Dual Centric Data Center Network Architectures

DAWEI LI, JIE WU (TEMPLE UNIVERSITY)

ZHIYONG LIU, AND **FA ZHANG** (CHINESE ACADEMY OF SCIENCES)

ICPP 2015

Agenda

Introduction

Preliminaries

Dual-Centric DCNs: FSquare And FRectangle

Comparison With Various DCN Architectures

Simulations

Conclusion And Future Works

Introduction

Data centers have become important infrastructures to support various cloud computing services,

varying from web search, email, video streaming social networking, to distributed file systems and data processing engines.

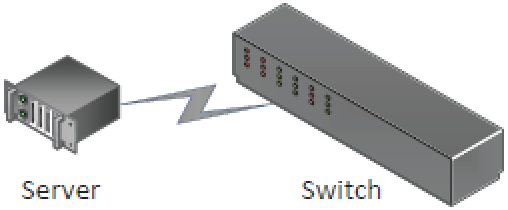
Introduction

Three types of connections:

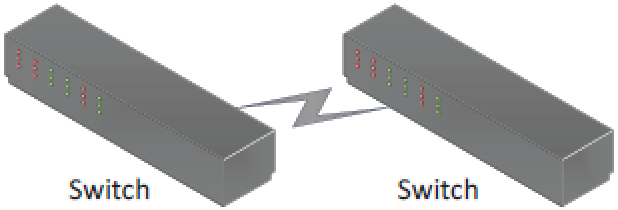
- Server-switch connection (a)
- Switch-switch connection (b)
- Server-server connection (c)

Two classes of DCN architectures:

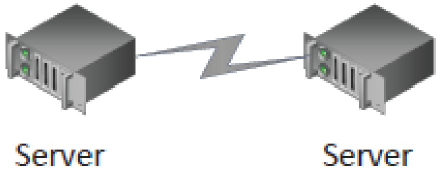
- Switch-centric architecture
  - Only server-switch and switch-switch connections (a and b), no server-server connections. Eg, Fat-Tree Flattened Butterfly
- Server-centric architecture
  - Mostly, only server-switch and server-server connections (a and c), no switch-switch connections. Eg: BCube, FiConn, DCell



(a) Server-switch connection



(b) Switch-switch connection



(c) Server-server connection

Introduction

Switch-centric vs. Server-centric.

Switch-centric	Switch-centric
➢Fast switching capability ➢Less programmability ➢Expansive ➢Power hungry	➢Larger processing delays ➢High programmability

Can we combine the advantages of both categories?

fast switching capability + High programmability

How about **dual-centric** DCN architectures?

Introduction

Contributions

- We formally introduce a new class of DCN architectures: dual-centric DCN architectures, and propose two novel architectures: **FSquare and FRectangle**.

- To enable **fair and meaningful comparisons** among existing DCN architectures and our proposed ones, we propose a **unified path length definition** and a **unified diameter definition**, for general DCNs. Also, to characterize the power efficiency of a general DCN, we propose a **unified DCN power consumption** model.

- By investigating the two proposed architectures and by comparing them with existing DCN, we show that dual-centric architectures can have appealing properties for practical DCN designs.

Agenda

Introduction

Preliminaries

Dual-Centric DCN Architectures

More On FSquare And FRectangle

Simulations

Conclusion And Future Works

Preliminaries

- Packets on switches and servers experience 3 important delays: *processing delay, transmission delay, and queuing delay*.

  $d_{w,p}$, $d_{w,t}$, $d_{w,q}$, and $d_{v,p}$, $d_{v,t}$, $d_{v,q}$.

- Switches can operate in 2 modes: store-and-forward & cut-through.
  - Store-and-forward, $d_w = d_{w,p} + d_{w,t},\ d_{w,t,} = S_{packet} / r_{bit}$
  - Cut-through, $d_w = d_{w,p} = 2\mu s$

Preliminaries

- Unified Path Length Definition:

$$d_P = n_{P,w} d_w + (n_{P,v} + 1) d_v,$$

$n_{P,w}$: the number of switches in a path.
$n_{P,v}$: the number of servers in a path (excluding the source and the destination).
d_w : the processing delay on a switch.
d_v : the processing delay on a server.

- Unified Diameter in a DCN:

$$d = \max_{P \in \{\mathcal{P}\}} d_P,$$

Preliminaries

- DCN power consumption per server:

$$p_V = p_{dcn}/N_v = p_w N_w/N_v + n_{nic} p_{nic} + \alpha p_{fwd}.$$

p_w : Power consumption of a switch
N_w : The number of switches in a DCN
N_v : The number of servers in a DCN
n_{nic} : The average number of NIC ports each server uses
p_{nic} : The power consumption of a NIC port
α : Whether the server is involved in packet relaying
p_{fwd} : Power consumption of a server's packet forwarding engine.

