Evaluation of Data Center Network Structures Considering Routing Methods

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

***Abstract*—In a data center, servers communicate with each other to handle a large amount of data, and the network within the data center should provide sufficient bandwidth. In addition, the traffic pattern in a data center changes in a short interval, and the data center network should accommodate such frequently changing traffic. Since it is hard to obtain traffic information of the whole network in a short interval, the routing methods using local traffic information are suitable for a large data center network. Though there are many researches to construct data center networks, none of them discuss the characteristics of the data center network structures that can provide sufficient bandwidth considering the routing methods. In this paper, we evaluate the network structures constructed by setting various parameters of the Generalized Flattened Butterfly (GFB), FatTree and Torus considering the routing methods. The results show that the network constructed in a hierarchical manner and having multiple links from a node in each layer can provide sufficient bandwidth between all servers.**

Keywords*—*data center; network structure; routing method

I. INTRODUCTION

In recent years, online services such as cloud computing have become popular, and the amount of data, required to be processed by such online services, is increasing. To handle such a large amount of data, large data centers with hundreds of thousands of servers have been built.

In a data center, servers handle a large amount of data by communicating with each other. The performance of the data center depends on the network connecting between servers. There are many researches to construct data center networks [1-5]. A network structure called FatTree, which can provide large bandwidth between all servers using commodity switches with a small number of ports, has been proposed by M. Al-Fares et al. [1]. A cost-efficient structure for a high-radix network, which can provide large bandwidth and a small average number of hops between all servers, has also been proposed [2]. The methods to connect a large number of servers with a small number of switches have also been proposed by C. Guo et al. [3, 4]. In addition, we have proposed the method, named GFB, which can construct appropriate network structures by setting parameters to meet the demands of applications in a data center [5]. However, these papers aimed to propose the specific network structures, and did not discuss the characteristics of network structures suitable to a data center sufficiently.

In a data center, handling significant traffic changes and failures is another problem. Traffic in a large data center changes frequently [6]. Failures are also common in a large data center [7]. Even when such traffic changes or failures occur, we should keep the performance of the data center. One of the important methods to keep the performance of the data center is the routing methods. By changing the routes based on the current traffic and network status, we can keep the large bandwidth between servers even in case of traffic changes or failures.

To calculate the optimal routes of the traffic between all server pairs, we require the traffic information of the whole network. In a larger data center, however, it is difficult to collect the traffic information of the whole network in a short interval, while the traffic changes frequently. Thus, in a data center, a routing method should control routes between servers based on the local information which can be obtained by each switch. There are several researches to control routes between servers for a data center [8]. A. Greenberg et al. proposed a method to distribute loads by randomly choosing switches in high layers at the Tree structure [8, 1]. M. Al-Fares et al. use the Equal-Cost Multi-Path routing (ECMP) to distribute loads to multiple paths in the FatTree [1]. However, these methods consider only a particular network structure and none of them clarify the network structure suitable to the routing methods.

In this paper, we evaluate the various network structures considering the routing methods for a data center. Then, we clarify the characteristics of the network structures where the routing methods can provide a large bandwidth between all servers.

The rest of this paper is organized as follow. Sections II and III explain the network structures and the routing methods used in this paper, respectively. In Section IV, we describe the evaluation process. Then we discuss the results of the evaluation in Section V. Finally, we conclude this paper in Section VI.

II. NETWORK STRUCTURE

In this paper, we compare the performance of various network structures constructed with switches which have the same number of 10 Gbps ports, and clarify the characteristics of the network structures that can provide sufficient bandwidth between all servers. In this paper, we use the following network structures.

A. GFB

The GFB [5] is a network structure which is constructed hierarchically; the upper-layer GFB is constructed by connecting the lower-layer GFBs. The GFB has the following parameters.

- Number of layers: k
- Number of links per node used to construct layer-k GFB: L_k

- Number of layer-$(k-1)$ GFBs used to construct layer-k GFB: N_k

By setting these parameters, we can construct various network structures.

The layer-k GFB is constructed based on the ID assigned for each layer-$(k-1)$ GFB. First, the GFBs having the nearest ID are connected to construct a ring topology. Then the residual links are used to connect the layer-$(k-1)$ GFBs so that the interval of the IDs of the layer-$(k-1)$ GFBs connected to a certain layer-$(k-1)$ GFB is equal.

