Data Center Network Structure using Hybrid Optoelectronic Routers

Yuichi Ohsita, and Masayuki Murata
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
{y-ohsita, murata}@ist.osaka-u.ac.jp

***Abstract*—Large data centers hosting hundreds of thousands of servers have been built to handle huge amounts of data. In a large data center, servers cooperate with each other, and thus the data center network must accommodate a large amount of traffic between servers. However, the energy consumption of large data center networks is a major problem because several large-capacity switches or many conventional switches are required to handle the traffic. A low-energy hybrid optoelectronic router has been proposed to provide high bandwidth between data center servers. The hybrid optoelectronic router has an optical packet switching functionality and packet buffering in the electronic domain. The optical packet switching provides large bandwidth communication between the optical ports. In addition, the router can be directly connected to server racks by its electronic port. The packets from the server racks are stored in an electronic buffer in the router. Then, the packets are converted into optical packets and sent to another router. Finally, when the packets arrive at the router connected to the destination server rack, they are stored in the buffer, converted to electronic packets, and sent to the destination server rack. In this paper, we discuss a data center network structure that contains hybrid optoelectronic routers. We propose a method for constructing a data center network that uses hybrid optoelectronic routers efficiently. Furthermore, we discuss the effect of the network structure on the number of routers required to accommodate a large amount of traffic in a mega data center.**

Keywords*—Data Center Network; Topology; Optical Packet Switch

I. Introduction

Large data centers with tens and even hundreds of thousands of servers have been built to handle the vast amount of data generated by various online applications. Data center servers communicate with each other to handle the data. A lack of bandwidth or large delay prevents communication between servers and increases the time taken to retrieve data. This degrades the performance of the data center.

The energy consumption of data centers, which increases with the data center size, is another major problem, and the energy consumption of the network is a substantial proportion of total energy usage [1]. Therefore, data center networks with high energy efficiency and high communication performance are required [2].

Optical networking is a promising approach to constructing networks with high energy efficiency [3]. Optical network devices provide low latency communication with low energy consumption because they relay optical signals without conversion to electrical signals. Optical networking also provides high bandwidth via technologies such as wavelength division multiplexing (WDM).

An optical packet switch architecture called the *hybrid optoelectronic router* was proposed for data centers by Ibrahim et al. [4]. The hybrid optoelectronic router has optical packet switching functionality and packet buffering in the electronic domain. The optical packet switching functionality provides large bandwidth communication between the optical ports by relaying the optical packets without conversion to electronic signals unless packet collision occurs. Even if a collision occurs, the router can retransfer the packets after storing them in the electronic buffer.

In addition, the hybrid optoelectronic router can be connected directly to server racks via its electronic port. The packets from the server racks are stored in an electronic buffer in the router, and then converted into optical packets and sent to another router. Finally, if the packets arrive at the router connected to the destination server rack, the packets are stored in the buffer, converted to electronic packets, and sent to the destination server rack.

The network structure is important for constructing a large data center that uses hybrid optoelectronic routers and should use the large bandwidth of the hybrid optoelectronic routers efficiently. However, if each server rack can be connected to multiple routers, the connection from the server racks may have a large effect on the network performance. We have proposed a network structure that uses optical packet switches and multiple connections from the server racks [5]. However, our previous work used multiple connections from the server racks only to provide connectivity when optical packet switches fail, and did not discuss the effect of using multiple connections from the server racks on the performance of the data center network.

In this paper, we propose a method to construct a data center network structure that accommodates a large amount of traffic by using the hybrid optoelectronic routers and multiple connections from the server racks efficiently. We evaluate our network structure and demonstrate that our method can accommodate more traffic than a torus network. In addition, we discuss the importance of using multiple links from the server racks, and show that multiple links are necessary to construct a large data center with sufficient bandwidth between server rack pairs.

The rest of this paper is organized as follows. Section II explains related work about data center networks using optical network technologies. Section III provides an overview of hybrid optoelectronic routers and the data center network that uses hybrid optoelectronic routers. Section IV proposes a method to construct data center network structures using hybrid optoelectronic routers. Section V discusses building a data center network that can accommodate more servers without decreasing bandwidth based on our method of constructing data center network structures. Finally, we conclude this paper in Section VI.

