Relieving Hotspots in Data Center Networks with Wireless Neighborways

Liqin Shan*, Chang Zhao*, Xiaohua Tian*, Yu Cheng†, Feng Yang* and Xiaoying Gan*
*Department of Electronic Engineering, Shanghai Jiao Tong University
{tianwochazhuangshan, messi0802, xtian, yangfeng, ganxiaoying}
@sjtu.edu.cn
†Department of Electrical and Computer Engineering, Illinois Institute of Technology
{cheng}@iit.edu

***Abstract*—Recent studies show that the 60GHz wireless technology could help resolving the hotspot issue in data center networks (DCNs). However, transmissions over 60GHz suffer from limitations of short transmission range and blockage, which makes it a new challenge how to appropriately establish wireless links in the DCN. In this paper, we propose to integrate wireless links into the DCN with the *wireless neighborway* scheme. Each top-of-rack switch (ToR) has multiple 60GHz wireless links connecting to its neighboring ToRs, which are termed as neighborways. The elephant flow from the sending ToR can be partially offloaded through neighborways, which are then delivered to the ToRs around the receiving ToR through wired links of the DCN, and finally converged to the receiving ToR. The fundamental challenge for the design is how to prevent the offloaded traffics from forming new hotspots in the network fabric. To this end, we developed a wireless network planning solution with corresponding IP address assignment and traffic engineering scheme, which can leverage the potential under-utilized wired links in the DCN and meanwhile avoid forming new hotspots. The simulation results show that the proposed scheme can notably relieve the hotspots in the DCN.**

I. Introduction

The data center network (DCN) has become the fundamental component of today's Internet infrastructure. In the current DCN, each rack contains multiple servers, which are connected by a top-of-rack switch (ToR); ToRs are then connected by a number of aggregation or core switches, so that all servers at different racks within the DCN are connected. Many applications within the DCN, such as MapReduce and search engine [1], require a server to exchange information with remote servers that may locate in different racks before proceeding the local computation.

Traditional tree-structured DCNs may suffer from oversubscription, which means that the actual throughput of the DCN is much lower than the available bandwidth of the server's network interface card (NIC). In order to provide the maximum and same communication bandwidth between any arbitrary pair of severs, efforts have been made to improve the DCN structure and design new routing mechanisms [1], [2], where more links and switches are added with multipath routing schemes proposed in order to avoid the oversubscription [2], [3].

However, even if any server in the DCN is able to communicate with any other server at full NIC bandwidth, the performance can also be affected by the hotspots in the DCN incurred by the unbalanced traffic, where some hot nodes need to transmit a high volume of traffic and incur congestion in the edge links of the DCN. Since the distribution of the hot nodes are non-deterministic, adding more wired links for certain nodes to alleviate the congestions is inefficient considering the costs of large amount of hardware, manual efforts and cabling complexity have to be paid.

Recent studies show that 60GHz wireless technology could help improve global job completion time in DCNs [4], [6], [9], [11], where communication capacity of wireless links can be flexibly and dynamically enhanced to fulfill the real-time traffic demands of hotspots. Nevertheless, the full utilization of wireless links in the DCN is hindered by two notable limitations of 60GHz technology: First, the transmission range of 60GHz wireless links is limited to a couple of meters, which makes it difficult for distant communication parties to reach each other in one hop; Second, the strength of signals at 60GHz attenuates significantly upon being blocked by any object larger than a couple of milimeters, which makes the multi-hop routing time-consuming and unstable [11]. These limitations incur the new challenge of how to appropriately establish wireless links in the DCN to fully utilize the 60GHz wireless transmission capability.

In this paper, we propose to integrate wireless links into the DCN with the *wireless neighborway* scheme. Each ToR has multiple 60GHz wireless links connecting to its neighboring ToRs, which are termed as neighborways. The high-volume traffic from the sending ToR can be partially offloaded through neighborways to neighboring ToRs, which are then delivered through appropriate wired links of the DCN to ToRs around the receiving ToR, and finally converged to the receiving ToR over the neighborways around. With the design, each sending ToR with potential high-volume traffic can be flexibly enhanced with multi-gigabit transmission capacity. The neighborway scheme accommodates the features of the 60GHz, where only the short-range one-hop wireless transmission around each ToR is needed. The wireless transmission capacity is appropriately integrated into the wired DCN fabric as the complement. The fundamental challenge for the design

978-1-4799-3512-3/14/$31.00 ©2014 IEEE

is how to prevent the offloaded traffics from forming new hotspots in the network fabric. To this end, we developed a wireless network planning solution with corresponding IP address assignment and traffic engineering scheme, which can leverage the potential under-utilized wired links in the DCN and meanwhile avoid forming new hotspots. The simulation results show that the proposed scheme can notably relieve the hotspots in the DCN.