In this paper, we construct and compare the various network structures by setting the parameters of the GFB.

B. FatTree

The method to construct the topology called *FatTree* by using switches with small number of ports was proposed by Al-Fares et al. [1]. The FatTree is a tree structure including multiple roots and multiple pods constructed of multiple switches.

Each pod is regarded as the switch with a large number of ports constructed by multiple switches with a small number of ports. Pods are constructed as the butterfly topology, where each switch uses a half of its ports to connect it to the switches of the upper layer, and the other half of its ports to connect it to the switches of the lower layer. The switches at the lowest layers are connected to the servers.

Though the method proposed by Al-Fares et al. [1] constructs the 3-layer FatTree, which is constructed of the root switches and the pods with two layers, we can construct the FatTree topologies with more layers. The k-layer FatTree constructed of switches with n ports includes $(2k-1)\frac{n}{2}^{k-1}$ switches.

C. Torus

Torus is a network structure constructed by locating switches in multidimensional grid. The n-dimensional Torus is constructed based on the IDs of the switch which are the n-dimensional vectors. In the n-dimensional Torus, the switch A is connected to the switch B whose ID is next to the ID of the switch A in a dimension and equals the ID of the switch A in the other dimensions. The n-dimensional Torus requires switches with $2n$ ports.

III. Routing method

A routing method selects the route passed by each packet from the candidates of the routes by using the traffic information. In this paper, we classify the routing methods based on the traffic information used by them.

The first one is a method that does not use the traffic information. In this method, each switch selects the next node passed by a packet randomly from the candidates of the next hop. We call this method the *random routing*.

The second one uses only the local traffic information that can be obtained by each switch. In this method, each switch selects the next hop whose corresponding link has the least utilization among the candidates. We call this method the *local routing*.

The third method uses traffic information of the whole network. In this routing method, a server, which is used to calculate a route of traffic, collects the overall traffic information and determines a route of traffic based on this information. We call this method the *global routing*.

In this paper, we use the above three types of the routing methods. In each method, we use the two types of the candidates of the routes; the first type of the candidates includes only the shortest paths, and the second one includes the shortest paths and the paths which are one hop longer than the shortest path.

The random routing and the local routing immediately adapt the route of a packet so as to suite the current traffic without collecting overall traffic information when traffic changes. On the other hand, the global routing cannot change the routes before the traffic information of the whole network is collected. For the global routing, we evaluate both cases that the traffic information after the traffic changes is collected or only the traffic information before the traffic changes is obtained.

In this paper, we also evaluate the case that link failure occurs. In all of the routing methods used in this paper, we calculate the routes after eliminating the routes including the failed links from the candidates. If no candidates remain, the traffic cannot be accommodated.

IV. Evaluation process

A. Overview

In this paper, we evaluate the combination of the network structures and the routing methods. In this evaluation, we generate traffic between all servers. Then, the network structure accommodates the traffic along the routes calculated by the routing methods. If we cannot find the routes without congestion, the traffic cannot be accommodated. In this paper, we focus on the case of the maximum amount of traffic that can be accommodated by the combination of the network structures and the routing methods, to evaluate the maximum bandwidth provided between all servers.

B. Traffic

A data center has two kinds of traffic; the mice traffic and the elephant traffic. The mice traffic is generated when a server exchanges a small message with the other servers. The elephant traffic is generated when a server exchanges a big data, such as files.

In this paper, the traffic is generated as the combination of the mice and elephant traffic. The mice traffic is generated between all servers and the amounts of the mice traffic are set to the random values so as to make the mice traffic occupying around 10 % of the bandwidth of all links in a network. The elephant traffic is generated between randomly selected server pairs. The amount of the elephant traffic is set to the largest value that can be accommodated by the combination of the network structure and the routing method. To set the amount

 ISBN: 978-1-61208-256-1

TABLE I
NETWORK STRUCTURES WITH VARIOUS NUMBERS OF LINKS AT EACH LAYER

name	N_0	N_1	N_2	L_0	L_1	L_2
GFB224	4	6	14	2	2	4
GFB242	4	6	14	2	4	2
GFB314	4	6	14	3	1	4
GFB323	4	6	14	3	2	3

of elephant traffic, we increase the amount of the elephant traffic unless the congestion occurs.