II. Related work

Farrington et al. proposed a data center network architecture that uses optical path switches [6]. In this network, the optical path switches are placed at the core of the data center network and are configured to connect the server rack pairs that are generating a large amount of traffic. Similar architecture using optical circuit switches was proposed by Wang [7]. However, the configuration of the optical path switches takes time, and this architecture cannot handle frequent traffic changes that often occur in a data center [8].

Another approach is to use optical packet switches [4], [9]–[13]. Optical packet switches contain arrayed waveguide grating routers (AWGRs) and wavelength converters. Because the output port of the input signal depends on the wavelength of the input signal in AWGRs, the destination port is changed by changing the wavelength. Optical packet switches relay optical packets based on the optical labels attached to the packets. Because optical packet switches do not require the establishment of paths, a network constructed of optical packet switches can handle frequent changes in traffic.

Optical packet switches with a large number of ports have also been constructed by connecting optical packet switches with a small number of ports. Xi et al. constructed an optical packet switch with a large number of ports by connecting the optical switches in a Clos topology [12]. Liboiron-Ladouceur et al. built a large data center by connecting the optical switches in a Tree topology [14]. However, they considered only the connections between optical switches. In a data center, servers may have electronic ports instead of optical ports. Therefore, we need to consider the connection from the servers or server racks that have electronic ports.

An optical packet switch architecture proposed by Ibrahim et al. [4], called a hybrid optoelectronic router, has electronic ports that can be connected directly to the servers. By using the connections between optical switches, and connections between optical switches and server racks, the data center network can accommodate a large amount of traffic between a large number of servers. However, a data center network that efficiently uses the both of these connections has not been reported. Therefore, we discuss a data center network structure that accommodates a large amount of traffic by efficiently using hybrid optoelectronic routers.

III. Data center networks with hybrid optoelectronic routers

In this section, we introduce the hybrid optoelectronic router, and the data center network constructed of them.

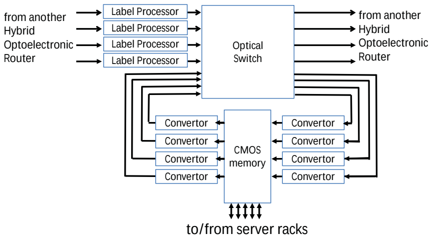

Figure 1. Hybrid optoelectronic router.

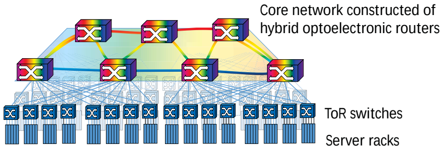

Figure 2. Network with hybrid optoelectronic routers in a data center.

A. Hybrid optoelectronic routers

Figure 1 shows a hybrid optoelectronic router. Each hybrid optoelectronic router has ports connected to other hybrid optoelectronic routers (hereafter called *optical ports*), and ports connected to the server rack (hereafter called *electronic ports*).

When an optical packet arrives through the optical port, the label processors identify its label and destination, and the controller controls the switching fabric to relay the packet to the destination. If the destination port is busy, the packet is stored in the electronic buffer after the optical packet is converted to an electronic packet. The packet is re-sent after it is converted into an optical packet.

Electronic packets from a server rack arrive through the electronic port. The electronic packet is buffered and converted into an optical packet. The converted packet is relayed in the same way as optical packets from the other optical packet switches. Optical packets sent to a server rack are also buffered and converted to electronic packets that are relayed to the server rack.

B. Data center network using hybrid optoelectronic routers

Optical networks containing hybrid optoelectronic routers provide large bandwidth communication. Therefore, we place an optical network containing hybrid optoelectronic routers at the core of the data center network. Each server rack has an ToR switch and is connected to the core network by connecting its electronic ToR switch to multiple hybrid optoelectronic routers. Figure 2 shows an example of a data center network containing hybrid optoelectronic routers.

In this network structures, the connections between the hybrid optoelectronic routers, and between the server racks and the hybrid optoelectronic routers must be set, which is discussed in Section IV.

C. Routing in the data center network by using hybrid optoelectronic routers

In the data center, the traffic changes frequently. If the routes are controlled by a central controller, the central controller has to collect the traffic information about the whole network frequently, which requires a large overhead.