The rest of this paper is organized as follows. In Section II, we present more description of DCNs with wireless links. In Section III, we propose the neighborway service model and the core design issue. In Section IV, we present the network planning solution for the neighborway, with corresponding IP address assignment and traffic engineering scheme in order to realize the service model. In Section V, we evaluate the performance of the neighborway scheme by providing simulation results. Finally, in Section VI, we present conclusion remarks.

II. DCNs with Wireless Links

A. Integrating Wireless Links into DCNs

Integration of wireless links into the DCN is initially proposed to reduce the complexity of cabling [15]. In order to partially or completely replace wires for the connectivity in the DCN, the 60GHz wireless connectivity is chosen, which can possibly achieve the same level of bandwidth, reliability and security as its wired counterpart. The 60GHz DCN architecture design and corresponding challenges are presented in [15], but concrete technical solutions are not provided.

Taking the radical position, Shin et al. investigate the feasibility of complete wireless data centers [14], where a system-level architecture that incorporates a rack-level topology and a dedicated geographic routing protocol is demonstrated. With a number of significant challenges confronted with the complete wireless DCN [14], it is more practical to integrate wireless links into a hybrid Ethernet/wireless DCN architecture. Cui et al. propose to realize the wireless DCN in a simple tree structure, where a distributed scheduling scheme is developed to arrange the wireless links [4], [5].

Halperin et al. propose the *flyway* mechanism, where the 60GHz wireless links called flyways are dynamically setup and combined with the base wired DCN links to alleviate hotspots incurred by oversubscription [8], [9]. The flyway scheme is oblivious of the topology of the wired link base DCN network. Each ToR switch is installed with multiple 60GHz wireless devices with directional antennas, which can be electronically steered to different directions to establish wireless connections with other ToR switches. However, accurate steering the directional antenna imposes notable engineering difficulties to the control and management of the DCN; moreover, the direct wireless communication between ToR switches far away from each other may incur interference and signal attenuation, which may leave the wireless communication capability not fully exploited.

In order to fully exploit the augmented wireless links, two inherit limitations, short range and blockage as mentioned earlier, must be addressed appropriately. Zhou et al. propose the *3D beamforming* scheme for the DCN, where the 60GHz signals bounce off the DCN ceiling to establish the wireless connection with other ToR switches [11]. While the 3D beamforming scheme could resolve the drawbacks of 60GHz wireless links to some extent, the ceilings of DCNs have to be refurnished to provide perfect reflection without degrading energy or changing path loss characteristics. Moreover, it will be challenging for the 3D beamforming scheme to be applied in the modular DCN [12], where the racks of servers are installed in containers and containers are placed in the open space.

B. Relieving Hotspots with Wireless Links

Besides the oversubscription effect studied in [9], [11], the unbalanced traffic or elephant flow can also incur the hotspot in the DCN. The DCN switch is basically performing the functionality of *store-and-forward*. As the buffer size and the output rate of the switch are both limited, the input rate may exceed the output rate, which could make the proportion of the occupied switch buffer, termed as *cache occupancy rate*, to be high. It is straightforward that the buffer will easily overflow with high cache occupancy rate according to the queuing theory. Moreover, since most of flows in the DCN are over TCP protocol, the lost packets will be retransmitted thus potentially causing switch even more unstable. Hotspots in the DCN are actually those switches with high cache occupancy rates, where a small amount of extra packets can cause the switch's buffer to overflow thus significantly impact the performance of the DCN.

Consequently, even if the current DCN is able to guarantee any server can communicate with any other server at full NIC bandwidth, the hotspots also exist. As the flow distribution is dynamic in the DCN, the distribution of the hotspots is also non-deterministic. In this case, it is natural to utilize the flexible wireless links to relieve the hotspot by temporarily enhancing it with multi-gigabit output capacity. Our perspective in this paper is that the wireless links should be complementary to the wired DCN. While existing work on the hybrid wired/wireless DCN pays little attention to the effect of wired DCN topology on the overall performance, we initiate to explore the flexibility of wireless links to effectively utilize the potentially under-utilized wired transmission capacity.

