
Proceedings Paper:
Hassen, F and Mhamdi, L (2017) High-radix Packet-Switching Architecture for Data Center Networks. In: 2017 IEEE 18th International Conference on High Performance Switching and Routing (HPSR). 18th International Conference on High Performance Switching and Routing (HPSR), 18-21 Jun 2017, Campinas, Brazil. IEEE. ISBN: 978-1-5090-2839-9. EISSN: 2325-5609.

https://doi.org/10.1109/HPSR.2017.7968672

High-radix Packet-Switching Architecture for Data Center Networks

Fadoua Hassen Lotfi Mhamdi
School of Electronic and Electrical Engineering
University of Leeds, UK
Email: {elfha, L.Mhamdi}@leeds.ac.uk

***Abstract*—We propose a highly scalable packet-switching architecture that suits for demanding Data Center Networks (DCNs). The design falls into the category of buffered multistage switches. It affiliates the three-stage Clos-network and the Networks-on-Chip (NoC) paradigm. We also suggest a congestion-aware routing algorithm that shares the traffic load among the switch's central modules via interleaved connecting links. Unlike conventional switches, the current proposal provides better path diversity, simple scheduling, speedup and robustness to load variation. Simulation results show that the switch scales well with the port-count and traffic fluctuation and that it outperforms different switches under many traffic patterns.**

***Index Terms*—Data Center Networks switching fabric, Clos-network, Multi-Directional NoCs, Packets scheduling**

I. Introduction

Future DCN designs call for the replacement of conventional capacity-limited switches/routers with more scalable ones to increase the reliability and the throughput of the network, and to reduce the deployment and expansion costs. Data center traffic is frequently reported to be unstable across a variety of time-scales. Hence, congestion is likely to happen at any point of the DCN, especially if the switching substrate is incapable of handling the skewed traffic.

Single-stage packet-switches have been long adopted for their simplicity. They can fit for small networks. Yet, they are unsuitable for data center networks, as scaling a single-stage crossbar switch is rather costly, than infeasible. Multistage switches were proposed to get over the limitations of single-stage designs by interconnecting a number of Switching Elements (SEs) in a particular fashion. The Clos-network has been a typical solution, extensively studied and tested by both academia and industry (Cisco CRS-3 and Junipers T600 [1], [2]).

Multistage switches are classified with reference to the connection type, number of stages, and mainly the buffering strategy, etc. This criterion gave rise to all sorts of three-stage Clos switches, ranging from Space-Space-Space (S^3) to Memory-Memory-Memory (MMM) [3], [4]. Other combinations have also been investigated [5], [6]. Despite their scalability potential, multistage switches call for an excessive number of separate queues in the input modules running faster than the external line rate, to resolve the Head-of-Line blocking problem [5]. In addition to their cost, almost all existing Clos packet-switches perform poorly under unbalanced traffic. Memory-Space-Memory (MSM) is a popular design that presents good compromise of cost/complexity. Still, the bufferless nature of the middle stage mandates a centralized scheduling to perform the global matching between the set of input/output ports [5]. In an MMM switch, the contention is all absorbed by means of distributed buffers [4], dismissing any need for a central arbitration. All the same, this alternative is unscalable, since large buffers are required to enhance the switch performance under skewed traffic. The Clos switch with Uni-Directional NoC (UDN) fabric – Clos-UDN– was proposed in [7], in an attempt to build a large-scale switch for DCNs. The design calls for the interesting features of NoCs to reduce the complexity of the switching hardware, and scheduling process while achieving high performance. In [8], a wrapped-around Clos switch was described. The switch brings good scalability features. It is easily configurable. However, increasing the port count pushes up the size of the central NoC modules and leads to substantially raising the design cost.

Contributions and content

Motivated by the shortcomings of the previous proposals, we propose a highly scalable packet-switching architecture. In particular, we describe the Clos-MDN: A three-stage Clos-network switch with Multi-Directional NoC (MDNs) fabric [9]. Our contribution can be summarized in two main points:

- First, we made a radical change to the Clos-network by changing classical crossbar SEs by MDN modules. The MDN is a compact NoC fabric with small on-chip buffers, input-queued mini-routers, buffered flow-control, and Virtual Channels (VCs). Unlike the UDN fabric, MDN allows traffic flowing in all directions through deterministic routes, with no deadlocks. A central stage SE connects to its adjacent modules using interleaved links, leading to a significant extension of the switching facility.
- Our second contribution lies in the implementation of a *proactive* congestion-control scheme that tightly works with an appropriate scheduling algorithm, to enhance the throughput performance.