Agenda

Introduction

Preliminaries

Dual-Centric DCNs: FSquare And FRectangle

Comparison With Various DCN Architectures

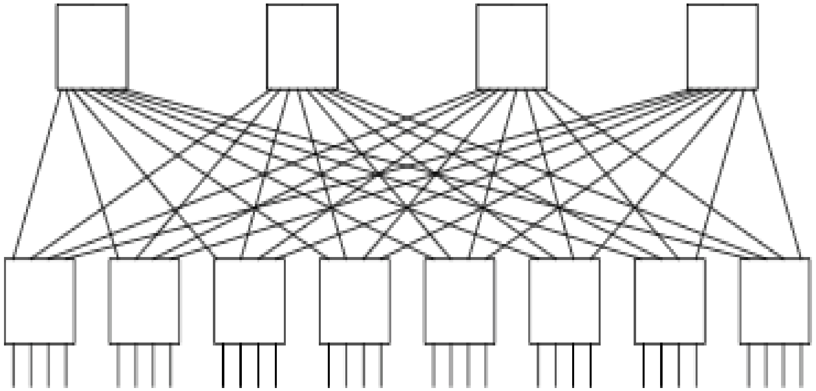
Simulations

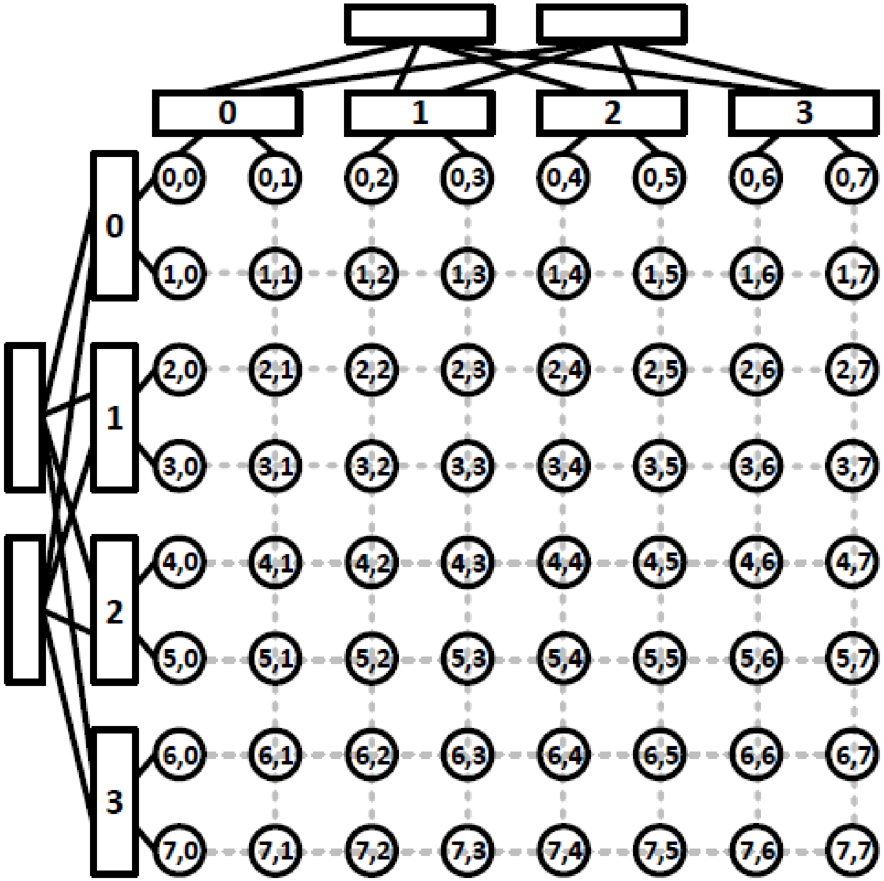
Conclusion And Future Works

Dual-Centric DCNs: FSquare And FRectangle

FSquare Construction:

The switches and servers in each row and each column form a simple instance of the folded Clos topology.





Dual-Centric DCNs: FSquare And FRectangle

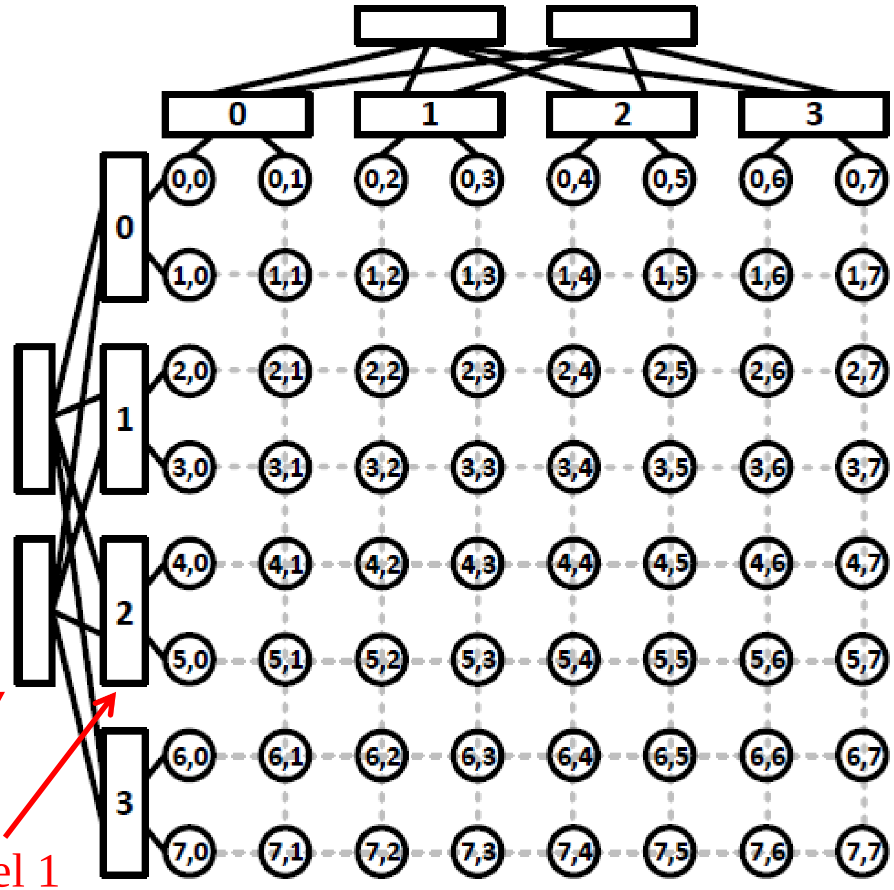
FSquare (n):

Built from servers with 2 NIC ports and n-port switches

n level 1 switches

n/2 level 2 switches

The set of n/2 level 2 switches and the set of n level 1 switches form a complete bipartite graph.



level 2

level 1

Dual-Centric DCNs: FSquare And FRectangle

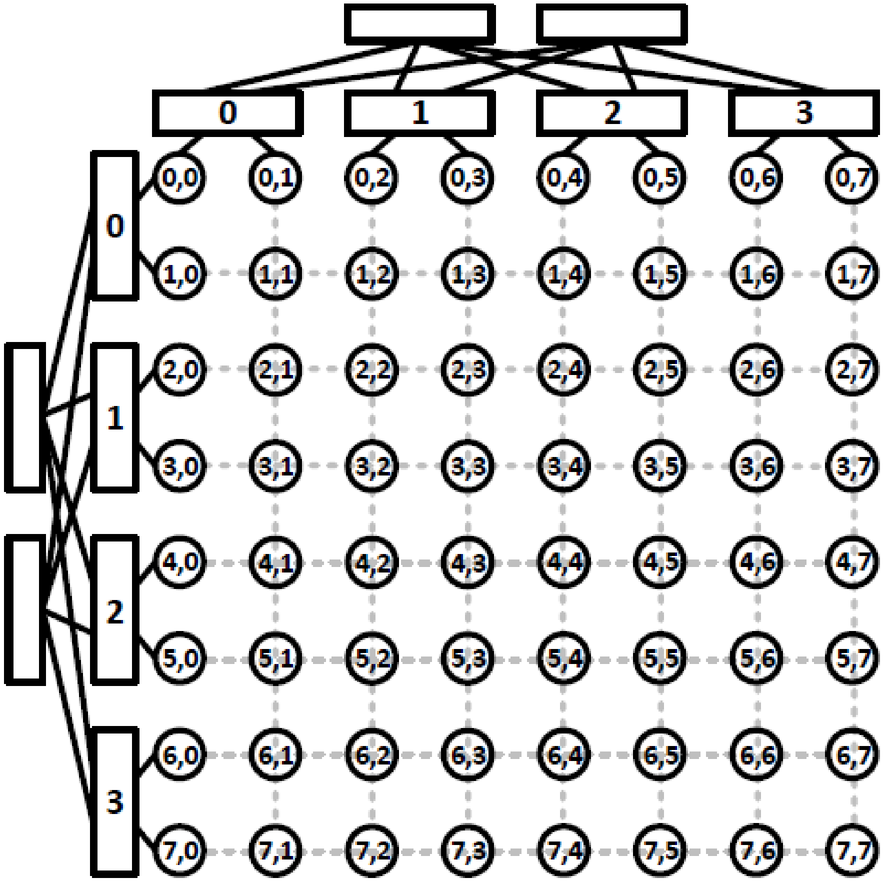
FSquare (n):