In this evaluation, we generate the traffic change by regenerating the mice traffic and newly selecting the server pairs where the elephant traffic is generated.

C. Metrics

In this paper, we focus on the case that the largest amount of the elephant traffic is generated. Then we investigate the smallest amount of the elephant traffic between server pair among the server pairs where the elephant traffic is generated. By investigating the smallest amount of the elephant traffic, we compare the bandwidth between servers that can be provided at least. In addition, when we generate link failures, we also investigate the communication failure ratio which is defined by the ratio of flows between switch pairs that have no routes from the source server to the destination server.

V. EVALUATION

In this section, we discuss the characteristics of a network structures which can provide a large bandwidth between all servers. In this evaluation, to clarify the characteristics of the network structures suitable to a data center, we compare the network structures constructed by setting various parameters of the GFB. The GFB constructs various network structures by setting its parameters. By comparing the network structures constructed by setting the parameters of the GFB, we clarify the characteristics of the network structures suitable to a data center. First, we compare the performances of network structures with various numbers of links at each layer. Then, we compare the performances of network structures with various numbers of layers.

We also compare the performances of network structures when link failures occur. Finally we compare the performances of the GFB with that of the Torus and the FatTree.

A. Comparison of network structures with various link at each layer

In this subsection, we use the network structures shown in Table I, constructed by connecting 336 switches with 8 ports. The network structures are constructed by setting the parameters in the GFB. All of the network structures used in this subsection, N_i are set to the same value. We change L_i to clarify the impacts of the number of the links in each layer on the bandwidth provided between servers. In this subsection, we refer each network structure by the number of links at each layer. The elephant traffic to one destination server per one switch is generated. The server pairs where the elephant traffic is generated are selected randomly.

Figure 1 shows the minimum amount of the elephant traffic between a certain server pair which can be accommodated to each network structure by using each routing method. As shown in this figure, the global routing can provide the largest bandwidth between all servers at any network structures if the accurate traffic information is collected. The global routing, however, can provide only as large bandwidth between all servers as the random routing if we have only the traffic information before the traffic change. It is difficult to collect accurate traffic information of all links in a short time interval, though traffic in a large data center changes frequently[6]. Therefore, in the data center, the routes should be calculated based on the local traffic information that can be obtained by each switch.

As shown in Fig. 1, the local routing can provide as large bandwidth between all servers as the random routing. This is because each switch does not have the information of the links connected to the next switches. Thus, each switch cannot know whether the next switch has the congested links, and may select the next switch having the congested links. As a result, the local routing provides only the similar bandwidth to the random routing.

This figure also indicates that the candidate routes including the one hop longer paths provide larger bandwidth. This is because each switch has more candidates by including one hop longer path. Thus, it is easy to avoid the congested links.

In the following evaluation, we focus on the characteristics of each network structure. We compare the bandwidth provided by the random routing with the candidates including the shortest path and the one hop longer path.

As shown in Fig. 1, GFB224 can provide the largest bandwidth between all servers. The largest bandwidth between all servers depends on the maximum link utilization. To discuss the link utilization, we model the probability that traffic between each switch pair passes the link by Eqs. (1) and (2). In this model, we assume that the link passed by the traffic is selected randomly among the link in the network. We also assume that the probability that the traffic between each switch pair passes the link depends only on the layer of the link, because each layer of the GFB is symmetric.

P_k is the probability to select a link in the layer k as the link passed by the traffic. P_1 is calculated by

$$P_1 = p \times \left(1 - \prod_{k=0}^{H_1-1} \left(1 - \frac{1}{l-k}\right)\right), \qquad (1)$$

where p is the probability that the flow passes a particular layer-1 GFB, l is the number of links at the layer-1 GFB and H_1 is the number of hops at the layer-1 GFB.