Both the wrapped-around design suggested in this paper, and the congestion-aware routing, are motivated by a relevant topic: Load balancing in DCNs. Actually, load-balancing has been long devoted to centralized controllers [10], network edge modules [11], [12], or end-hosts [13]. These solutions mandate a global traffic information to redistribute the load, making the

response delays too long as compared to short-lived congestion events encountered in DCNs. Latest works suggested solutions to amend congestion management in data centers by conveying part of the job to switches/routers [11], [13]. This approach is referred to as the *micro load-balancing* [14]. It allows fine time scale decisions and enhances the network performance when combined with the common practice *macro load-balancing*. The Clos-MDN switch has many architectural and scheduling advantages over the MSM, MMM and the two variations of the Clos-UDN switch as described in [7] and [8] – respectively. In general, the Clos-MDN switch:

1) Obviates the need for complex and costly input modules, by means of few, yet simple, input FIFO queues.
2) Avoids the need for a complex and synchronized scheduling process over a high number of input/output modules and port pairs.
3) Provides speedup[1], load balancing and path-diversity thanks to the NoC based fabric nature.
4) Allows the switch size to grow faster than with UDN modules for less design cost.
5) Deals better with skewed traffic thanks to the inter-CM links and the adaptive routing scheme.

The remainder of the paper is structured as follows. In Section II, we highlight the switch architecture and we describe the routing process. Section III is reserved for assessing the Clos-MDN switch performance under a range of traffic patterns. In Section IV, we compare the current proposal to the state-of-the-art multistage switches, and we conclude the paper in Section V.

II. Clos-MDN switching architecture

In this section, we outline the multistage switch topology and the packet buffers. Next, we provide a full description of the MDN modules. We also give details of the packet dispatching process and the routing scheme.

A. Switch terminology and packet buffers

We made a radical change to the conventional Clos switching architecture by plugging multi-directional NoC-based modules in the middle stage of the network. A typical flattened[2] multistage packet-switch design tends to make the data traffic flowing horizontally from the input modules (at the first stage of the network) until the output modules (at the third stage). We distribute the set of input/output ports such that each of the first and last stages of the Clos-network regroups n input and n output buffers, as depicted in Fig.1. An input FIFO is associated with an input port. It can receive at most one packet and sends at most one packet to a central module at every time slot. In the same analogy, each of the output

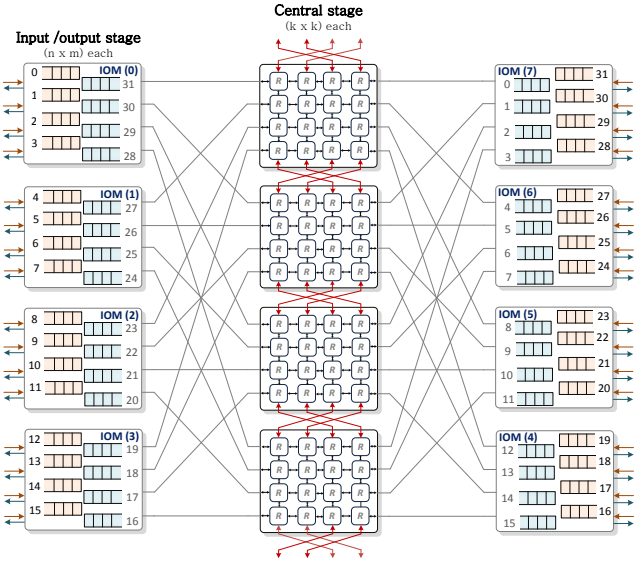


Fig. 1: An example of a 32 × 32 Clos-MDN packet-switching architecture.

queues is dedicated to an output port. It can receive at most n packets (from the different MDN modules) and forwards one packet to the output line card at every time slot. Throughout the paper, the first and last stage blocks are referred to as Input/Output Modules (IOMs). Our switching architecture[3] has $2k$ IOMs and m central MDNs, each of dimension $(k \times k)$. We consider the simple design case of $Bene's$ network for which $n = m$. This makes the switch architecture, the lowest-cost rearrangeably non-blocking Clos-network, and avoids the need for an insertion policy to distribute packets among the input buffers at the traffic arrival phase[4]

