A Virtual Network to Achieve Low Energy Consumption in Optical Large-scale Datacenter

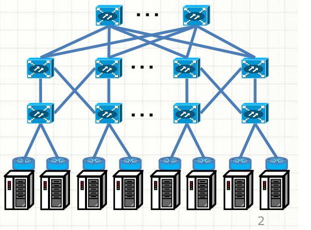
Graduate School of Information Science and Technology, Osaka University, Japan

Yuya Tarutani, Yuichi Ohsita, and Masayuki Murata

Advanced Network Architecture Research Group
http://www.anarg.jp/

Datacenter

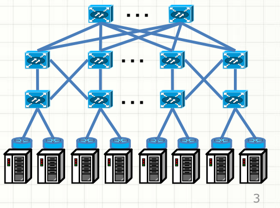
- A large datacenter have been built to handle a large amount of data
  - A Datacenter is constructed by many servers and a network between servers
  - Servers handle a large amount of data by communicating with each other
- The network has a large impact on the performance of the datacenter



Requirements of datacenter network

- Provide communication with large bandwidth and small delay between servers
  - The lack of bandwidth or large delay between servers degrades the performance of handling datacenter applications
- Reduce the energy consumption in the datacenter
  - One approach is to turn on only the minimal devices

The traditional datacenter cannot satisfy both requirements
  - Many devices are required to provide communication with large bandwidth and small delay



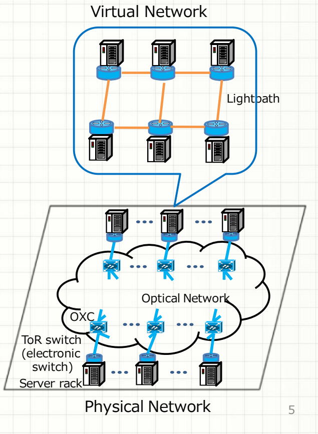
Research goal

- Goal:
  Construct the datacenter network to satisfy both requirements
  - Provide communication with large bandwidth and small delay
  - Reduction of energy consumption
- Approach
  - Dynamically reconfigure the network to satisfy the current requirements with the small number of devices

Optical datacenter network

- Physical network
  - The core network is constructed by using the OXCs and optical fibers
  - ToR(Top of Rack) switches are connected to the core network
- Virtual Network
  - The virtual network is constructed by connecting the lightpaths between ToR switches

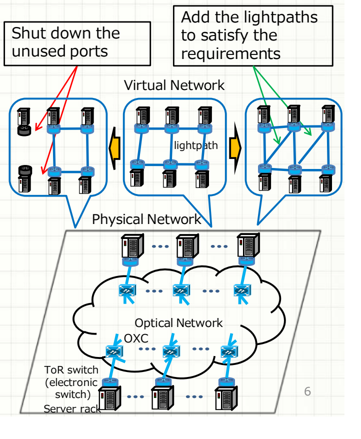
The virtual network can be reconfigured dynamically



Reduction of the energy consumption by reconfiguring the virtual netowrk

The energy consumption of optical switch is much smaller than that of electronic switch

- Our approach
  - Minimize the number of the lightpaths used by the virtual network
  - Shut down the unused ports of electronic switch



Virtual Network reconfiguration method

- Requirements
  - Minimize the number of lightpaths
    - To minimize the energy consumption.
  - Small calculation time
    - To accommodate the traffic that changes in a short period
- Existing virtual network reconfiguration methods does not satisfy the above requirements
  - They require the large calculation time to optimize the virtual network
  - They cannot reconfigure the virtual network so as to suit the current condition

7

Our approach to reconfigure the Virtual Network

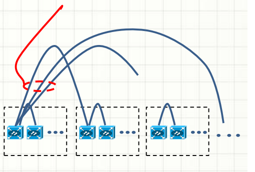
- Reconfiguration of the virtual network by setting the parameters of the base topology
  - The calculation time is small even in a large data center
- Key issues
  - The Base topology
    - Various topologies should be constructed by setting the parameters
    - The network conditions can be calculated from the parameters
  - The method to set the parameter to satisfy the requirements

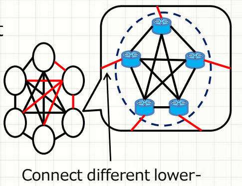
8

Existing topologies

- Flattened Butterfly (FB)
  - Each switch uses the links to connect the all lower-layer FBs
  - Advantage : Provide the sufficiently large bandwidth
  - Disadvantage : Require the large number of links
- Switch-based DCell
  - Each switch uses one link to connect the different lower-layer DCell
  - Advantage : Require the small number of links
  - Disadvantage : Cannot provide the sufficiently large bandwidth