Therefore, we use routing that balances the load among the shortest paths by using only the local information. Each hybrid optoelectronic router has a routing table, which includes multiple next hybrid optoelectronic routers on the shortest paths to each destination hybrid optoelectronic router. When a hybrid optoelectronic router receives a packet, the hybrid optoelectronic router selects the next router with the smallest load from the routing table.

When a ToR switch sends a packet to the core hybrid optoelectronic router network, the ToR switch encapsulates the packet by attaching the destination hybrid optoelectronic router. The packet is sent to one of the hybrid optoelectronic routers connected to the ToR switch. The first and destination router pair is selected from candidate pairs with the smallest number of hops for the first router connected to the source server rack and the destination hybrid router connected to the destination server. If there are multiple candidate pairs, the pair is selected randomly to balance the loads.

IV. Construction of the network structure

In this section, we propose a method for constructing the data center network structure by using hybrid optoelectronic routers. The server racks are grouped, and the racks in each group are connected to the same hybrid optoelectronic routers. Each server rack is connected to R hybrid optoelectronic routers, and each hybrid optoelectronic router is connected to S groups of server racks. We construct a network containing S^R groups of server racks and RS^{R-1} hybrid optoelectronic routers.

First, we connect the server racks to hybrid optoelectronic routers, and then set the connections between the routers.

A. Connection from server racks

We aim to maximize the amount of traffic that can be accommodated. A large number of hops between server racks decreases the amount of traffic that can be handled because hops consume bandwidths for a large number of links.

We connect server racks to hybrid optoelectronic routers to maximize the number of server racks that can communicate through only one hybrid optoelectronic router. We use a network architecture similar to the BCube [15].

We separate hybrid optoelectronic routers into R layers, and each server rack is connected to one hybrid optoelectronic router in each layer. To determine the connections between server racks and hybrid optoelectronic routers, we set the ID to the server rack group. Similarly, we set the ID to the hybrid optoelectronic routers in each layer.

We connect the ith hybrid optoelectronic router at the rth layer to the jth server rack when $\lfloor \frac{i}{S^{r-1}} \rfloor = \lfloor \frac{j}{S^r} \rfloor$ and $i \bmod S^{r-1} = j \bmod S^{r-1}$ are satisfied. By doing so, the server rack groups sharing hybrid optoelectronic routers are different for the different layers.

B. Connections between hybrid optoelectronic routers

After connecting the hybrid optoelectronic routers, we construct the connections between hybrid optoelectronic routers. To search for the best connections, we generate and select candidate connections.

1) Candidate connections: The number of available connections between hybrid optoelectronic routers is ${}_{(RS^{R-1})^2}C_{PRS^{R-1}}$, where P is the number of optical ports of each hybrid optoelectronic router. The number of available connections is too large, and it takes a long time to select the best one from all available connections. Therefore, we focus on candidates for which the number of hops between server racks is small, because networks where the number of hops is large cannot accommodate a large number of hops because the hops waste the bandwidth of many links. In addition, we focus on candidates where all hybrid optoelectronic routers play the same role, because if there are hybrid optoelectronic routers that play a special role, such as the root node of the tree topology, the loads on the special routers becomes large.

We construct the candidates through the following steps.

1) Set the network structure, where all server racks are connected to the hybrid optoelectronic routers but no links between hybrid optoelectronic router are constructed, as a candidate.
2) Add one link per hybrid optoelectronic router for each candidate if the candidate has an empty optical port. If no candidate has an empty optical port, end the process.
3) Select N candidates based on the number of hops between server racks.
4) Go to step 2.

In Step 2, we add links to each hybrid optoelectronic router so that all hybrid optoelectronic routers play the same role. After adding one link from the first hybrid optoelectronic router, we add links that have the same properties as the first link. In this paper, we regard the links satisfying the following constraints as links with the same properties.

- The source router for the link is included in the same layer.
- The destination router for the link is included in the same layer.
- The number of hops from the source router to the destination router on the network before adding the links is the same.

In Step 3, we select N candidates based on the number of hops between server racks. In this paper, we select the candidates with the largest number of hops between the server racks. If multiple candidates have the same largest number of hops. we compare the average number of hops between all server rack pairs.