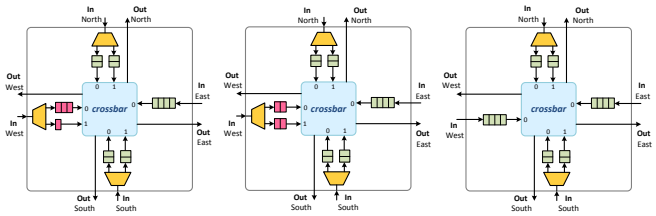
The major part of the switch is the MDN central modules. The single stage MDN crossbar switch was introduced in [15] as an extension of the UDN proposal [9]. The MDN design makes good use of the NoC pattern, and it efficiently builds a compact switching fabric. The geometry is a regular 2-D mesh of size $(k \times k)$, where inlets and outlets are placed on the perimeter of the squared layout. The MDN can be viewed as an optimized concatenation of two UDN fabrics where data traffic flow in two opposite directions (East/West and West/East). To preserve the integrity of packets, we use the store-and-forward switching mode. Two separate VCs are implemented to isolate traffic flows and to avoid deadlocks. Packets cross the first virtual channel VC1, if their corresponding output destination is located eastern to its input port. The second channel VC2 is used whenever the packet destination is located western to the input port. The on-chip routers are equipped with small input queues and a Round Robin (RR) arbitration unit that resolves the input contention. We consider evaluating

[1]We will use the term speedup to refer to the speed ratio at which the on-chip links of the NoC fabric can run with respect to the external links speed. Saying that the NoC switching elements run at a speedup SP, is equivalent to the on-chip routers removing up to SP packets from one input buffer, and sending up to SP packets to one output per time-slot.

[2]Other 3D architectures such as the hypercube and layered switches, would obviously allow traffic circulation in cubic way or in-between layers.

[3]For an arbitrary non-blocking Clos-network, the number of outlets in any of the first-stage modules (m) can differ from the number of its inlets (n).

[4]Generally, a non-blocking Clos-network switch can be of any size, where $m \geq 2n-1$. This would simply require packets insertion policy in the FIFOs input queues, should we need to maintain low-bandwidth buffers at the IOMs. We consider this to be out of the scope of the current work.



(a) Asymmetrical West router (b) Symmetrical West router (c) Center, North, South

Fig. 2: Buffers distribution in the MDN mini-routers. In the example, input buffers of on-chip-routers are of size 4 packets, each.

the performance of the Clos-MDN switch while we consider two ways to distribute the on-chip buffering space of mini-routers, per input: Symmetrical[5], and asymmetrical[6]. Fig.2 shows a high-level diagram of the on-chip routers.

B. Packet routing in the MDN fabric

We consider a static packet dispatching scheme from the IOMs, for which every input FIFO constantly delivers packets to the same MDN module on the connecting link. Packets travel in all directions in the central stage modules until they reach the external links bridging their corresponding IO modules. Based on their destination ports, packets are routed inside the MDN modules as following: The first step consists on finding out the IO module index to which is related the packets' ultimate destination port. Upon their entry to a CM, packets are locally routed using a combination of two algorithms: "XY" algorithm and the "Modulo" routing. The "XY" algorithm has been long ago introduced for mesh NoCs [16][7]. The "Modulo" algorithm is an improved version of the basic "XY". It introduces an extra turn in one intermediate column before the last one to better balance the traffic in the mesh. It is used in the MDN switch if the local input and output ports are parallel.

C. Proactive congestion management

Our previous results have shown that a static packet dispatching and an oblivious routing scheme, are irrelevant to skewed traffic arrivals [8]. In fact, NoC-based switches can get congested under some traffic patterns causing the packet delay to become longer and the switch throughput to deplete. Therefore, we make the central stage modules of the Clos-MDN switch capable of sharing the traffic via intermediate links (see the connection algorithm below). We use two VCs on each link to transport packets depending on the flow direction and to prevent deadlocks. The additional connections extend the advantage of the NoC geometry to the Clos-network and make the multistage switch architecture a wrapped-around network.