III. Service Model of Neighborways

We perform our investigation in the context of the Fat-Tree structured DCN [1], which could shed light on the design of DCNs with other topologies. A simple Fat-Tree structured DCN with 4-port switches is illustrated in Fig.1 [1]. The rectangle in the figure represents the concept of pod in the Fat-Tree, which means some part of the IP address of the switches are the same [1]. In the proposed service model, the 60GHz wireless transceivers with directional antennas are installed on the top of each rack to avoid blockage by human beings. We assign 4 radios to be associated with each ToR switch to establish wireless neighborways with 4 neighboring

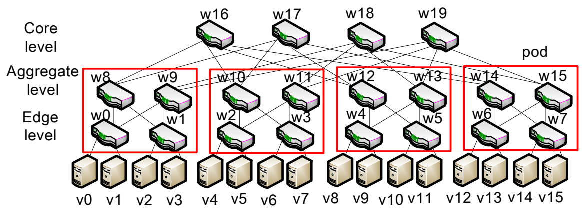


Fig. 1. Fat-Tree with 4-port switches.

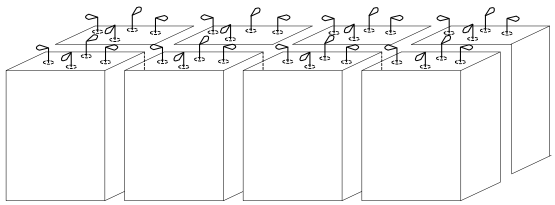


Fig. 2. Radio transceivers are placed on top of each rack.

ToR switches around. The racks in the DCN are supposed to be deployed as shown in Fig.2 [10] [16].

The idea of the neighborway scheme is illustrated in Fig.3. When the hotspot issue occurs at any ToR switch in the edge level of the DCN, the equipped wireless links can be activated to enhance the local transmitting capacity. Since there are 4 radio transceivers associated with each ToR switch, the hotspot switch can offload part of the traffic to the nearby ToR switches through the wireless neighborways as shown in Fig.3. The solid line in the aggregation and core level denotes the wired links to be used in the DCN without the wireless links. The offloaded traffics handled by the surrounding ToRs of the sender ToR are delivered to corresponding neighboring ToRs around the receiver ToR through wired links that are denoted by the dashed lines. Note that the dashed lines and the solid line in the aggregation and core level are independent to each other, otherwise it means the paths for the offloaded traffics may share some wired transmission capability, which could incur new hotspots in the DCN fabric.

The neighborway service model has the following two notable advantages:

- There is no need to worry the limitations of short range and blockage of 60GHz wireless links, since every link is an one-hop transmission in a short distance.
- The hotspots are sparsely populated in the practical DCN [8], thus using independent wired links to offload traffic can exploit the potentially under-utilized wired links in the DCN.

In order to realize the service model, *the core design issue is how to guarantee the wired links used to deliver offloaded traffics are independent with each other*, which will be addressed in the following section.

IV. Design of Neighborways

A. Place Neighbors in Different Pods

The logical topology of the Fat-Tree DCN can be presented as shown in Fig. 1; however, it is non-trivial mapping the logical topology into the physical layout as shown in Fig. 2

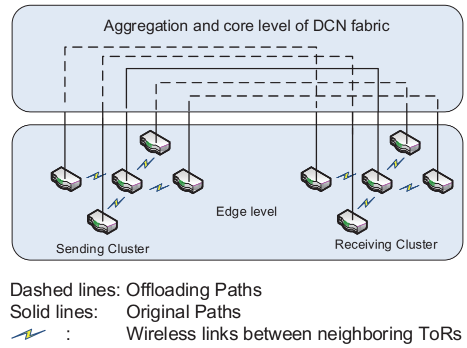


Fig. 3. Service model of the wireless neighborway scheme.

with satisfying the path-independent requirement in the service model.

Rule 1: Any ToR and its 4 neighbors around should be placed in different pods.