The layer-2 GFB uses $(N_1 \times L_2)$ links per layer-1 GFB for the connection between the layer-1 GFBs. We assume that the sufficient number of links are used to connect the layer-1 GFBs, and the layer-1 GFBs are fully connected. Thus, no traffic passes multiple links between layer-1 GFBs. Therefore,

 ISBN: 978-1-61208-256-1

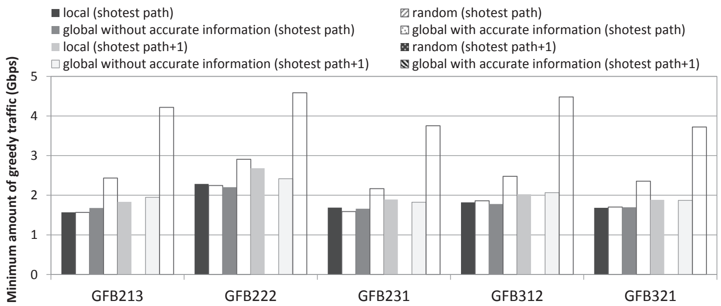

Fig. 1. Minimum amount of the elephant traffic between a certain server pair which can be accommodated by each network structure with various numbers of links at each layer

P_2 is calculated by

$$P_2 = q \times \frac{1}{l_{GFB}}, \quad (2)$$

where q is the probability that the flow passes a particular link between layer-1 GFB-pair and l_{GFB} is the number of the links between a certain layer-1 GFB-pair.

A network structure whose P_1 and P_2 in these equations is small can provide large bandwidth between all servers. As shown in Eq. (1), to make P_1 smaller, a network structure should have smaller H_1. We can reduce H_1 by reducing the number of switches in the layer-1 GFB or increasing the number of the links in the layer-1 GFB since the maximum number of hops at the layer-1 GFB is calculated by $\left\lceil \frac{N_1}{2\times(L_1-1)} \right\rceil$ according to Tarutani et al. [5]. Increasing the number of links at the higher layer also reduces H_1, because if multiple switches are connected to the laye-1 GFB including the destination, the flow goes out the GFB including the source from the switch nearest to the source among the switches connected to the destination GFB.

As shown in Eq. (2), to make P_2 smaller, a network structure should have large l_{GFB}. To increase l_{GFB}, we need to increase the number of links at layer-2 GFB, or reduce the number of the layer-1 GFBs connected at the layer-2 GFB because l_{GFB} is calculated by $\frac{N_1 \times L_2}{N_2}$.

The above discussion is also applicable when the number of layers is more than 2. The probability that the flow passes the links at the higher layer is small when the number of the links between the GFB pair is large.

According to the above discussion, to provide a large bandwidth, the number of links at each layer L_k should be sufficiently large compared with N_k. In this subsection, we set N_1 to 4, N_2 to 6, and N_3 to 14. In the GFB242, the number of links at the layer 3, L_3 is small compared with $N_3 = 14$. In the GFB314, the number of links at the layer 2, L_2 is small compared with $N_2 = 6$. Thus, these network structures cannot provide a large bandwidth between servers.

The GFB224 provides a larger bandwidth than the GFB323. This is because the number of links at the lowest layer is sufficient even when $L_1 = 2$, since the GFB of the lowest layer includes only 4 switches. In this case, by adding more links to the upper layers, we reduce the number of hops to the other layer-1 GFBs, and increase the bandwidth provided between servers compared with adding more links to the lowest layer.

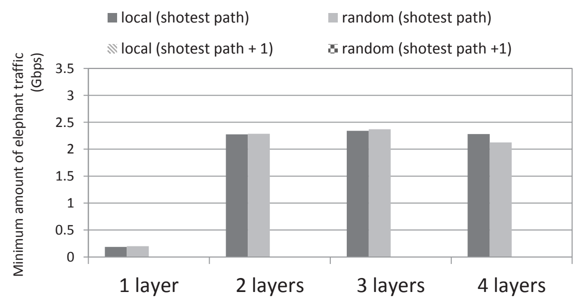

Fig. 2. Minimum amount of the elephant traffic between a certain server pair which can be accommodated by each network structure with various numbers of layers

To summarize the above discussions, the GFBs where the number of links at a particular layer is small cannot provide large bandwidth between all servers. Therefore we need a network structure which has sufficient number of links at all layers.