Algorithm 1 : Inter-CM interleaved connections

Require: The coordinates of the mini-router MR^r in the CM of index r
1: **for** $r \in \{1, \ldots, m-1\}$ **do**
2: $r' \leftarrow ((r+1) \ mod \ m)$ and $r'' \leftarrow ((r-1) \ mod \ m)$
3: **for** $i \in \{1, \ldots, k-1\}$ **do**
4: $j \leftarrow ((\frac{k}{2} + i) \ mod \ k)$
5: $\text{MR}^r(k-1, i)$ connects to $\text{MR}^{r'}(0, j)$
6: $\text{MR}^r(0, i)$ connects to $\text{MR}^{r''}(k-1, j)$
7: **end for**
8: **end for**

Choosing an interleaved configuration as described in Algorithm 1, is made to ensure that sending packets from their original congested CMs to a neighbouring module does not increase the remaining hop count[8]. Our ultimate goal is to maximize the switch throughput under coarse traffic without affecting the delay performance. Therefore, we adopt a metric that is suitable for the routing scheme to correlate well with the global Clos-network congestion status while being inexpensive to compute. We consider the Regional Congestion Awareness (RCA) [17] to evaluate and to propagate congestion information proactively[9], across the central module of index r and its direct neighbours (blocks of indexes $((r-1) \ mod \ m)$ and $((r+1) \ mod \ m)$). The congestion metric weights both distance (hop count until the exit port) and buffers occupancy to make sure that the traffic is adaptively transferred through minimal paths, and that the average packet delay is little affected by the inter-module routing decision.

III. Performance Analysis

We evaluate the performance of the Clos-MDN switch under a wide range of traffic, and we compare it to the state-of-the-art multistage switches. Our simulation models are built on top of an event-driven simulator written in C language. For all of the simulation scenarios, the capacity of input buffers of the mini-routers (buff) is 4 packets[10] each; unless it is otherwise stated. In what follows, we tried to adjust settings of the Clos-UDN and Clos-MDN switches to make the comparison fair. We consider the same buffering space, and we make the Clos parameters n, m equal for both switches configurations – in which case the performance disparity is mainly attributed to the NoC modules. The essence of the Clos-MDN is in the prospect of building high-capacity switching architectures with small sized NoC modules. Note that for any switch valency, a central stage UDN[11] module uses four times as many mini-routers as an MDN module employs. Therefore, we consider trading area by speedup in the Clos-MDN switch, since it is

[5]In a symmetrical buffer space distribution, the channels VC1 and VC2 are allocated the same buffer space

[6]In an asymmetrical buffer space distribution, the west routers have 2/3 of the buffer depth for VC1 and 1/3 for VC2, and east routers use 1/3 of the port buffering space for VC1 and 2/3 of it for VC2

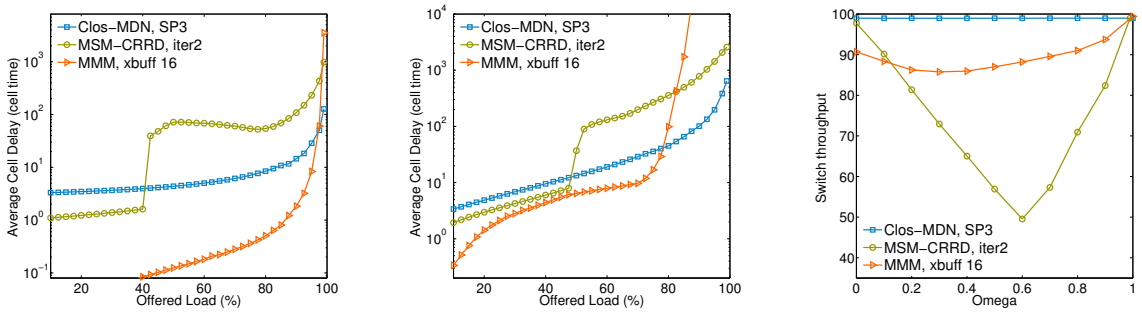
[7]The "XY" algorithm is used to route packets in the MDN whenever the local output port is perpendicular to its input port. It simply starts by forwarding packets horizontally to the correct column (x-coordinate) and then vertically to the right row (y-coordinate).

[8]In the worst case scenario, a packet will do the same number of hops in the neighbour CM as it would have in its non-congested CM for two reasons: First, the inter-module routing algorithm considers the distance metric and second packets are minimally routed within a single MDN.

[9]Details of the congestion-aware routing algorithm used in the Clos-MDN switch are available in [8].

[10]All packets are assumed to have the same size.

[11]For full mesh design where the number of the unidirectional NoC stages is equal to the number of inlets/outlets [7].



(a) Bernoulli Uniform traffic (b) Bursty uniform (c) Throughput stability

Fig. 3: Performance for 256-ports MSM, MMM and Clos-MDN Switches.

not expensive to run short on-chip links fast using the currently available technology [9], [18].