Level 1 switch also called Top of Rack (ToR) switch.

Each server connect 2 ToR switches.

$a_{i,j}$'s row ToR switch:

- $\lfloor j/(n/2) \rfloor^{th}$ ToR Switch in *i* row
- $\lfloor i/(n/2) \rfloor^{th}$ ToR Switch in *j* column



Dual-Centric DCNs: FSquare And FRectangle

Routing in FSquare (n):

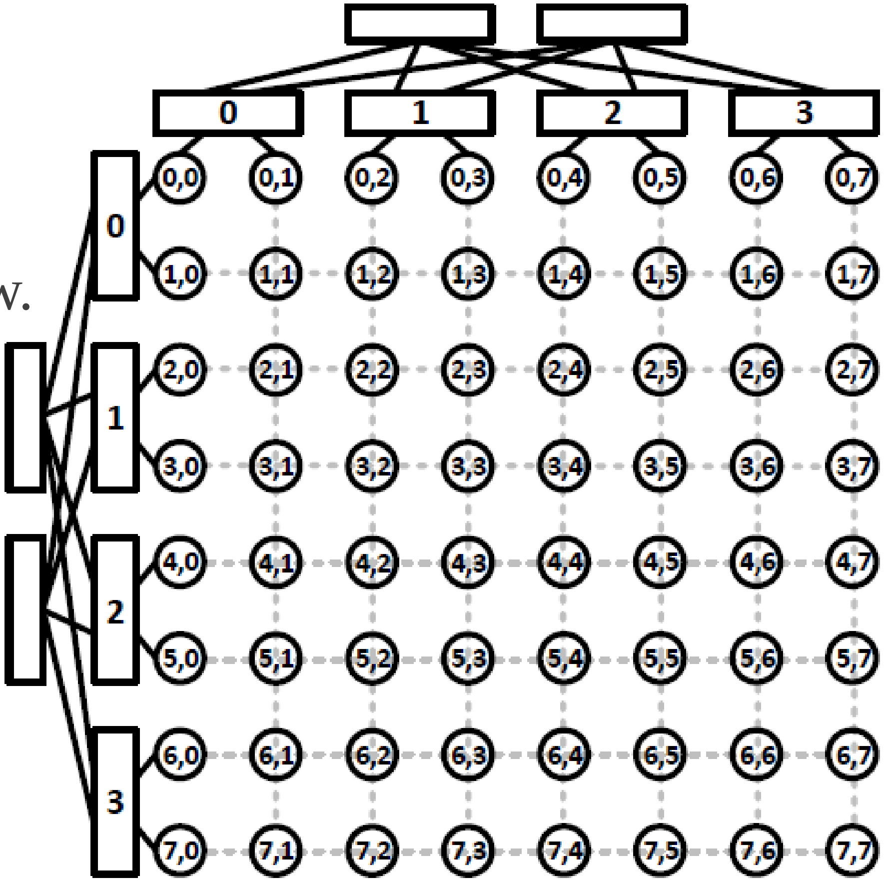
Source : $a_{i,j}$, Destination: $a_{k,l}$

If $i=k$:

the shortest path is within this row.

if $\lfloor j/(n/2) \rfloor = \lfloor l/(n/2) \rfloor$, $a_{i,j}$ and $a_{k,l}$ are connected the same row ToR switch, the shortest path consists of one switch.

If $\lfloor j/(n/2) \rfloor \neq \lfloor l/(n/2) \rfloor$, the shortest path consists of three switches.



Dual-Centric DCNs: FSquare And FRectangle

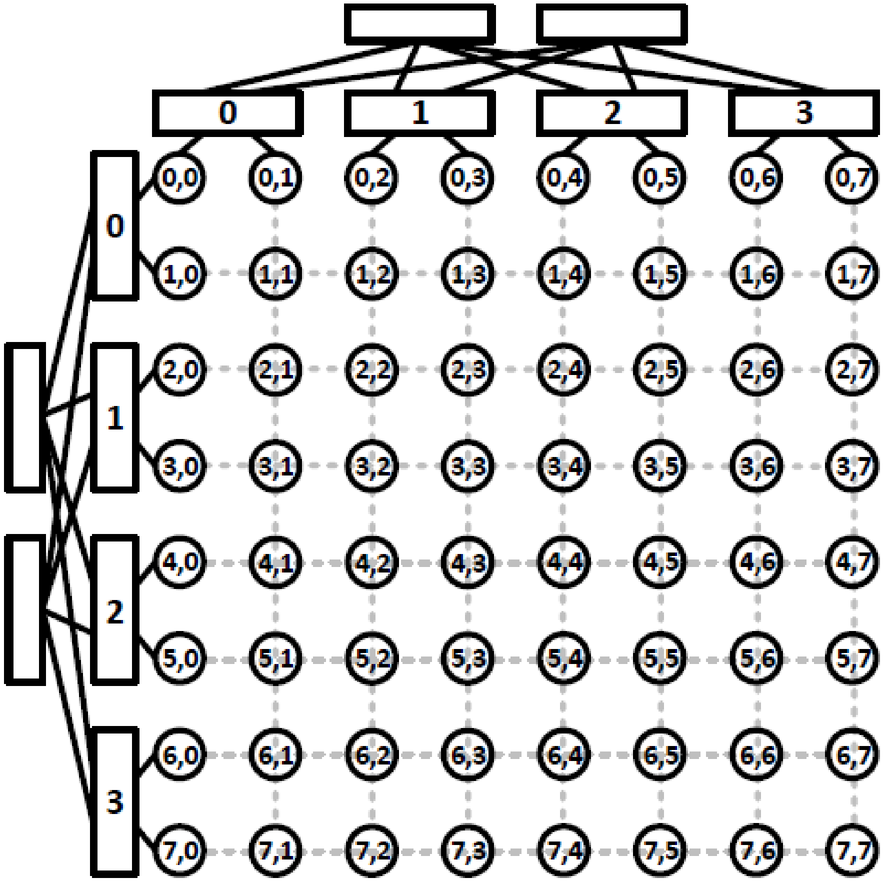
Routing in FSquare (n):

Source : $a_{i,j}$, Destination: $a_{k,l}$

If $i \neq k$ and $j \neq l$:

We can choose one from 2 intermediate servers $a_{i,l}$ and $a_{k,j}$

Row first or column first, or based on traffic condition within the row or column



Dual-Centric DCNs: Fsquare And FRectangle

FSquare Basic Properties

Property 1. *In an FSquare*(n)*, the number of servers is* $N_v = n^4/4$*, and the number of switches is* $N_w = 3n^3/2$.