Algorithm 1 shows the pseudo code for generating N candidates. In this pseudo code, R is the set of the hybrid optoelectronic routers, c^{init} is the initial candidate network topology where all server racks are connected to the hybrid optoelectronic routers, but no links between hybrid optoelectronic routers are constructed. In each iteration from Line 3 to 35, we update the list of the candidates C by adding one optical link per hybrid optoelectronic router. To add one link per hybrid optoelectronic router, we first decide the router pair where the new link is added at Line 10. Then, for each router, we find target routers that have the same properties as the link added at Line 10, and add links between found route pairs from Line 12 to 23. At the end of each iteration, we save only N candidates from Line 33 to 34.

In these steps, we continue the iteration from Line 3 to 35 P times. In each iteration, we generate and evaluate at most $N(RS^{R-1})^2$ where RS^{R-1} is the number of hybrid optoelectronic routers. That is, to generate the candidates, we evaluate at most $PN(RS^{R-1})^2$ candidates. The number of generated and evaluated candidates becomes large as the number of hybrid optoelectronic routers becomes large. However, the number of hybrid optoelectronic routers is much smaller than the number of server racks. In addition, the candidates can be evaluated in parallel. Moreover, the candidate topologies are generated only once before constructing a data center. Therefore, we believe that the calculation time for generating the candidate topologies should not be a major problem.

2) Selection of the best candidate: Finally, we select the best candidate from the generated candidates. We simulate the routing in the data center when traffic is generated between all server rack pairs. Next, we select the candidate with the smallest link utilization to construct the network structure that can accommodate the largest amount of traffic. When constructing the data center network, the traffic within a data center is unknown. Thus. we set the traffic between all server rack pairs so that traffic between all server rack pairs equal.

Algorithm 1 Generation of N candidates for connections between hybrid optoelectronic routers.

```
 1: Clear the list of candidates, $C$.
 2: Add $c^{\mathrm{init}}$ to $C$.
 3: while True do
 4:   Clear the list of newly generated candidates, $C'$
 5:   for $c \in C$ do
 6:     Select the router, $r_1 \in R$, with the largest number of
        remaining ports.
 7:     for $r_2 \in R$ do
 8:       if $r_2$ has remaining ports in $c$ then
 9:         Clear the list of temporal candidate $C^{\mathrm{tmp}}$
10:         Construct a new candidate, $c^{\mathrm{new}}$, by adding link
            between $r_1$ and $r_2$ to $c$.
11:         All $c^{\mathrm{new}}$ to $C^{\mathrm{tmp}}$
12:         for $r_3 \in R$ do
13:           Clear the list of temporal candidate $C^{'\mathrm{tmp}}$
14:           for $c^{\mathrm{tmp}} \in C^{\mathrm{tmp}}$ do
15:             for $r_4 \in R$ do
16:               if $r_4$ has the remaining ports in the
                  network, $c^{\mathrm{tmp}}$ then
17:                 if link $r_3, r_4$ has the same property as
                    link $r_1, r_2$ then
18:                   Construct a new candidate, $c^{'\mathrm{new}}$,
                      by adding a link between $r_3$ and $r_4$
                      to $c^{\mathrm{tmp}}$.
19:                   Add $c^{'\mathrm{new}}$ to $C^{'\mathrm{tmp}}$
20:                 end if
21:               end if
22:             end for
23:           end for
24:           $C^{\mathrm{tmp}} \leftarrow C^{'\mathrm{tmp}}$
25:         end for
26:         Add all candidates $c^{\mathrm{tmp}} \in C^{\mathrm{tmp}}$ to $C'$
27:       end if
28:     end for
29:   end for
30:   if $C'$ is empty then
31:     return $C$
32:   end if
33:   Constructing the list of candidates $C''$ by selecting $N$
      candidates from $C'$
34:   $C \leftarrow C''$
35: end while
```

C. Incremental construction

Constructing a data center network with a large number of server racks simultaneously is difficult. The data center should be incrementally constructed by adding server racks and routers. We propose a method to construct the data center network structure incrementally. We first calculate a suitable network structure including the maximum number of server racks and hybrid optoelectronic routers. Hereafter, we call this calculated network structure the largest network structure. Next, we construct the subset of the largest network structure that can connect the currently required number of server racks. When it is necessary to add more servers to the current network, we find the best network structure that includes the current routers and server racks and is a subset of the largest network structure. In the rest of this subsection, we explain the steps to finding the best network structure in detail.