B. *Comparison of network structures with various numbers of layers*

In this subsection, we evaluate the performances of the network structures when we change the number of layers. In this evaluation, we use the network structures shown in Table II. In all network structures used in this subsection are constructed of 336 switches with 8 ports. The network structures constructed with various numbers of layers by setting the parameters in GFB. In this subsection, we refer each network structure by the number of layers. In this subsection, the elephant traffic to one destination server per one switch is generated. The server pairs where the elephant traffic is generated are selected randomly.

Figure 2 shows the minimum amount of elephant traffic which can be accommodated by each network structure. As shown in Fig. 2, GFB(1 layer) can provide the smallest bandwidth between all servers. GFB(2 layers) and GFB(3 layers) can provide lager bandwidth between all servers. This is because the large number of hops between servers in the

TABLE II
NETWORK STRUCTURES WITH VARIOUS NUMBERS OF LAYERS

name	the number of layer	N_0	N_1	N_2	N_3	L_0	L_1	L_2	L_3
GFB (1 layer)	1	336	-	-	-	8	-	-	-
GFB (2 layers)	2	14	24	-	-	4	4	-	-
GFB (3 layers)	3	4	6	14	-	2	2	4	-
GFB (4 layers)	4	3	4	4	7	2	2	2	2

1-layer GFB increases the link utilization.

The number of hops between a switch pair is calculated as follows. $H_k^{(i)}$ is the number of hops between a switch pair at the layer-k GFB with ID(i), and P is the set of ID of the layer-$(k-1)$ GFBs where a flow passes.

$$H_k = \sum_{i \in P} (H_{k-1}^{(i)} + 1) - 1 \tag{3}$$

As shown in Eq. (3), the number of hops between a switch pair depends on the number of GFBs at the lower layer passed by the traffic and the number of hops at the lower layer GFBs. The 1-layer GFB is constructed by adding links to a ring topology so that the interval of the IDs of the switches connected to a certain switch is equal. When the number of switches is 336, the interval of the IDs is 56, and the number of hops between switches is large. As shown in Eq. (1), the large number of hops causes the high probability that a flow passes a particular link. As a result, the 1-layer GFB cannot provide the smallest bandwidth.

In the 2-layer GFB used in this evaluation, the layer-1 GFB includes 14 switches, and the layer-2 GFB includes 24 fully connected layer-1 GFBs. Thus, the number of hops in each layer is significantly smaller than the 1-layer GFB, and the number of hops between a switch pair is small. As a result, the probability that a flow passes a particular link is smaller, and a network structures constructed in a hierarchical manner can provide larger bandwidth between all servers than the 1-layer GFB.

In the 4-layer GFB, we cannot provide as large bandwidth as the 2-layer or 3-layer GFB. This is because the number of hops in the 4-layer GFB becomes larger than that in the 2-layer or 3-layer GFB. As shown in Eq. (3), the number of hops between a switch pair is calculated by adding the number of hops at lower layer recursively. The number of the recursive calculations increases if the number of the layers increases. If the increase of the number of the layers does not reduce the number of hops in each layer sufficiently, the increase of the number of the layers makes the number of hops between a switch pair large. As a result, though the 4-layer GFB can provide larger bandwidth than the 1-layer GFB, the 4-layer GFB can provide only smaller bandwidth than the 2-layer or 3-layer GFB.

As discussed above, to provide larger bandwidth between all servers, the network structure should connect any switch pairs with a small number of hops. It is effective to reduce the number of GFBs at each layer by constructed in a hierarchical manner. However, if the number of layers is too large, the traffic between a switch pair passes many layers, and its number of hops becomes large. Therefore, we need to set the number of layer to as small value as possible without a large number of hops in each layer.

C. Evaluation in the case of link failures

In a large data center, failures such as link failure are common. Thus, the network should be robust to failures to keep the service provided in a data center even when failure occurs. In this subsection, we evaluate the communication failure ratio of various network structures when links have failed, and discuss the characteristics of a network structure robust to failures.

Figure 3 and 4 show the communication failure ratio of the network structures shown at Tables I and II when several links failed. As shown in Figs. 3 and 4, each switch having the candidates of the routes including the one hop longer path can achieve the smaller communication failure ratio than the case of the candidates including only the shortest path. This is because each switch has more candidates by including one hop longer path and it is easy to bypath the failed links.