Property 2. *FSquare*(n) *has a diameter of* $\mathrm{d} = 6d_w + 2d_v$.

Property 3. *The bisection bandwidth of an FSquare*(n) *is* $B = N_v/2$.

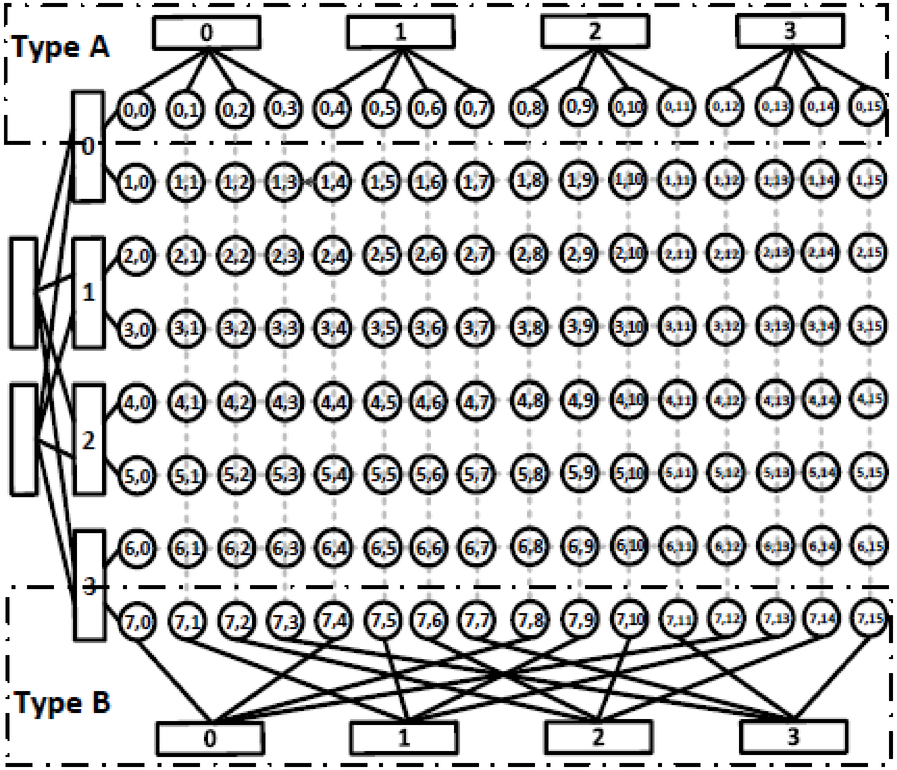
Property 4. *The DCN power consumption per server of an FSquare*(n) *is* $p_V = 6p_w/n + 2p_{nic} + p_{fwd}$.

Dual-Centric DCNs: Fsquare And FRectangle

FRectangle Construction:

The switches and servers in each column form a simple instance of the folded Clos topology.

Switches in each row can adopt Type A or Type B connections.



Dual-Centric DCNs: Fsquare And FRectangle

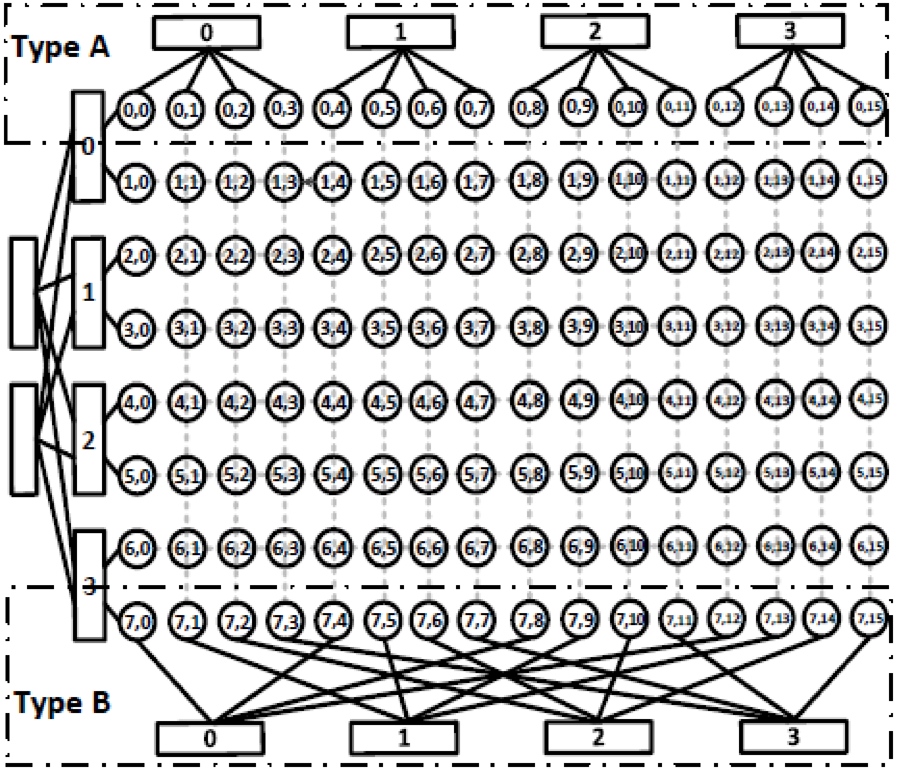
FRectangle Construction:

<u>Type A connections:</u>

For $a_{i,j}$ servers in the i^{th} row, if $kn \le j \le kn+n-1$, $a_{i,j}$ connect to the k^{th} switch.

<u>Type B connections:</u>

For $a_{i,j}$ servers in the i^{th} row, if $j \% n = k$, $a_{i,j}$ connect to the k^{th} switch.