1) Construction of the network structure when the number of the required server rack groups is given: We construct a network structure that can connect the number of the required server rack groups. The suitable network structure is calculated by generating candidates for the network structure that is a subset of the largest network structure, and selecting one of them. The candidates are generated by the following steps.

1) Construct the initial network. If there is a current working network, the current network is set as the initial network. Otherwise, select the server rack group with the smallest ID, and construct the initial network including the selected server rack group and all hybrid optoelectronic routers to which the groups are connected.
2) Add the constructed initial network to the *list of incom-*

plete candidates.

3) For each network in the list of incomplete candidates, generate the new networks by adding one hybrid optoelectronic router that is not included in the network, but has a link to one router included in the network in the largest network structure, and by adding links between hybrid optoelectronic routers included in the new network. Then, save the new network in the *list of newly constructed candidates*.
4) For each network in the list of newly constructed candidates, count the number of server rack groups connected to the hybrid optoelectronic router. If the number of server racks is more than the number of required server rack groups, add the network to the *list of candidates*.
5) If the list of candidates includes more than one candidate, end the process. Otherwise, replace the list of incomplete candidates with the list of newly constructed candidates, and go back to step 3.

In Step 1. we select the server rack group with the smallest ID, because all server rack groups play the same role in the largest network structure.

In these steps, we look for a network structure that can connect the required number of server rack groups by adding one hybrid optoelectronic router during each iteration. We end the process if at least one candidate is found. As a result, we can find candidates that can connect the required number of server racks with the smallest number of hybrid optoelectronic routers. The pseudo code for these steps is shown in Algorithm 2.

Next, we select the candidate that can accommodate the largest amount traffic through simulating the routing, similar to Section IV-B2.

Finally, we add optical links between hybrid optoelectronic routers or electronic links between server rack groups and hybrid optoelectronic routers if there are unused ports. By adding links, we can accommodate more traffic. In this paper, we generate all patterns of added links and select the best pattern to accommodate the largest amount of traffic through the routing simulation.

2) Construction of the network structure when the number of the required server racks is given: The network structure should be constructed to accommodate any possible traffic rate. To guarantee this, we use valiant load balancing (VLB) [16]. In VLB, we select the intermediate nodes randomly regardless of the destination to avoid concentrating traffic on certain links, even when traffic volume of certain node pairs is large. In the data center network, the intermediate node is selected from the server racks. The packet is encapsulated by attaching a header whose destination is the selected intermediate server rack. When the intermediate server rack receives the packet, it relays the packet to the final destination after removing the header.

By applying VLB, the traffic rate between server rack T satisfies the following inequality.

$$T \leq \frac{2B}{N}$$

Here, B is the bandwidth from a server rack and N is the number of server racks. That is, we construct a network structure that can accommodate the flow of data with $\frac{2B}{S}$ between all server rack pairs.

Algorithm 2 Construction of candidate subnetworks of the largest network structure.

```
Clear $C^{\text{incomplete}}$.
Add the initial candidate, $c^{\text{init}}$, to $C^{\text{incomplete}}$.
while True do
  Clear $C^{'\text{incomplete}}$
  for $c \in C^{\text{incomplete}}$ do
    for $r \in R$ do
      if $r$ is not included in $c$ but has a link to a router included in $c$ then
        Generate new candidate $c^{\text{new}}$ by adding the router, $r$, and the server racks connected to $r$ to $c$.
        Add $c^{\text{new}}$ to $C^{'\text{incomplete}}$
      end if
    end for
  end for
  Clear $C^{\text{complete}}$
  for $c \in C^{'\text{incomplete}}$ do
    if $c$ include more than the required number of server rack groups then
      Add $c$ to $C^{\text{complete}}$
    end if
  end for
  if $C^{\text{complete}}$ is not empty then
    return $C^{\text{complete}}$
  end if
  $C^{\text{incomplete}} \leftarrow C^{'\text{incomplete}}$
end while
```

Based on this, when the number of required server racks, N, is given, we construct the data center network that can accommodate N server racks by the following steps.