We use Fig. 1 to explain the rule. The hotspot occurs at the edge layer ToRs and aggregation layer ToRs. The former ones normally are the origins of the high-volume traffic [7], while the latter ones could worsen the performance of traffic from the former ones. For example, if the hotspot occurs at $w0$, it is possible that the 4 links $w0 \rightarrow w8 \rightarrow w16$, $w0 \rightarrow w8 \rightarrow w17$, $w0 \rightarrow w9 \rightarrow w18$ and $w0 \rightarrow w9 \rightarrow w19$ are loaded with considerable traffics. If $w0$ establishes the neighborway with $w1$ at this time, $w1$ may use the same wired links as $w0$ to offload part of the elephant traffic, which could possibly incur congestion in the aggregation layer.

Even if the core layer may use some expensive switches for larger capacity, offload traffic to the switches in the same pod may also incur congestions between the edge and the aggregation layer. If $w0$ offloads traffic to $w1$ in the same pod as in Fig. 1, it is very possible that links $w0 \rightarrow w8$, $w0 \rightarrow w9$, $w1 \rightarrow w8$ and $w1 \rightarrow w9$ will be overloaded. The traffic from $v2$, $v3$ to $v0$, $v1$ can be affected. It worth noting that the transmission capacity between the edge and the aggregation layer is more precious than that between the aggregation and the core layer. This is because the former also needs to take care of the traffic in the same pod. Many applications such as the social networking prefer to have more intra-pod traffic exchange for the real time consideration [13]. Therefore, it could be better if ToR switches for offloading are from different pods in the DCN, which can save more edge-aggregation layer bandwidth to facilitate the intra-pod traffic exchange.

Rule 2: Offloading ToRs on the sender side and corresponding offloading ToRs on the receiver side should be placed in different pods.

Recall that we have 4 neighboring ToRs for offloading the traffic from the sending ToR and 4 corresponding neighboring ToRs around the receiving ToR as shown in Fig. 3. Each offloading switch delivers the traffic to the receiving switch through wired links. If the two parties are with in the same pod, for example, $w0$ and $w1$ in Fig. 1. The links $w0 \rightarrow w8 \rightarrow w1$ and $w0 \rightarrow w9 \rightarrow w1$ will be used, which may affect the intra-pod traffic exchange as mentioned earlier. It is preferred if the

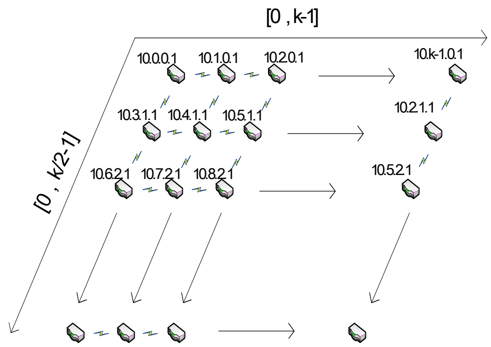


Fig. 4. Network planning for the wireless neighborway scheme.

transmission resource from the core to the aggregation layer can be utilized, therefore; it could be more efficient if the two parties are from different pods.

With ***Rule 1***, the sending switch and its surrounding neighbors, as well as the receiving switch and its surrounding neighbors are in different pods, respectively. With ***Rule 2***, each pair of offloading ToRs on the sender and receiver side are in different pods.

B. Network Planning and IP Addresses Assignment

We here present the network planning solution and corresponding IP addresses assignment, which can map the logical topology of the Fat-Tree into the physical layout, with obeying the two rules above. Suppose we use k-port switches to construct a Fat-Tree structured DCN, there will be k pods, each of which contains two layers of $\frac{k}{2}$ switches. Each k-port switch in the edge layer directly connects to $\frac{k}{2}$ servers, and each of the remaining $\frac{k}{2}$ ports is connected to the $\frac{k}{2}$ of k ports in the aggregation layer of the hierarchy. Figure 1 gives an example of the Fat-Tree DCN with $k = 4$. The wireless transceivers are installed on top of the ToR switches, thus each ToR switch can be considered as a wireless node in the overhead view as shown in Fig. 4.

The Fat-Tree structured DCN uses the $10.0.0.0/8$ block of IP addresses. The switches in the pod are given addresses of the form $10.pod.switch.1$, where the pod denotes the pod number with the range $[0, k-1]$, and the $switch$ denotes the position of the switch with the range $[0, \frac{k}{2}-1]$ in the pod starting from left to right and bottom to top.