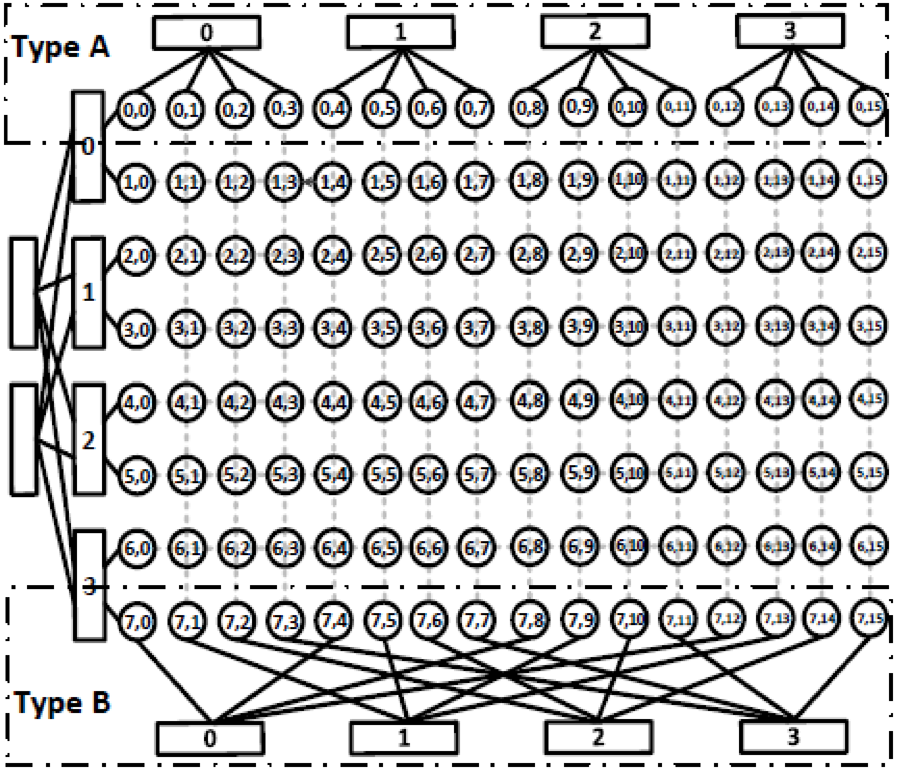
Dual-Centric DCNs: Fsquare And FRectangle

FRectangle Construction:

Let FRectangle choose form the 2 types interconnection in an interleaved fashion.

If i % 2 = 0, type A

If i % 2 = 1, type B



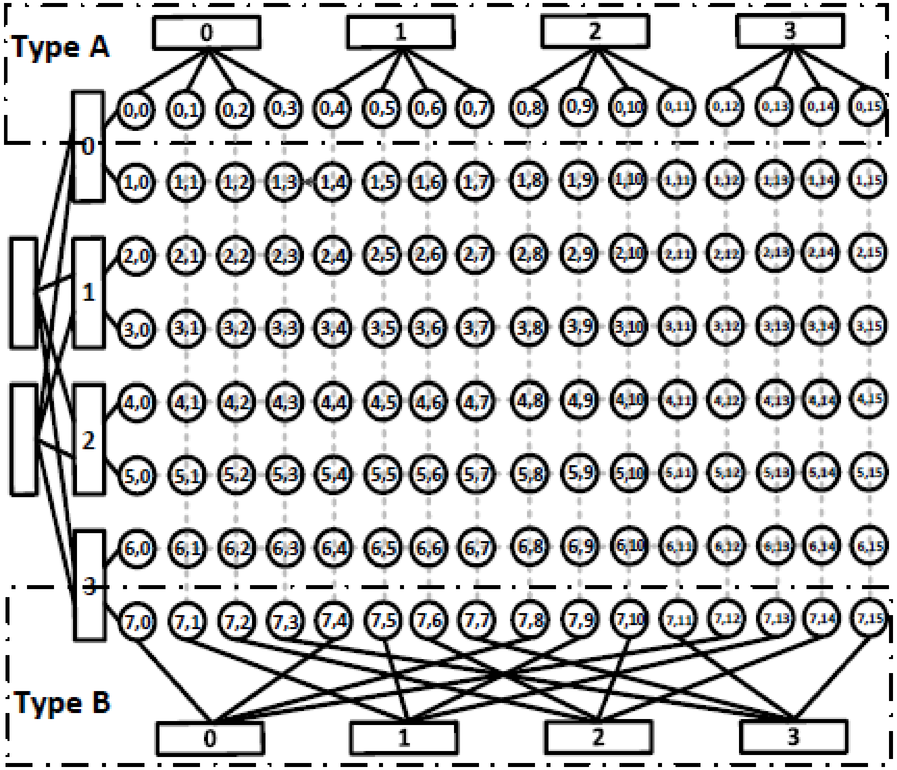
Dual-Centric DCNs: Fsquare And FRectangle

Routing in FRectangle:

Source : $a_{i,j}$, Destination: $a_{k,l}$

If $a_{i,j}$ & $a_{k,l}$ belong to different row types, only need one relay server (along with multiple switches) to forward the packet.

If $a_{i,j}$ & $a_{k,l}$ belong to the same row types, may need three servers to relay the packet.



Dual-Centric DCNs: Fsquare And FRectangle

Routing in FRectangle:

Source : $a_{i,j}$, Destination: $a_{k,l}$

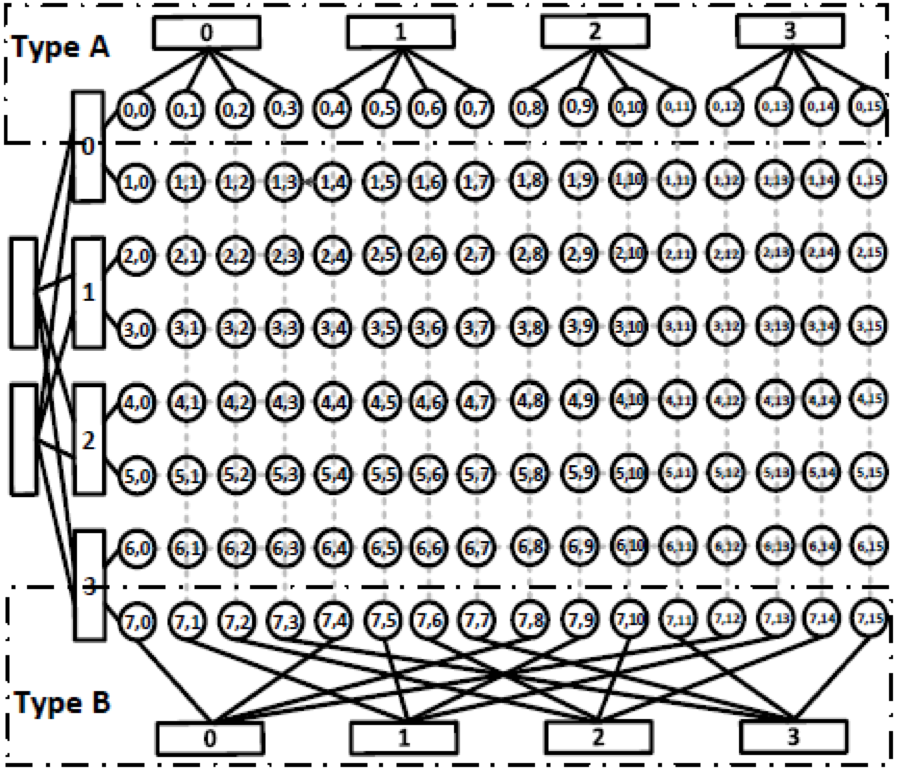
If $i \% 2 = 0$, $k \% 2 = 1$,

Find the column # c^*,

$$c^* = \lfloor j/n \rfloor n + (l \% n)$$

a_{i,c^*} and $a_{i,j}$ connect same row ToR switch; a_{k,c^*} and $a_{k,l}$ connect same row ToR switch.

The shortest path from $a_{i,j}$ to $a_{k,l}$ consists of 3 segments:



$a_{i,j} \rightarrow a_{i,c^*}$, $a_{i,c^*} \rightarrow a_{k,c^*}$, $a_{k,c^*} \rightarrow a_{k,l}$

Dual-Centric DCNs: Fsquare And FRectangle

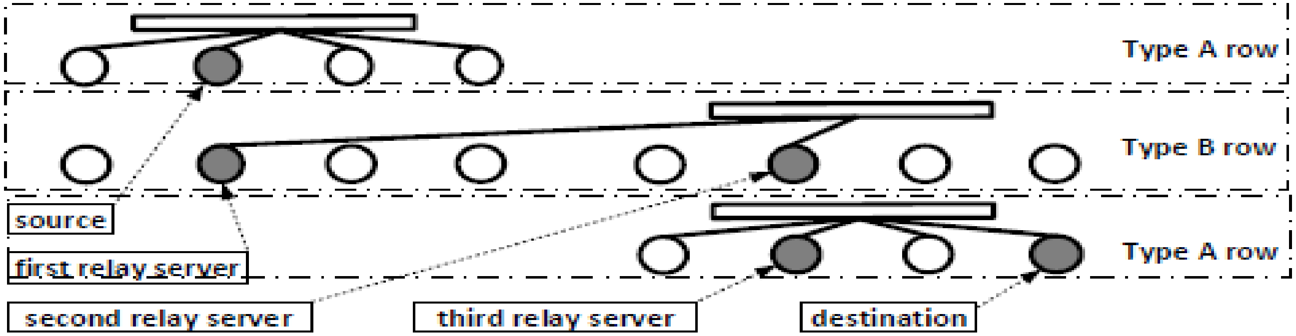
Routing in FRectangle:

Source : $a_{i,j}$, Destination: $a_{k,l}$

<u>If i % 2 = k % 2 = 0 or i % 2 = k % 2 = 1,</u>

$a_{i,j}$ → 1st relay server, 1st relay server → *2nd* relay server

2nd relay server → *3rd* relay server, *3rd* relay server → $a_{k,l}$



Dual-Centric DCNs: Fsquare And FRectangle

FRectangle Basic Properties

Property 1. *In an FSquare*(n), *the number of servers is* $N_v = n^4/4$, *and the number of switches is* $N_w = 3n^3/2$.

Property 2. *FSquare*(n) *has a diameter of* $\mathrm{d}{=}6d_w{+}2d_v$.

Property 3. *The bisection bandwidth of an FSquare*(n) *is* $B{=}N_v/2$.

Property 4. *The DCN power consumption per server of an FSquare*(n) *is* $p_V{=}6p_w/n + 2p_{nic} + p_{fwd}$.

Dual-Centric DCNs: Fsquare And FRectangle

FRectangle Basic Properties

Property 5. *In an FRectangle*(n)*, the number of servers is* $N_v = n^4/2$*, and the number of switches is* $N_w = 2n^3$.

Property 6. *FRectangle*(n) *has a diameter of* $d = 6d_w + 4d_v$.

Property 7. *The bisection bandwidth of an FRectangle*(n) *is* $B = N_v/4$.

Property 8. *The DCN power consumption per server of an FRectangle*(n) *is* $p_V = 4p_w/n + 2p_{nic} + p_{fwd}$.

Comparisons Of Various DCN Architectures

Some Existing Architectures

Switch-centric

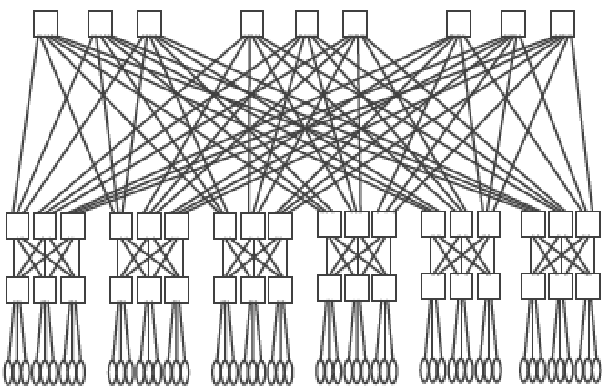
Fat-tree (FDCL)

Fattened Butterfly (FBFLY)