As shown in Fig. 3, the communication failure ratio in the GFB224 is the smallest. As discussed in the previous subsection, the GFB242, the GFB314, and the GFB323 have the links passed by many flows. The failures of such links passed by many flows cause the large communication failure ratio in these network structures.

When the candidate routes include only the shortest paths, the communication failure ratio is the lowest in the 4-layer GFB. This is because the number of routes between switches is large in the 4-layer GFB. In the 2-layer GFB or the 3-layer GFB, though the number of hops is smaller than the 4-layer GFB, the number of the shortest paths between servers is small.

By including the one hop longer paths in the candidates of the routes, the 2-layer GFB and the 3-layer GFB also achieves the similar communication failure ratio to the 4-layer GFB. However, the communication failure ratio of the 1-layer GFB is large even when the candidate routes include the one hop longer paths. This is because the probability that each flow passes the link is large in the 1-layer GFB. Thus, the failure of the links in the 1-layer GFB has a large impact on many flows, and causes the large communication failure ratio.

As discussed above, a network structure, which has multiple routes between servers and the small probability that each flow passes each link, are robust to failures.

D. Comparison of GFB, Torus and FatTree

In this subsection, we compare the performances of the network structures constructed by setting parameters in the

(a) In the case of the candidate routes including only the shortest paths

(b) In the case of the candidate routes including one hop longer paths

Fig. 3. Communication failure ratio of network structures with various numbers of links at each layer in the case of link failures

(a) The case of candidate routes including only the shortest paths

(b) The case of candidate routes including one hop longer paths

Fig. 4. Communication failure ratio of network structures with various numbers of layers in the case of link failures

TABLE III
COMPARISON OF THE NETWORK STRUCTURE USED IN OUR EVALUATION

name	the number of switches	the number of links	average hops	maximum hops
GFB(5,5,6)	150	450	4.09	7
Torus	150	450	4.93	8
FatTree	189	486	6.62	7

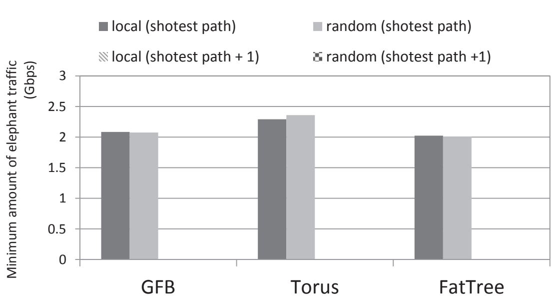

Fig. 5. Minimum amount of elephant traffic which can be accommodated by GFB, Torus and FatTree when the elephant traffic to one destination server per one switch is generated

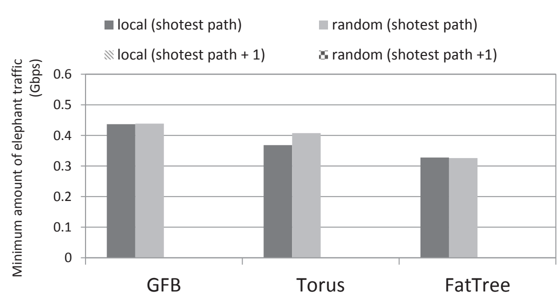

Fig. 6. Minimum amount of elephant traffic which can be accommodated by GFB, Torus and FatTree when the elephant traffic to ten destination server per one switch is generated

GFB with that of the Torus and the FatTree. The Torus is the well-studied network structure. Compared with the GFB, the Torus has a large number of hops, but it has more routes between a switch pair. The FatTree is the network structure used in the existing data centers. The FatTree has the same number of maximum number of hops as the GFB, but the

 ISBN: 978-1-61208-256-1

average number of hops between servers is larger than the GFB or the Torus. Similar to the Torus, the FatTree has multiple routes between server pairs. By comparing the GFB with these network structures, we evaluate the impacts of the average number of hops and the number of routes between switch pairs.