We lay out those racks in $\frac{k}{2}$ rows and k columns geographically as shown in Fig. 4. The IP address of each ToR will be assigned by the algorithm below, which ensures that if the sending ToR switch and the receiving ToR switch are different, the two rules above must be satisfied.

According to our algorithm, the switch at row m and column 0 should be assigned the pod number $(3m)\%k$, where $0 \leq m \leq \frac{k}{2}-1$. The switch at row m and column n, denoted as (m, n), should be assigned the pod number $((3m)\%k+n)\%k$, where $0 \leq n \leq k-1$. For neighboring switches around the switch at (m, n), the addresses assigned should be:

- $(m-1, n)$: $((3m)\%k+n)\%k-3$;
- $(m+1, n)$: $((3m)\%k+n)\%k+3$;
- $(m, n-1)$: $((3m)\%k+n)\%k-1$;

Algorithm 1: IP address assignment for ToR switches

Data: Layout of racks
Result: IP addresses assignment of each ToR switch
for $row \leftarrow 0$ **to** $\frac{k}{2}-1$ **do**
  pod = $(3 \times row)\%k$;
  for $column \leftarrow 0$ **to** $k-1$ **do**
    SetIPAddress: $10.pod.row.1$;
    pod = $(pod+1)\%k$;
  end
end

- $(m, n+1)$: $((3m)\%k+n)\%k+1$.

Since $m \leq \frac{k}{2}-1$ and $n \leq k-1$, any 5-switch cluster will be assigned in different pod numbers, which means that Rule 1 must be satisfied. Consequently, if the sending ToR switch and the receiving ToR switch are different, the Rule 2 must be satisfied.

C. Optimizing the Network Planning

In the design above, the racks of the DCN are laid in $\frac{k}{2}$ rows and k columns geographically; however, the racks may need to be laid in different ways due to the limitation of the space accommodating the DCN. Moreover, we may come across the special case that the sending and the receiving switch are in the same pod, which makes the four neighboring switches around the sending switch and the corresponding switches around the receiving switch are in the same pod. Such cases can degrade the performance of the neighborway scheme.

We here define the *offloading factor*

$$\alpha = n_s \cdot n_d - n, \tag{1}$$

to measure the performance of neighborway offloading scheme on one sending-receiving switch pair. n_s and n_d are the numbers of neighboring switches around the sending and receiving switch following rule 1 which are in different pods, respectively; n is the number of overlapped pods accommodating switches around the sending and receiving switch. For example, if the switches around the sending one are in pods 1, 2, 3, 4, and the switches around the receiving one are in pods 1, 2, 5, 6, respectively, then $n = 2$. For a given sending-receiving switch pair, the number of offloading paths is $n_s \cdot n_d$, but the offloading pairs should not be in the same pod according to rule 2, consequently, α is actually the number of effective offloading paths for such pair, which indicates to what extent this pair could benefit from neighborway offloading.

In order to maximize the utility of the neighborway scheme, we now develop an optimization scheme to maximize the offloading factor of switch pairs in the entire DCN. Suppose the $\frac{k^2}{2}$ edge-switches is laid out in an $a \times b$ grid, where $a \times b = \frac{k^2}{2}$ and a and b denote the row and column index in a the DCN, respectively. We use $p_{x,y}$ to represent the pod the ToR in the position (x,y) belongs to. According to the fat-tree topology, $0 \leq p_{x,y} \leq k-1$. We use $\Delta_{x,y}$ to represent the amount of ToRs in different pods around the ToR in the position (x,y). We define an indicator $d_{a,b}$, which is set to 0 if $p_{a,b} = p_{x,y}$ and 1 if $p_{a,b} \neq p_{x,y}$, thus

$$\Delta_{x,y} \triangleq d_{(x-1),y} + d_{(x+1),y} + d_{x,(y-1)} + d_{x,(y+1)}. \quad (2)$$