1) Set the number of server rack groups S to the number of the server rack groups included in the current network. If the data center network is newly constructed, set S to 1.
2) Set the number of server racks in each server rack group to $\frac{N}{S}$.
3) Construct a network that can accommodate S server rack groups by the steps in IV-C1.
4) Check whether the constructed network can accommodate the traffic $\frac{2B}{N}$ between all server rack pairs. If yes, designate the current network as a suitable network. Otherwise, go back to step 2 after incrementing S.

D. Properties of the constructed data center network

1) Constructed network structure: First, we show the network structure constructed by our method. Figure 3 shows the network constructed from 10 hybrid optoelectronic routers with four optical ports and 25 server rack groups that have two

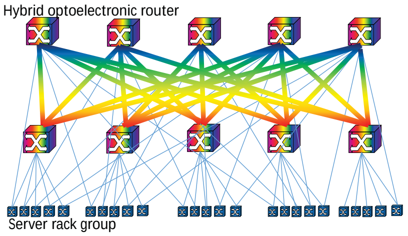

Figure 3. Example of the constructed network.

TABLE I
NUMBER OF HOPS IN THE CONSTRUCTED NETWORK STRUCTURE.

Number of hybrid optoelectronic routers on the path	Number of paths
1	192
2	288
3	96

electronic ports connected to hybrid optoelectronic routers. In this network structure, all hybrid optoelectronic router pairs can communicate without converting optical packets into electronic packets.

In this network structure, the number of hops between server racks is small (Table I). Most of the server rack pairs are connected to the same hybrid optoelectronic router or to hybrid optoelectronic routers that are directly connected to each other. Only 16 % of the server rack pairs require three hops to communicate with each other. In the network structure, the server rack pairs requiring three hops have three disjoint paths with the smallest number of hops. That is, this network structure can provide sufficient bandwidth by balancing the loads.

2) Delay: We investigate the delay between the server rack pairs through a simulation. We compare the delay in our network structure with that of the torus network constructed from the same number of hybrid optoelectronic routers and server racks [4]. We demonstrate the effectiveness of the data center network constructed considering the links from the server racks.

We model the delay for the hybrid optoelectronic router as follows. If packet collision does not occur, the hybrid optoelectronic router can relay the packet without buffering it. We assume that the delay in this case is 240 ns. If packets collide, the hybrid optoelectronic router converts the optical packet into an electronic packet, and stores it in the buffer. We assume that converting the optical packet into an electronic packet, and relaying it via the buffer takes 240 ns, similar to the case where the optical packets are relayed without collision. Storing the packet to the buffer and reading the packet from the buffer takes 180 ns. Converting the electronic packet into the optical packet and relaying it to the optical switch takes 240 ns. In addition, the packets stored in the buffer wait until the destination ports become available. We use the M/M/1 model to simulate the queuing delay in the buffer, where the average service time is 240 ns.

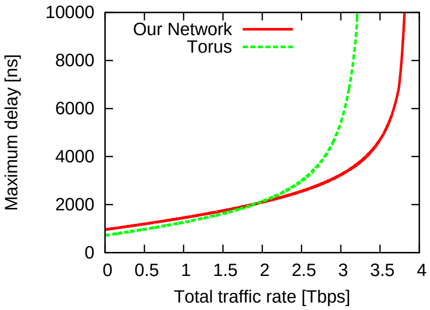

Figure 4. Comparison of the maximum delay among the server rack pairs.

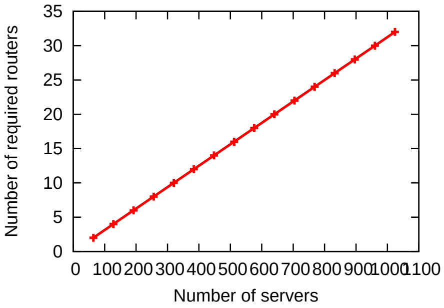

Figure 5. Number of required hybrid optoelectronic routers.

We construct both kinds of the network from 20 hybrid optoelectronic routers with four optical ports and 100 server racks with two electronic ports connected to hybrid optoelectronic routers. We generate the same traffic rate between all server rack group pairs.

Figure 4 shows the maximum delay among the server rack pairs. This figure indicates that our network structure can accommodate more traffic with a smaller delay than the torus network. This is because in our network structure, the connections between the server racks are decided considering the server rack groups connected to the hybrid optoelectronic routers.