The first set of simulations compares the performance of the Clos-MDN switch to the MSM switch (using the Concurrent RR Dispatching scheme- CRRD [5]) and the MMM switch[12] as has been described in [3]. The Clos-MDN architecture fits into the category of buffered multistage switches. However, we strongly think that comparing its performance to MSM, helps to analyse the response of the current design, with respect to its features (size, buffering space, scheduling complexity, etc.). In Fig.3 (a) and Fig.3 (b), are shown the simulation results for a uniform Bernoulli *i.i.d* and bursty traffic – respectively. MSM performs well under light and medium uniform loads. However, its average packet delay rises sharply (at around 40% of the load for a Bernoulli *i.i.d* traffic, and 55% of the load, for a uniform arrival of bursts), and never pulls down. The MMM switch outperforms both MSM and Clos-MDN thanks to its large internal buffers. Yet, its performance noticeably degrades under bursty traffic (throughput saturation at 84%, for crosspoint buffers worth of 16 packet, each). The Clos-MDN switch, experiences relatively higher delay than MSM and MMM, under light-to-medium loads. The pipelined structure of the NoC-based central modules is behind this initial cumulative delay. Still, the delay variation is quasi-stable showing a good scalability of the Clos-MDN with the load increase. We also note that the switch throughput is much higher than the MSM and MMM switches, though a reasonable speedup is used (Fig.3 (c)).

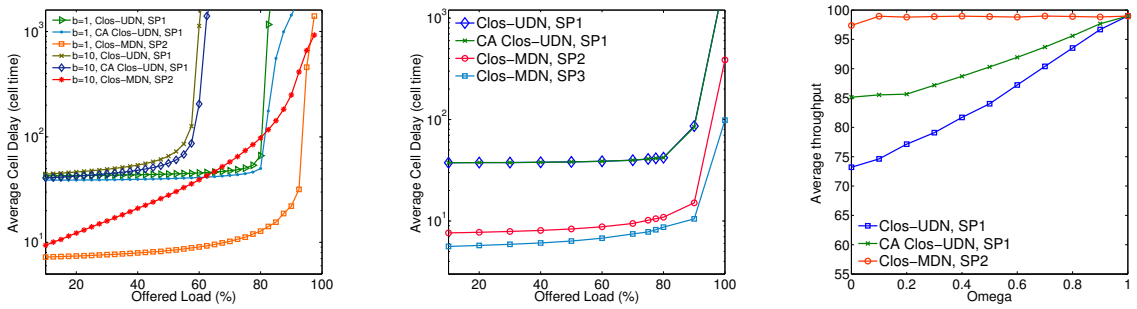
Next, we compare the performance of the Clos-UDN, the Congestion-Aware Clos-UDN (that we denote CA Clos-UDN in the graphs), and the Clos-MDN proposals. Fig.4 (a), shows that under uniform traffic[13], The Clos-UDN and CA Clos-UDN switches have comparable performance. They both yield a higher latency under light-to-medium loads, as the number of NoC stages is much higher than in the Clos-MDN switch. Clearly, the initial delay correlates with the number of NoC stages at the middle stage of the Clos-network. Filling in the pipeline takes few time slots before the latency variation becomes quasi-constant. We note that with few on-chip mini-routers and a small speedup factor ($SP = 2$), the Clos-MDN proposal outperforms the Clos-UDN switch variations, and achieves high throughput. Overall, trading the area by speedup improves the Clos-MDN throughput by approximately 20% under Bernoulli *i.i.d* traffic, and 30% under bursty traffic. Under hot-spot traffic, Clos-MDN still defeats Clos-UDN switches in terms of packet latency (Fig.4 (b)) and throughput (Fig.4 (c)).

We further study the scalability and robustness of the MDN-based multistage design under a bursty traffic, by varying the architectural settings. We investigate the effect of speedup, load, on-chip buffers capacity, buffers distribution, and switch valency. Increasing the burst size entails higher packet delay, and throughput degradation, as shown in Fig.5 (a) and Fig.5 (b). Two conclusions can be drawn: First, increasing the speedup factor boosts the switch performance, even though large bursts of packets break into the switch. Second, increasing the capacity of the on-chip buffers (from 4 packets to 6 packets, each, in our simulations), also lifts up the Clos-MDN's performance. Under bursty non-uniform traffic, the Clos-MDN performs a little bit better when the buffering capacity is asymmetrically distributed among VCs. However, this proves to have no remarkable effect on the switch's performance under hot-spot traffic as depicted in Fig.5 (c).