We use ϕ_{x_s,y_s,x_d,y_d} to denote the number of overlapped pods accommodating switches around (x_s, y_s) and around (x_d, y_d). We define another indicator $s_{a,b,c,d}$, which is set to 0 if $p_{a,b} \neq p_{c,d}$ and 1 if $p_{a,b} = p_{c,d}$, and

$$\begin{aligned} f_{a,b,x_d,y_d} =& s_{a,b,x_d+1,y_d} + s_{a,b,x_d-1,y_d} \\ &+ s_{a,b,x_d,y_d+1} + s_{a,b,x_d,y_d-1}, \end{aligned} \quad (3)$$

thus

$$\begin{aligned} \phi_{x_s,y_s,x_d,y_d} &= f_{x_s-1,y_s,x_d,y_d} + f_{x_s+1,y_s,x_d,y_d} \\ &+ f_{x_s,y_s-1,x_d,y_d} + f_{x_s,y_s+1,x_d,y_d}. \end{aligned} \quad (4)$$

We define $l_{x,y,C}$:

$$l_{x,y,C} = \begin{cases} 1, & \text{if } p_{x,y} = C, \\ 0, & \text{if } p_{x,y} \neq C, \end{cases}$$

where C is the pod number. The problem of maximize the effectiveness of neighborway scheme now is transformed into the following optimization problem. The objective function is actually the sum of *offload factors* of all the pairs in the DCN:

$$max \sum_{s_i=1}^{a} \sum_{s_j=1}^{b} \sum_{d_i=1}^{a} \sum_{\substack{d_j=1 \\ d_i,d_j \neq s_i,s_j}}^{b} (\Delta_{s_i,s_j} \cdot \Delta_{d_i,d_j} - \phi_{s_i,s_j,d_i,d_j}), \quad (5)$$

$$s.t. \qquad 0 \leq p_{i,j} \leq k-1, \quad \forall i \in [0,a], j \in [0,b] \quad (6)$$

$$\sum_{i=1}^{a} \sum_{j=1}^{b} l_{i,j,C} = \frac{k}{2}, \quad \forall\ C \in [0, k-1] \quad (7)$$

Constraint (6) represents in k-port fat-tree DCN, there are k pods in all, ranging from 0 to $k-1$. (7) represents that each pod has $k/2$ edge switches. The solution of this problem is the assignment of the pod number of each ToR. This problem is solvable because there are finite assignment in a certain layout, thus a optimized assignment of pod number exists.

D. Traffic Engineering

After network planning and IP address allocation, we now study how to perform the traffic engineering in the hybrid wired/wireless DCN so that the hotspots can be relieved. We consider the time in the system is slotted, and we want to reveal how to perform the traffic engineering over all wired/wireless links.

Suppose a given edge-level switch u is processing $\sum_{i=1}^{p} F_{st}(v_i, u) + \sum_{i=1}^{q} F_{st}(u, v_i)$ amount of wired data in each time slot, and the first item represents for the wired input-data while the second item represents for the wired output-data. $F_{st}(u, v)$ denotes a wired flow from switch u to switch v, and the original source of this flow is s while the final destination is t. p denotes the amount of the switches transmit data to switch u and q denotes the amount of switches that switch u transmit data to. Thus, we can define the *cache occupancy rate* of an edge-level switch as

$$\beta \quad \triangleq \quad \beta' + \frac{F_{in} - F_{out}}{T} \quad (8)$$

where T is the maximum amount of data a switch can buffer at one time slot and β' is the *cache occupancy rate* of the switch in the prior time slot. $F_{in} - F_{out}$ is the amount of the data that cannot be processed timely in one time slot, which will increase the cache occupancy rate. Note that the cache occupancy rate should be no greater than 1. If F_{in} is extremely high in a time slot and cause the value of β to exceed 1, it would incur buffer overflow and packet loss. The cached data could be transmitted in the next time slot, and the amount of data can be denoted as $\beta' T$. Thus, $\frac{F_{in} - F_{out}}{T}$ is the rate of change of the cache occupancy. (8) denotes the case that there is no wireless offloading scheme.

Now we take the wireless neighborway scheme into consideration. Consider a certain ToR switch with high *cache occupancy rate* β'_s in the prior time slot. Among its four neighboring ToRs, there are m ToRs' *cache occupancy rate* is lower than β'_s, where m is an integer and $1 \leq m \leq 4$. Each cache occupancy rate of such ToR is denoted as β'_i where $1 \leq i \leq m$. We use the following equation set to calculate the traffic load f_i in each wireless link:

$$\begin{cases} f_i \propto (1 - \beta'_i), \\ \sum\limits_{i=1}^{m} f_i = (\beta'_s - \frac{1}{m+1}(\sum\limits_{i=1}^{m} \beta'_i + \beta'_s))T. \end{cases} \quad (9)$$

while the updating formula for β is:

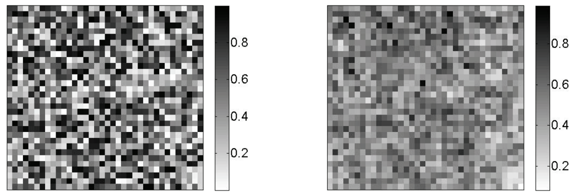
$$\beta_s = \beta'_s + \frac{F_{in} - F_{out} - \sum\limits_{i=1}^{m} f_i}{T}, \quad (10)$$

$$\beta_i = \beta'_i + \frac{F_{in} - F_{out} + f_i}{T}. \quad (11)$$

Note that for a certain ToR switch, it only transmits data through wireless links to those ToRs with lower cache occupancy rate than itself. Thus, after calculating such process for each ToR in the DCN, the wireless traffic load is assigned. Note that the traffic load on one wireless link is always calculated once, and this flow is from the high-β ToR to the low-β ToR.

(9) is a equation set to calculate how much traffic for a certain ToR to transmit to its m neighboring switches. $\frac{1}{m} \sum_{i=1}^{m} \beta'_i$ is the average cache occupancy rate of these switches, and the total data $\sum_{i=1}^{m} f_i$ transmitted from a certain ToR is to make its cache occupancy rate close to this average level.

When a high-β ToR transmits data to its offloading ToRs with wireless links, it could also be one of the offloading receiver of a neighboring ToR switch with higher β. According to proportional relationship in (9), the traffic load assigned between these two high-β ToRs will be small, which would not increase their traffic burden. The traffic load assignment



(a) Scenario without neighborways. (b) Scenario with neighborways.

Fig. 5. Hotspots map in different scenarios.

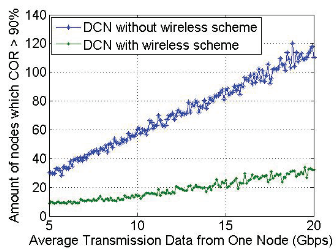


Fig. 6. Num of hotspots in different scenarios.

is based on the β' in the prior time slot and is updated by the traffic load according to (10) and (11).

V. Simulation Results

We use Matlab to evaluate the performance of our proposed scheme with comparison to the all wired DCN scenario. We create a Fat-Tree structured DCN with the 48-port switches, which consists of 1152 ToRs. A switch can handle at most 20Gb in one second, and wireless link capacity is 4Gbps. The source ToR and destination ToR are randomly selected. We measure *cache occupancy rate* of all switches in the all wired link scenario and that in the neighborway scenario, then the hotspots map is obtained in Fig.5. Each pixel in the graph represents a switch. The darker the pixel's color is, the higher cache occupancy rate the switch has.

In Fig. 5(a), the workloads of switches are seriously unbalanced. The effectiveness of neighborways are illustrated in Fig. 5(b) where more nodes are grey, which means that the traffic loads have been balanced, and the under-utilized network resources are exploited to relieve the hotspots.

Then we randomly choose sending switches, and assign output traffic from the sending node in the range between 5Gbps and 20Gbps. We count the number of switches whose cache occupancy rate is over 90% shown in figure 6. It is clear that the neighborway scheme can significantly reduce the number of hotspots in the DCN.

VI. Conclusion

In this paper, we have proposed to integrate wireless links into the DCN with the *wireless neighborway* scheme. Each top-of-rack switch (ToR) has multiple 60GHz wireless links connecting to its neighboring ToRs, which are termed as neighborways. The fundamental challenge for the design is how to prevent the offloaded traffics from forming new hotspots in the network fabric. To this end, we have developed a wireless network planning solution with corresponding IP address assignment scheme and traffic engineering scheme, which can leverage the potentially under-utilized wired links in the DCN and meanwhile avoid forming new hotspots. The simulation results show that the proposed scheme can dramatically relieve the hotspots in the DCN.

VII. Acknowledgements

This work is supported by NSF China (61202373, 61102051); SRF for ROCS by SEM; Shanghai Basic Research Key Project (No.13510711300, 12JC1405200, 11JC1405100); Open Foundation of State Key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) (No.SKLNST-2013-1-16). Yu Cheng was supported in part by the NSF under grant CNS-1053777.