3) Incremental construction: We construct the data center network incrementally. We set the total capacity of the electronic ports of the hybrid optoelectronic router to 32 Gbps, the bandwidth of the optical link to 100 Gbps, and the bandwidth of each server to 1 Gbps. That is, each hybrid optoelectronic router can connect 32 server racks. We set the largest network structure to the network structure constructed of 256 hybrid optoelectronic routers with four optical ports, and 32 server rack groups with two electronic ports. In the largest network structure, we connect 1024 servers without congestion.

Figure 5 shows that the number of required hybrid optoelectronic routers when we construct the network incrementally by adding 64 servers at each step. This figure indicates that we can always add 64 servers by adding only two hybrid optoelectronic routers.

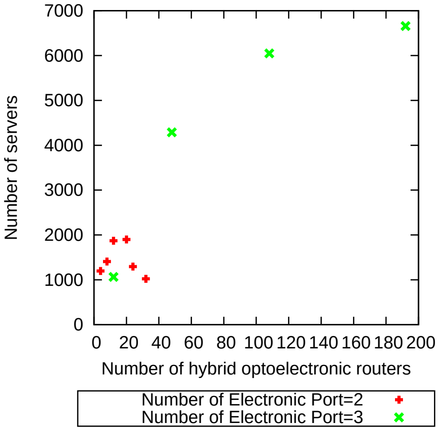

Figure 6. Number of hybrid optoelectronic routers vs number of accommodated servers (four optical ports).

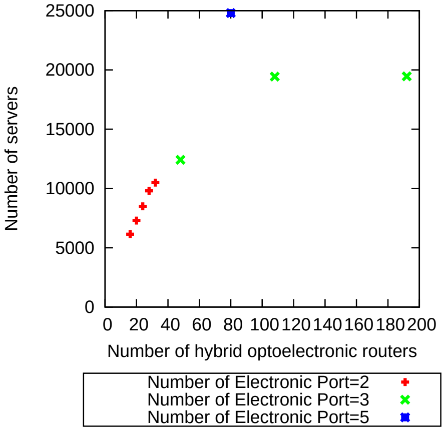

Figure 7. Number of hybrid optoelectronic routers vs number of accommodated servers (12 optical ports).

V. Discussion

In this section, we discuss the network structure that is suitable for a large data center based on our construction method. We investigate the relationship between the number of hybrid optoelectronic routers and the number of servers that can be accommodated without congestion for any traffic pattern by using VLB. In this investigation, we set the total capacity of the electronic ports of the hybrid optoelectronic router to a sufficiently large value, to focus on the bandwidth provided by the core network constructed of hybrid optoelectronic routers. We set the bandwidth of the optical link to 100 Gbps, and the bandwidth of the link from a server to 1 Gbps.

Figure 6 shows the relationship between the number of accommodated servers and the number of hybrid optoelectronic routers when each hybrid optoelectronic router has four optical ports. The figure shows network structures that include server racks with two or three electronic ports. As the number of hybrid optoelectronic routers increases, the number of optical links increases, which increases the capacity of the network and the number of accommodated servers. However, as the number of hybrid optoelectronic routers increases, the number of hops between hybrid optoelectronic routers increases. As a result, when the number of hybrid optoelectronic routers becomes sufficiently large, adding hybrid optoelectronic routers cannot greatly increase the number of servers that can be accommodated. In particular, when the number of electronic ports for each server rack is two, adding hybrid optoelectronic routers decreases the number of servers that can be accommodated. In contrast, the server racks with three electronic ports can accommodate more server racks. This is because the increase in the number of links from the server rack decreases the number of hops between server racks.

We also investigate the case where each hybrid optoelectronic router has 12 optical ports. Figure 7 shows the relationship between the number of accommodated servers and the number of hybrid optoelectronic routers when each hybrid optoelectronic router has 12 optical ports. Compared with Figure 6, the network constructed of hybrid optoelectronic routers with 12 optical ports can accommodate more servers because this network can provide more bandwidth owing to the larger number of optical links. In addition, the increase in the optical links decreases the number of hops between optical hybrid optoelectronic routers.