The last set of simulations shows that the Clos-MDN switch is flexible to size variation. With small additional buffering, and speedup, we can adequately tune the switch settings and push up its throughput under uniform (Fig.6 (a)), and non-uniform traffic patterns (Fig.6 (b) and Fig.6 (c)). Although a speedup of three proves enough for a (256×256) Clos-MDN switch to achieve full throughput, it is still insufficient to get full throughput for a 512-ports switch. Increasing the NoC speedup does not resolve the persistent backlogs that can form inside the MDN modules under heavy traffic loads. Under skewed traffic (hot-spot and diagonal traffic), the Clos-MDN architecture still perform well as Fig.6 (b) shows.

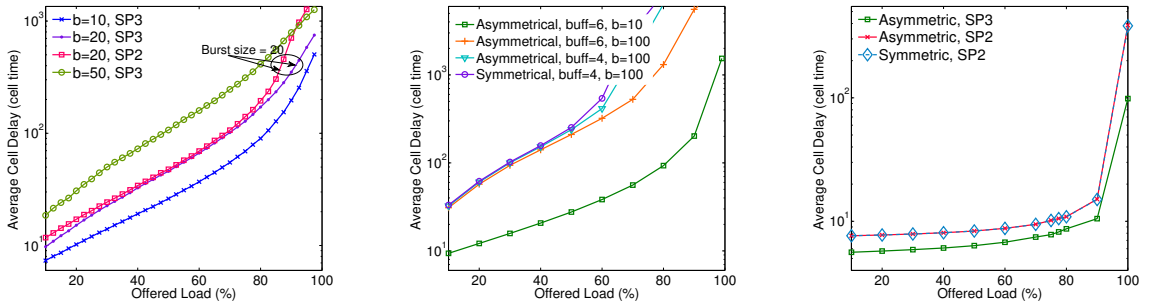
[12]We test MSM with 2-iterations CRRD matching since even with larger iterations the switch performance converges to nearly the same values [5]. We also set the MMM crosspoint buffers (xbuff) to 16 packets as with only one packet crosspoint buffering, the switch throughput does not exceed 65% under bursty traffic [3].

[13]Bernoulli *i.i.d*, for a burst size of 1 packet, and bursty uniform, for a burst size of 10 packets



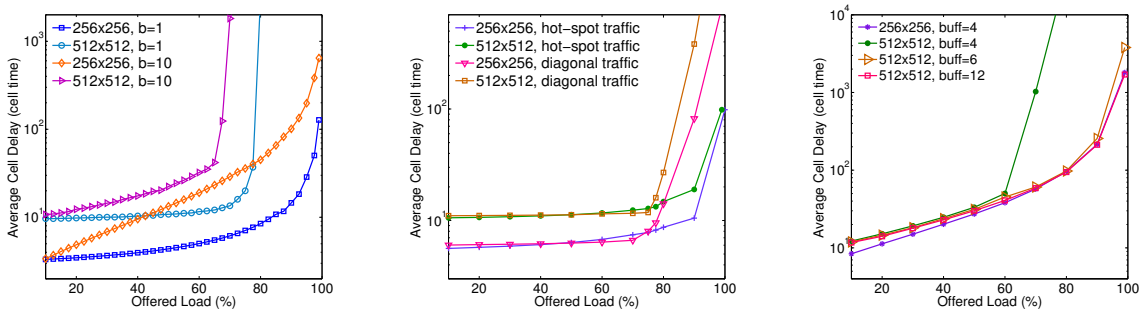
(a) Uniform traffic (b) Bernoulli Hot-spot traffic, $\omega = 0.5$ (c) Throughput stability

Fig. 4: Performance for 256-ports Clos-UDN/MDN Switches.



(a) Asymmetrical, Bursty non-unifrom (b) Bursty non-uniform (c) Bernoulli Hot-spot traffic, $\omega = 0.5$

Fig. 5: Delay performance for 256-ports Clos-MDN Switches under non-uniform traffic.



(a) Uniform traffic (b) Unbalanced traffic (c) Bursty non-uniform

Fig. 6: Impact of switch size on performance of Clos-MDN Switch, $SP = 3$.

In accordance with the previous conclusions, increasing the capacity of on-chip buffers effectively promotes the overall switch performance. However, simulations show that there is little interest in further increasing the buffers' capacity (the switch throughput converges with buff$= 6$ and there is little delay improvement if we rise buff to 12 packets as shown in Fig.6 (c)).