Connect all lower-layer FBs





Connect different lower-layer DCells

Topology between these topologies

9

GFB（Generalized Flattened Butterfly）

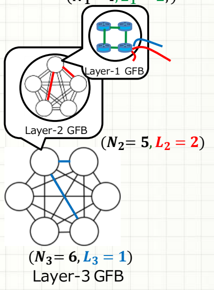
- The GFB is constructed hierarchically
  - The upper-layer GFB is constructed by connecting multiple lower-layer GFBs
- Parameters of the GFB
  - k : Number of layers
  - N_k : Number of layer-$k-1$ GFBs used to construct layer-k GFB
  - L_k : Number of links per switch use to construct layer-k GFB
- The following metrics are calculated from the parameters
  - The maximum number of hops
  - The number of flows passing a link

10

The structure of GFB

- Hierarchical Structure
  - The layer- k GFB is constructed of the N_k layer- $k-1$ GFBs.
  - Each node has L_k links to connect the other layer- $k-1$ GFBs
- Structure of each layer
  - The structure based on the ring topology
    - If fully connect all layer- $k-1$ GFBs, the layer- $k-1$ GFBs are fully connected.
    - Otherwise,
      1. Construct the ring topology
      2. Add links so that the interval of ID of the GFB connected to the same GFB is equal

$(N_1 = 4, L_1 = 2,)$



$(N_2 = 5, L_2 = 2)$

$(N_3 = 6, L_3 = 1)$

Layer-3 GFB

k : Number of layer
N_k: Number of layer-$k-1$ GFBs used to construct layer-k GFB
L_k : Number of links per switch use to construct layer-k GFB

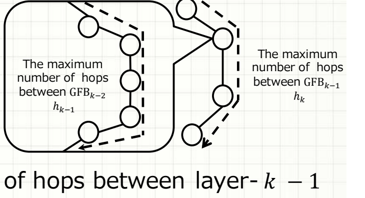
11

The maximum number of hops in the GFB

- Calculated by following equation

$$H_k = (h_k + 1)H_{k-1} + h_k$$

H_k : The maximum number of hops between switches in layer-k GFB
h_k : The maximum number of hops between layer- $k-1$ GFBs in layer-k



- The maximum number of hops between layer- $k-1$ GFBs in layer-k
  - Calculated by the interval of the IDs of layer-$k-1$ GFB connected to same layer-$k-1$ GFBs

$$h_k = \begin{cases} 1 & \left(L_k^{gfb} \geq (N_k - 1)\right) \\ \left\lceil \frac{p_k}{2} \right\rceil & \left(L_k^{gfb} \leq 3\right) \\ \left\lfloor \frac{p_k}{2} \right\rfloor + 1 & (otherwise) \end{cases}$$

L_k^{gfb} : The number of links used to connect one layer-$k-1$GFB to other layer-$k-1$ GFBs
p_k : The variable that indicates the interval of the IDs of the layer-$k-1$ GFB connected to same layer-k – 1 GFBs

12

The number of flows passing a link in the GFB

- Calculated by the following equation

$$X_k = F_k \frac{\sum_{i=1}^{h_k} i\, s_k(i)}{L_k \prod_{i=1}^{k} N_i}$$

F_k : The number of flows passing the link between layer-$k-1$ GFBs
h_k : The maximum number of hops between layer- k -1 GFBs in layer-k
X_k : The number of flows passing the links between ToR switches in layer-k GFB
$s_k(i)$: The number of layer-$k-1$ GFB pairs whose flow passes i links between layer-$k-1$ GFBs

  - Assumption: The flows between layer-$k-1$ GFBs are balanced among all links between layer-$k-1$ GFBs
- The number of flows passing the link between layer-$k-1$ GFBs
  - The sum of flows whose source and destination switches belong to the layer-$k-1$ GFBs

13

Steps to Calculate the parameters to satisfy the requirements

K←generate_candidate_of_layers(H)
For each k in K
  N_k, L_k ←set_parameters_based_on_hops(H)
  N_k, L_k ←set_parameters_to_provide_bandwidth(B)
End for
k, N_k, L_k ←Select_the_parameters_with_the_smallest_links()

- Requirements
  B : Required bandwidth between server
  H : The acceptable number of hops between servers

14

Calculate the candidate of the number of layers

- The candidate of the number of layers k must satisfy the following condition

$$H_{\max} \geqq 2^k - 1$$

H_{max} : The acceptable number of hops between switches

  - Because the maximum number of hops of the GFB is not less than the case that $h_k = 1$ for all layers

$$H_k = (h_k + 1)H_{k-1} + h_k$$

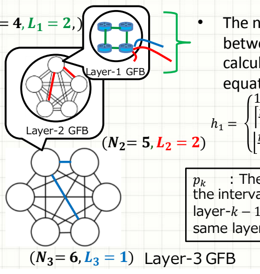
H_k : The maximum number of hops between switches in layer-k GFB
h_k : The maximum number of hops between layer- k -1 GFBs in layer-k

- We consider the all k satisfying the above condition as the candidates of the number of layers

15

Calculate the parameter to satisfy the acceptable number of hops

- Change the number of links only in the layer-1 GFBs
  - Because the number of links in the upper-layer GFBs become enough to connect all layer-$k-1$ GFB pairs, even in case of small L_k
- L_1 is set based on the maximum number of hops between switches calculated from the parameters.