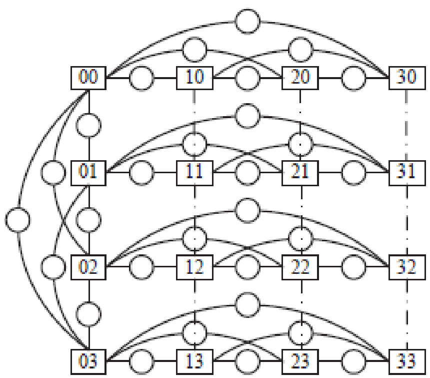
Server-centric

BCube, SWCube,

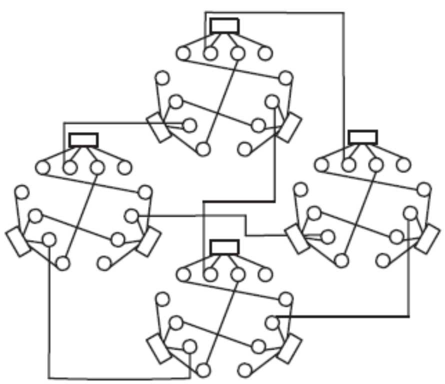
DPillar, DCell, FiConn



Fat-Tree



SWCube



FiConn

Comparisons Of Various DCN Architectures

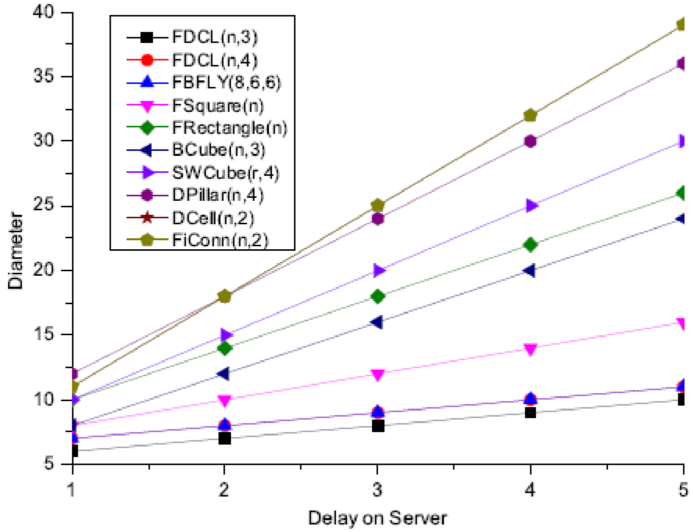
TABLE I
COMPARISON OF VARIOUS DCN ARCHITECTURES

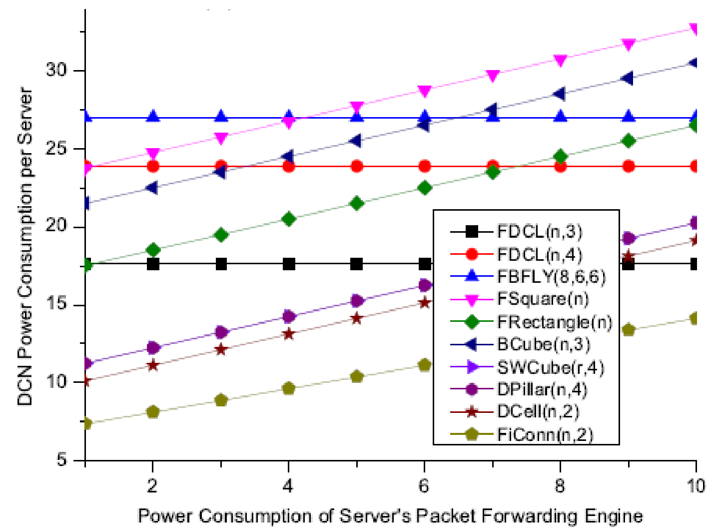
	$N_v(n{=}24)$	$N_v(n{=}48)$	N_w/N_v	d	B	p_V
FDCL$(n,3)$	3,456	27,648	$5/n$	$5d_w{+}d_v$	$N_v/2$	$5p_w/n + p_{nic}$
FDCL$(n,4)$	41,472	663,552	$7/n$	$7d_w{+}d_v$	$N_v/2$	$7p_w/n + p_{nic}$
FBFLY$(4,7,3)$	49,125	—	$8/24$	$8d_w{+}d_v$	$N_v/3$	$8p_w/n + p_{nic}$
FBFLY$(8,6,6)$	—	1,572,864	$8/48$	$7d_w{+}d_v$	$N_v/3$	$8p_w/n + p_{nic}$
FSquare(n)	82,944	1,327,104	$6/n$	$6d_w{+}2d_v$	$N_v/2$	$6p_w/n + 2p_{nic} + p_{fwd}$
FRectangle(n)	165,888	2,654,208	$4/n$	$6d_w{+}4d_v$	$N_v/4$	$4p_w/n + 2p_{nic} + p_{fwd}$
BCube$(n,3)$	331,776	5,308,416	$4/n$	$4d_w{+}4d_v$	$N_v/2$	$4p_w/n + 4p_{nic} + p_{fwd}$
SWCube$(r,4)$	28,812	685,464	$2/n$	$5d_w{+}5d_v$	$(N_v/8) \times r/(r-1)$	$2p_w/n + 2p_{nic} + p_{fwd}$
DPillar$(n,4)$	82,944	1,327,104	$2/n$	$6d_w{+}6d_v$	$N_v/4$	$2p_w/n + 2p_{nic} + p_{fwd}$
DCell$(n,2)$	360,600	5,534,256	$1/n$	$4d_w{+}7d_v$	$> N_v/(4\log_n N_v)$	$p_w/n + 3p_{nic} + p_{fwd}$
FiConn$(n,2)$	24,648	361,200	$1/n$	$4d_w{+}7d_v$	$> N_v/16$	$p_w/n + 7p_{nic}/4 + 3p_{fwd}/4$

From top (switch-centric) to bottom (server-centric),

- DCN power consumption per server decreases;
- Performances of architectures (bisection bandwidth) decreases;
- Diameter increasing

Comparisons Of Various DCN Architectures





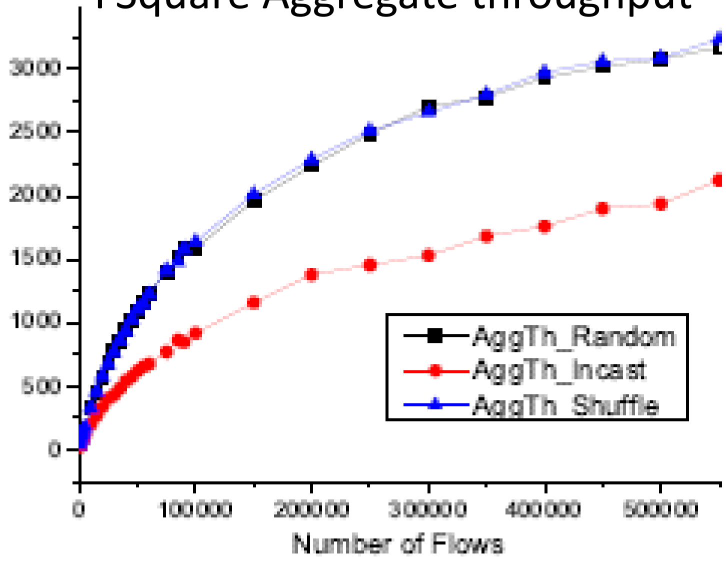
FSquare: lower diameter than all server-centric architectures; Larger bisection bandwidth than switch-centric; high power consumption.