IV. Related work

Scalable switches/routers have attracted attention since the early appearance of computing. Motivations to build large-scale switches are numerous. Primarily, there is an urgent need to integrate scalable and high-performance switches/routers in modern DCNs, to absorb the proliferating data traffic. Yet, commodity switches either lack scalability, or exhibit prohibitive cost and hardware complexity. Single-stage crossbar switches have been long used for their simplicity, cost effectiveness, and reliability. However, for an $N \times N$ switch, it is too difficult to scale both, the data path core (that has $2N$ links), and the N^2 crosspoints. It is also hard to scale the scheduling process that has a time complexity in the order of $\mathcal{O}(N^2)$. The Clos-network interconnects offer a good alternative to scale up switches for reasonable cost and complexity ratio. Bufferless multistage switches like the MSM

switch [5], [19] are appealing for their low cost. Yet, they need a global arbitration, to perform ports matching and paths allocation. So far, centralized scheduling algorithms are not only too complex to implement but also perform sub-optimally [20] [21]. Fully buffered Clos-switches, mandate large and expensive internal memories to accommodate packets in flight [3]. The advantage of buffers comes in terms of scheduling since no centralized arbitration is needed, and contention is merely absorbed. However, under unbalanced traffic or large-size bursts of packets, the switch's performance is susceptible to collapse, unless more buffering is provided. Common MSM and MMM switches, adopt many Virtual Output Queues (VOQs) that run fast[14] to alleviate the HoL blocking problem. On the scheduling/dispatching front, the cost and practicality are still an issue.

The NoC paradigm has emerged as an attractive candidate for designing packet-switches [9], [15], [18], [22]–[24]. It provides a natural backbone for switches design since it includes switching elements, buffering within on-grid routers, and built-in flow-control mechanisms. So far, NoCs have been used in single-stage switching architectures to overcome limitations of classical crossbar switches[15]. In 2015 Karadeniz et $al.$ suggested a single-stage stage switch with Output-Queued (OQ) NoC fabric in [25]. In more recent works, NoCs were part of the multistage packet-switching architectures design. In [7], [26], authors respectively described three-stage Clos packet-switches with IQ and OQ uni-directional NoC modules. They also suggested a congestion-aware routing for a fully connected Clos-network switch with NoC modules in [8]. The Clos-MDN switch discussed in this paper is a large-capacity multistage switch, with compact NoC fabric. The middle-stage modules are much more optimized than the UDNs blocks used in the Clos-UDN switch [7], [8]. They are built with full exploitation of the NoC concept. the Clos-MDN switch embroils VCs, and a proper distribution of the on-chip buffering space, to lower the packet latency, and to maximize the switch throughput. Moreover, it is easily amenable to scale to large sizes.

V. Conclusion

In this paper, we present a multistage packet-switching architecture with multi-directional NoC fabric. The switch overcomes some shortcomings of the conventional multistage switches. It obviates the need for complex and costly buffering structures such as VOQs. It also avoids highly complex scheduling algorithms of bufferless Clos switches and large crosspoint buffers of common MMM switches. Using an efficiently designed NoC fabric, the Clos-MDN switch proves scalable in port count and load fluctuation.

VI. Acknowledgement

This work was supported by the EU Marie Curie Grant (SCALE: PCIG-GA-2012-322250).

[14]Input queues need run $(n+1)$ times faster than the external line rate in an input module of dimension of $n \times m$.

[15]In addition to the scalability potential, the hardware cost and the complexity of the scheduling process are two related topics of investigation.

References

[1] "Cisco," 2016. [Online]. Available: http://www.cisco.com/c/en/us/products/switches/nexus-5000-series-switches/datasheet-listing.html

[2] "Juniper Networks," June 2015. [Online]. Available: http://www.juniper.net/assets/us/en/local/pdf/datasheets/1000414-en.pdf

[3] Z. Dong and R. Rojas-Cessa, "Non-blocking memory-memory-memory Clos-network packet switch," in *Sarnoff Symposium*, 2011, pp. 1–5.

[4] Y. Xia, M. Hamdi, and H. J. Chao, "A practical large-capacity three-stage buffered Clos-network switch architecture," *Trans. Parallel Distrib. Syst.*, vol. 27, no. 2, pp. 317–328, 2016.

[5] E. Oki, Z. Jing, R. Rojas-Cessa, and H. J. Chao, "Concurrent round-robin-based dispatching schemes for Clos-network switches," *ACM Trans. Netw*, vol. 10, no. 6, pp. 830–844, 2002.

[6] X. Li, Z. Zhou, and M. Hamdi, "Space-memory-memory architecture for Clos-network packet switches," in *ICC*, 2005, pp. 1031–1035.