- The number of links in the upper-layer GFBs is enough to connect all layer-k -1 GFB pairs
- The number of hops between switches is calculated by following equation

$$h_1 = \begin{cases} 1 & (L_1 \geq (N_1 - 1)) \\ \left\lceil \frac{p_1}{2} \right\rceil & (L_1 \leq 3) \\ \left\lfloor \frac{p_1}{2} \right\rfloor + 1 & (otherwise) \end{cases}$$

p_k : The variable that indicates the interval of the IDs of the layer-$k-1$ GFB connected to same layer-k - 1 GFBs

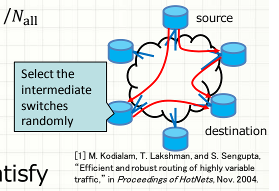
16

Calculate the parameter to provide the sufficient bandwidth

- Set parameters considering the load balancing
  - VLB(Valiant Load Balancing) [1]
    - Select the intermediate switches randomly regardless of destination to avoid the concentration of traffic on certain link
    - The amount of traffic between each ToR switch pair is calculated by following equation

$$T = (T^{SW_{to}} + T^{SW_{from}}) / N_{\text{all}}$$

N_{all}: The number of ToR switches
$T^{SW_{to}}$: The traffic amount of traffic to a ToR switches
$T^{SW_{from}}$: The amount of traffic from a ToR switches
T: The traffic amount between ToR switches using VLB
X : The number of flows passing links between ToR switches
B : Bandwidth of lightpath



[1] M. Kodialam, T. Lakshman, and S. Sengupta, "Efficient and robust routing of highly variable traffic," in *Proceedings of HotNets*, Nov. 2004.

- Calculate the parameters to satisfy the following condition

$$X \leq B \ / \ T_{VLB}$$

17

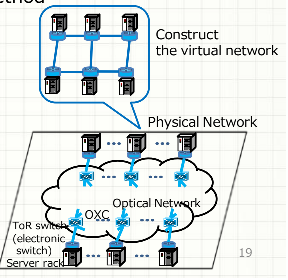
Evaluation settings

- Number of ToR switches : 420
- Bandwidth of lightpath : 10 Gbps
- Number of wavelengths on optical fiber is sufficient
- Requirements
  - Maximum amount of traffic from/to ToR switch
  - Maximum number of hops
- Metric
  - Number of used ports per ToR switches
    - Most of the energy is consumed by the electronic ports of ToR switches in the virtual network
    - Minimize the number of electronic ports to reduce the energy consumption

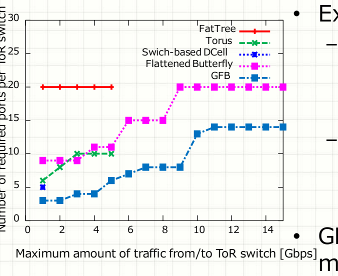
Compared methods

- We compare the virtual network constructed over the physical network
- Compared the following topologies
  - The GFB constructed by our method
    - The parameters of the GFB are set to satisfy the requirements
  - Existing topologies
    - FatTree
    - Torus
    - Flattened Butterfly
    - Switch-based DCell

    The parameters of these topologies are set to minimize the number of ports, considering the requirements



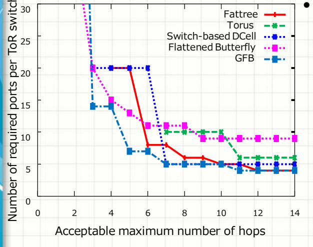
Evaluation Result



Evaluation settings
- Number of ToR switches : 420
- Bandwidth : of one link : 10Gbps

- Existing topologies
  - Some topologies cannot satisfy the requirement even if the number of ports is large
  - In the Flattened Butterfly, the number of required ports is large to satisfy the requirements in all cases
- GFB constructed by our method
  - Use the smallest number of required ports to satisfy the requirements in all cases

Evaluation Result



Evaluation settings
- Number of ToR switches : 420
- Bandwidth : of one link : 10Gbps

- Existing topologies
  - Some topologies cannot satisfy the requirement even if the number of ports is large
  - In the Flattened Butterfly, the number of used ports cannot be reduced even if the maximum number of hops is large
- GFB constructed by our method
  - Use the smallest number of required ports to satisfy the requirements in all cases

Conclusion and Future work

- Conclusion
  - We introduce the virtual network configured over the data center network constructed of the OXCs and the electronic switches
  - We propose a method to reconfigure the virtual network for the datacenter network
  - Numerical evaluations show that our method constructs the topology satisfying the requirements with small energy consumption
- Future work
  - Distributed autonomous virtual network control methods