In this evaluation, we use the network structures shown in Table III. Table III also includes the average number of hops and the max number of hops between all switch in each network structure. The GFB and the Torus are constructed of 150 switches with 6 ports, and the FatTree is constructed of 189 switches with 6 ports. In this evaluation we use a network structure by setting parameters $[N_0 = 5, N_1 = 5, N_2 = 6, L_0 = 2, L_1 = 2, L_2 = 2]$ in GFB. Also we use the 3-dimentional $5 \times 5 \times 6$ Torus, and the 4-layer FatTree which is constructed by allocating 54 switches at the lowest layer. In the FatTree, only the switches at the lowest layer are connected to servers. In the FatTree, we generate the same number of flows as in the GFB and the Torus. In the FatTree, we obtain only the case that each switch has the candidates of the routes including only the shortest paths because the one hop longer path does not exist in the FatTree.

In this subsection, we obtain the evaluation results of two cases of the ratios of the generated elephant traffic. In the first case, we generate that the elephant traffic to one destination server per one switch. In the other case, we generate the elephant traffic to ten destination servers per one switch is generated. Figure 5 shows the minimum amount of the elephant traffic which can be accommodated by each network structure when the elephant traffic to one destination server per one switch is generated. Figure 6 shows the minimum amount of the elephant traffic which can be accommodated by each network structure when the elephant traffic to ten destination servers per one switch is generated. As shown in Figs. 5 and 6, the GFB and the Torus can provide larger bandwidth between all servers than the FatTree. This is caused by the large average number of hops of the FatTree. In the FatTree, each flow passes more links, and requires the bandwidth of a large number of links. As a result, the bandwidth of each link is occupied with a small amount of traffic between servers.

Fig. 5 also indicates that the Torus can accommodate more traffic than the GFB, if the candidate routes include only the shortest path. This is because the Torus has a larger number of the shortest paths between switches than the GFB. However, in case of the candidate routes including one hop longer path, the GFB also has the sufficient number of routes between switches, and can accommodate the similar amount of traffic to the Torus.

As shown in Fig. 6, the GFB can provide the largest bandwidth between all servers by using the local routing with the candidates including one hop longer path. Though the Torus provides the largest bandwidth in Fig. 5, the Torus cannot provide as large bandwidth as the GFB in Fig. 6. This is because that the number of hops between switch pairs is larger in the Torus than the GFB. Thus, in the Torus, the bandwidth of each link is occupied with a small amount of traffic between servers.

As discussed above, the network structure with a large number of candidate routes between servers is required to provide a large bandwidth between servers when the number of flows in the network is small. However, when the number of flows in the network is large, the average number of hops of the flow becomes more important. Thus, the network structures should have small average number of hops to provide a large bandwidth when the number of flows is large.

VI. Conclusion

In this paper, we evaluated the data center networks considering the routing methods. According to the results, to provide a large bandwidth between servers, we should make the number of hops small by constructing the network in a hierarchical manner, and make the number of routes between servers large by adding multiple links from a node in each layer.

Acknowledgments

This work is a part of "Research & Development of Basic Technologies for High Performance Opto-electronic Hybrid Packet Router" supported by National Institute of Information and Communications Technology (NICT).

References

[1] M. Al-Fares, A. Loukissas, and A. Vahdat, "A Scalable, Commodity Data Center Network Architecture," ACM SIGCOMM Computer Communication Review, vol. 38, Oct. 2008, pp. 63–74.

[2] J. Kim, W. J. Dally, and D. Abts, "Flattened butterfly: a cost-efficient topology for high-radix networks," in Proceedings of the 34th annual international symposium on Computer architecture, vol. 35, Jun. 2007, pp. 126–137.

[3] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, and S. Lu, "DCell: A scalable and fault-tolerant network structure for data centers," ACM SIGCOMM Computer Communication Review, vol. 38, Aug. 2008, pp. 75–86.

[4] C. G. 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.

[5] Y. Tarutani, Y. Ohsita, and M. Murata, "A Virtual Network to Achieve Low Energy Consumption in Optical Large-scale Datacenter," in Proceedings of the 13th International Conference on Communication Systems IEEE ICCS 2012, Nov. 2012.

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

[7] P. Gill, N. Jain, and N. Nagappan, "Understanding network failures in data centers: measurement, analysis, and implications," ACM SIGCOMM Computer Communication Review, vol. 41, Aug. 2011, pp. 350–361.

[8] A. G. et al., "VL2: A scalable and flexible data center network," ACM SIGCOMM Computer Communication Review, vol. 39, Aug. 2009, pp. 51–62.

 ISBN: 978-1-61208-256-1