References

[1] M. Al-Fares, A. Loukissas and A. Vahdat, "A Scalable, Commodity Data Center Network Architecture," in *Proc. ACM SIGCOMM*, Aug. 2008, pp.63–74.

[2] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang and S. Lu, "DCell: A Scalable and Fault-Tolerant Network Structure for Data Centers," in *Proc. ACM SIGCOMM*, Aug. 2008, pp.75–86.

[3] A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel and S. Sengupta, "VL2: A Scalable and Flexible Data Center Network," in *Proc. ACM SIGCOMM*, Aug. 2009, pp. 51–62.

[4] Y. Cui, H. Wang and X. Cheng "Channel Allocation in Wireless Data Center Networks," in *Proc. IEEE INFOCOM*, 2011, pp.1395–1403.

[5] Y. Cui, H. Wang and X. Cheng, "Wireless data center networking, " *IEEE Wireless Commun. Mag.*, pp.46–53, Jun. 2011.

[6] P. Smulders, "Exploiting the 60 GHz Band for Local Wireless Multimedia Access: Prospects and Future Directions," *IEEE Commun. Mag.*, pp. 140–147, Jan. 2002.

[7] T. Benson, A. Anand, A. Akella and D. A. Maltz, "Network Traffic Characteristics of Data Centers in the Wild," in *Proc. IMC*, pp. 267–280, Nov. 2010.

[8] S. Kandula, J. Padhye and P. Bahl, "Flyways To De-Congest Data Center Networks," *ACM HotNets*, Nov. 2009.

[9] D. Halperin, S. Kandula, J. Padhye, P. Bahl and D. Wetherall, "Augmenting Data Center Networks with Multi-Gigabit Wireless Links," in *Proc. ACM SIGCOMM*, Aug. 2011, pp. 38–49.

[10] Y. Katayama, T. Yamane, Y. Kohda, K. Takano, D. Nakano and N. Ohba, "MIMO Link Design Strategy for Wireless Data Center Applications," in *Proc. IEEE WCNC*, Apr. 2012, pp.3302–3306.

[11] X. Zhou, Z. Zhang, Y. Zhu, Y. Li, S. Kumar, A. Vahdat, B. Zhao and H. Zheng, "Mirror Mirror on the Ceiling: Flexible Wireless Links for Data Centers," in *Proc. ACM SIGCOMM*, Aug. 2012, pp.443–454.

[12] N. Farrignton, G. Porter, S. Radhakrishnan, H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen and A. Vahdat, "Helios: A Hybrid Electrical/Optical Switch Architecture for Modular Data Centers," in *Proc. ACM SIGCOMM*, Aug. 2010, pp.443–454.

[13] J. Pujol, V. Erramilli, G. Siganos, X. Yang, N. Laoutaris, P. Chhabra and P. Rodriguez, "The Little Engine(s) That Could: Scaling Online Social Networks," in *Proc. ACM SIGCOMM*, Aug. 2010, pp.375–386.

[14] J. Shin, E. G. Sirer, H. Weatherspoon, D. Kirovski "On the Feasibility of Completely Wireless Data Centers, " in *Proceedings of the eighth ACM/IEEE symposium on Architectures for networking and communications systems*, 2012, pp. 3–14.

[15] K. Ranachandran . "60GHz data-center networking: wireless = worryless, " Tech. Rep. , NEC Laboratories America, Inc. , July, 2008.

[16] Y. Katayama, K. Takano, Y. Kohda , N. Ohba, D. Nakano "Wireless data center networking with steered-beam mmwave links.," in *Proc. IEEE WCNC,* Mar. 2011, pp. 2179-2184.

[17] Xinbing Wang, L. Fu, C. Hu, "Multicast Performance with Hierarchical Cooperation," *IEEE/ACM Trans. on Networking*, vol 20, no 3, pp. 917-930, 2012.

[18] Xinbing Wang, W. Huang, S. Wang, J. Zhang, C. Hu, "Delay and Capacity Tradeoff Analysis for MotionCast," *IEEE/ACM Trans. on Networking*, Vol. 19, no. 5, pp. 1354-1367, Oct 2011.

[19] Y. Cui, H. Wang and X. Cheng, "Channel Allocation in Wireless Data Center Networks," *in Proc. IEEE INFOCOM*, 2011, pp. 1395-1403.