[7] F. Hassen and L. Mhamdi, "A multi-stage packet-switch based on NoC fabrics for data center networks," in *Globecom Workshops*, 2015, pp. 1–6.

[8] ——, "Congestion-aware multistage packet-switch architecture for data center networks," in *Proc. GLOBECOM*, 2016, pp. 1–7.

[9] L. Mhamdi, K. Goossens, and I. V. Senin, "Buffered crossbar fabrics based on networks on chip," in *CNSR*, 2010, pp. 74–79.

[10] J. Perry, A. Ousterhout, H. Balakrishnan, D. Shah, and H. Fugal, "Fastpass: A centralized zero-queue datacenter network," in *ACM/SIGCOMM Computer Communication Review*, vol. 44, no. 4, 2014, pp. 307–318.

[11] M. Alizadeh, T. Edsall, S. Dharmapurikar, R. Vaidyanathan, K. Chu, A. Fingerhut, F. Matus, R. Pan, N. Yadav, G. Varghese *et al.*, "CONGA: Distributed congestion-aware load balancing for datacenters," in *ACM/SIGCOMM Computer Communication Review*, vol. 44, no. 4, 2014, pp. 503–514.

[12] P. Wang and H. Xu, "Expeditus: Distributed load balancing with global congestion information in data center networks," in *Proc of ACM CoNEXT on Student Workshop*, 2014, pp. 1–3.

[13] K. He, E. Rozner, K. Agarwal, W. Felter, J. Carter, and A. Akella, "Presto: Edge-based load balancing for fast datacenter networks," *ACM/SIGCOMM Computer Communication Review*, vol. 45, no. 4, pp. 465–478, 2015.

[14] S. Ghorbani, B. Godfrey, Y. Ganjali, and A. Firoozshahian, "Micro load balancing in data centers with DRILL," in *HotNets 14th*, 2015, p. 17.

[15] K. Goossens, L. Mhamdi, and I. V. Senin, "Internet-router buffered crossbars based on networks on chip," in *DSD*, 2009, pp. 365–374.

[16] W. Zhang, L. Hou, J. Wang, S. Geng, and W. Wu, "Comparison research between XY and odd-even routing algorithm of a 2-dimension 3x3 mesh topology network-on-chip," in *IEEE WRI Global Congress on Intelligent Systems.*, vol. 3, 2009, pp. 329–333.

[17] P. Gratz, B. Grot, and S. W. Keckler, "Regional congestion awareness for load balance in networks-on-chip," in *HPCA*, 2008, pp. 203–214.

[18] T. Karadeniz, L. Mhamdi, K. Goossens, and J. Garcia-Luna-Aceves, "Hardware design and implementation of a network-on-chip based load balancing switch fabric," in *ReConFig*, 2012, pp. 1–7.

[19] J. Kleban and A. Wieczorek, "CRRD-OG: A packet dispatching algorithm with open grants for three-stage buffered Clos-network switches," in *Workshop on HPSR*, 2006, pp. 6–pp.

[20] N. Chrysos, C. Minkenberg, M. Rudquist, C. Basso, and B. Vanderpool, "Scoc: High-radix switches made of bufferless clos networks," in *HPCA*, 2015, pp. 402–414.

[21] H. J. Chao, Z. Jing, and S. Y. Liew, "Matching algorithms for three-stage bufferless Clos network switches," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 46–54, 2003.

[22] A. Bitar, J. Cassidy, N. Enright Jerger, and V. Betz, "Efficient and programmable Ethernet switching with a NoC-enhanced FPGA," in *Proc of ACM ANCS'10*, 2014, pp. 89–100.

[23] F. Moraes, N. Calazans, A. Mello, L. Möller, and L. Ost, "HERMES: An infrastructure for low area overhead packet-switching networks on chip," *INTEGRATION, the VLSI journal*, vol. 38, no. 1, pp. 69–93, 2004.

[24] E. Bastos, E. Carara, D. Pigatto, N. Calazans, and F. Moraes, "MOTIM-A scalable architecture for Ethernet switches," in *ISVLSI*, 2007, pp. 451–452.

[25] T. Karadeniz, A. Dabirmoghaddam, Y. Goren, and J. Garcia-Luna-Aceves, "A new approach to switch fabrics based on mini-router grids and output queueing," in *ICNC*, 2015, pp. 308–314.

[26] F. Hassen and L. Mhamdi, "A scalable packet-switch based on output-queued NoCs for data centre networks," in *ICC*, 2016, pp. 1–6.