volume 1, issue 2, page no. 57–73, Apr. 2010.
13. "Characterization of the doe mini-apps," National Energy Research Scientific Computing Center. [Online]. Available: http://portal.nersc.gov/project/CAL/doe-miniapps.htm
14. J. Kim et al., "Flattened butterfly: A cost-efficient topology for high-radix networks," 34th annual international symposium on Computer architecture, San Diego, CA, page no. 126–137, June 2007.
15. J. Kim et al., "Microarchitecture of a high-radix router," 32nd International Symposium on Computer Architecture (ISCA'05), Madison, Wisconsin, page no. 420–431, June 2005.
16. J. Ahn et al., "Hyperx: topology, routing, and packaging of efficient large-scale networks," SC '09 Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, Portland, Oregon, Article 41, November 2009.