This figure also indicates that even if we use the hybrid optoelectronic router with 12 optical ports, we cannot accommodate more than 20,000 servers for server racks with two or three electronic ports. However, by connecting each server rack to four hybrid optoelectronic routers, we can connect 25,000 servers by using 80 hybrid optoelectronic routers. That is, to construct a large data center, multiple links from server racks are necessary.

VI. Conclusion

In this paper, we discussed a data center network structure using hybrid optoelectronic routers. We proposed a method to construct a data center network that uses hybrid optoelectronic routers efficiently. We investigated the effect of the network structure on the number of routers required to accommodate a large amount of traffic in a mega data center. The results indicate that multiple links from server racks are necessary to construct a large data center, even if hybrid optoelectronic routers are used.

Acknowledgments

This work was funded by the National Institute of Information and Communications Technology (NICT) R&D program, “Basic technologies for high-performance optoelectronic hybrid packet router”.

References

[1] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, “The cost of a cloud: research problems in data center networks,” ACM SIGCOMM Computer Communication Review, vol. 39, Jan. 2009, pp. 68–73.

[2] D. Abts, M. Marty, P. Wells, P. Klausler, and H. Liu, “Energy proportional datacenter networks,” in Proceedings of the 37th annual international symposium on computer architecture (ISCA 2010), June 2010, pp. 338–347.

[3] S. J. B. Yoo, “Optical packet and burst switching technologies for the future photonic internet,” Journal of Lightwave Technology, vol. 24, Dec. 2006, pp. 4468–4492.

[4] S. A. lbrahim et al., “100-Gb/s optical packet switching technologies for data center networks,” in Proceedings of Photonics in Switching, July 2014, pp. 1–2.

[5] Y. Ohsita and M. Murata, “Data center network topologies using optical packet switches,” in Proceedings of DCPerf, June 2012, pp. 57–64.

[6] N. Farrington et al., “Helios: a hybrid electrical/optical switch architecture for modular data centers,” ACM SIGCOMM Computer Communication Review, vol. 40, Oct. 2010, pp. 339–350.

[7] G. Wang et al., “c-through: Part-time optics in data centers,” in Proceedings of ACM SIGCOMM, Oct. 2010, pp. 327–338.

[8] T. Benson, A. Anand, A. Akella, and M. Zhang, “MicroTE: Fine Grained Traffic Engineering for Data Centers,” in Proceedings of ACM CoNEXT, Dec. 2011, pp. 1–12.

[9] C. Guillemot et al., “Transparent optical packet switching: The european ACTS KEOPS project approach,” Journal of Lightwave Technology, vol. 16, Dec. 1998, pp. 2117–2134.

[10] Z. Zhu et al., “Rf photonics signal processing in subcarrier multiplexed optical-label switching communication systems,” Journal of Lightwave Technology, vol. 21, Dec. 2003, pp. 3155–3166.

[11] Z. Pan, H. Yang, Z. Zhu, and S. J. B. Yoo, “Demonstration of an optical-label switching router with multicast and contention resolution at mixed data rates,” IEEE Photonics Technology Letters, vol. 18, Jan. 2006, pp. 307–309.

[12] K. Xi, Y. H. Kao, M. Yang, and H. J. Chao, “Petabit optical switch for data center networks.” Technical Report, Polytechnic Institute of New York University, http://eeweb.poly.edu/~chao/publications/petasw.pdf.

[13] X. Ye et al., “DOS: a scalable optical switch for datacenters,” in Proceedings of ANCS, Oct. 2010, pp. 1–12.

[14] O. Liboiron-Ladouceur, I. Cerutti, P. G. Raponi, N. Andriolli, and P. Castoldi, “Energy-efficient design of a scalable optical multiplane interconnection architecture,” IEEE Journal of Selected Topics in Quantum Electronics, vol. 17, Mar. 2011, pp. 377–383.

[15] C. Guo et al., “BCube: A high performance, server-centric network architecture for modular data centers,” ACM SIGCOMM Computer Communication Review, vol. 39, Aug. 2009, pp. 63–74.

[16] M. Kodialam, T. V. Lakshman, and S. Sengupta, “Efficient and robust routing of highly variable traffic,” in Proceedings of HotNets, Nov. 2004, pp. 1–6.