FRectangle: less power consumption than switch-centric; less diameter than most of server-centric; Larger bisection bandwidth than most of server-centric;

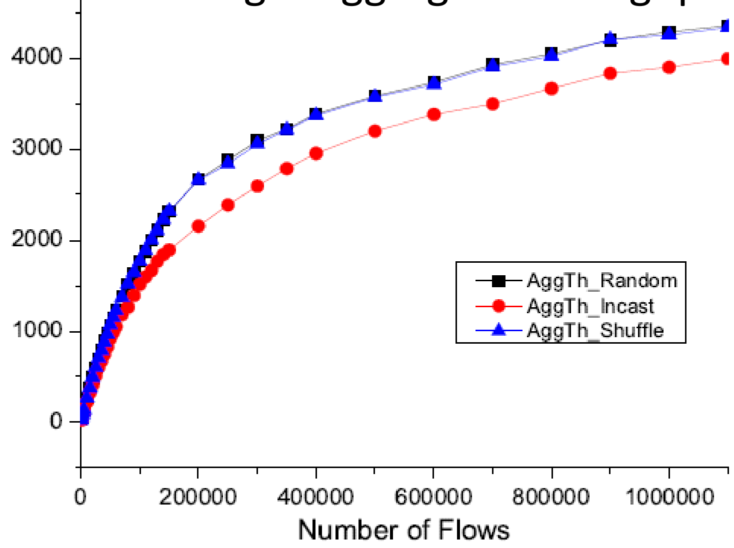
Simulations

We conduct simulations on FCell for both random traffic and bursty traffic.

FSquare Aggregate throughput



FRectangle Aggregate throughput



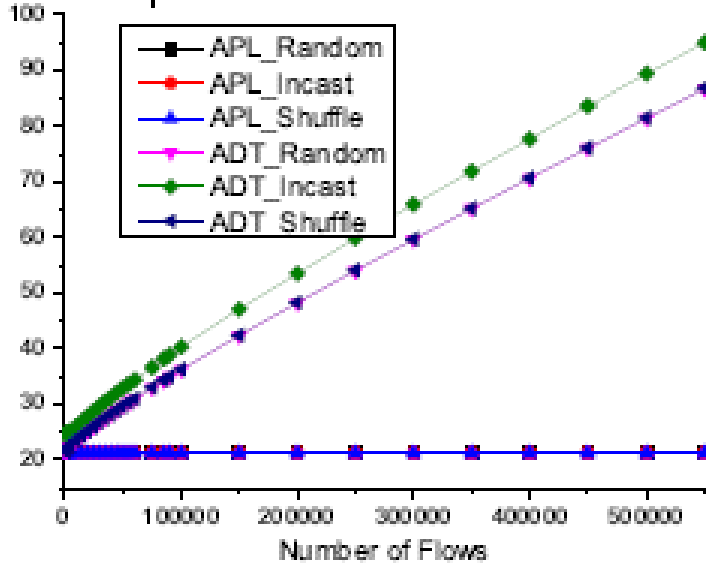
when flows number is **small**, the AggTh values increase almost linearly.

when flows number is **large**, the increasing rates of the AggTh becomes smaller and smaller. (Network become congested)

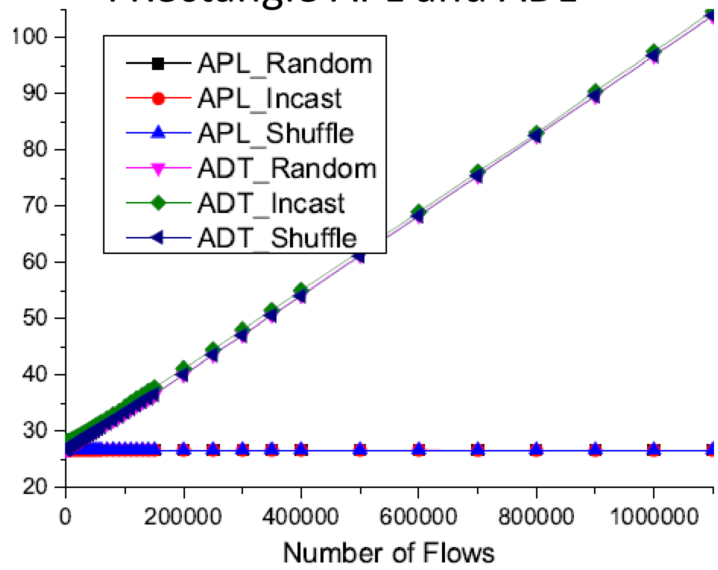
Simulations

We conduct simulations on FCell for both random traffic and bursty traffic.

FSquare APL and ADL



FRectangle APL and ADL



when flows number increases significantly and become congested, ADTs in both architectures only increase linearly.

Conclusion and Future Works

In this paper, we formally introduce a new category of DCN architectures: the dual-centric DCN architectures, where outing intelligence can be placed on both switches and servers.

We propose two typical dual-centric DCN architectures: FSquare and FRectangle.

By comparing them with existing architectures and by investigating themselves, we show that these two dual-centric DCN architectures have various nice properties for practical data centers, and provide flexible choices in designing DCN architectures.

The proposed dual-centric design philosophy will certainly become a potential candidate in future DCN architecture designs.

Conclusion and Future Works

Future works can be cast in, but are not limited to, the following directions:

- 1.) designing efficient and/or adaptive routing schemes for FSquare and FRectangle;
- 2.) exploring other possible dual-centric architectures that also have appealing properties;
- 3.) designing dual-centric architectures where each server uses more than 2 NIC ports; and
- 4.) exploring the limitations of the dual-centric design philosophy, and how to control and apply them in practical DCN designs.

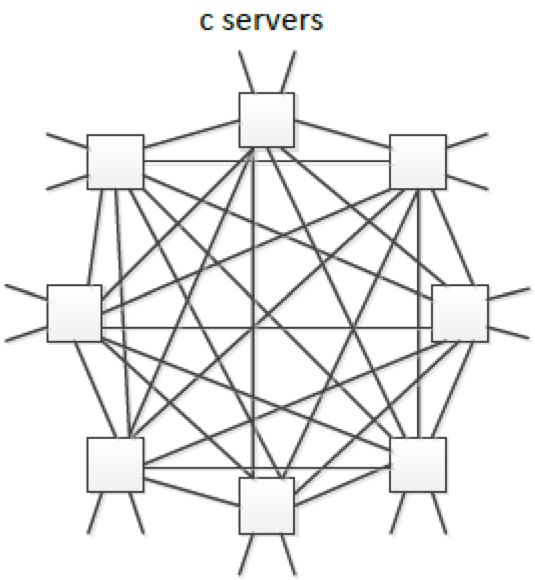
Thank you!

Questions can be sent to:

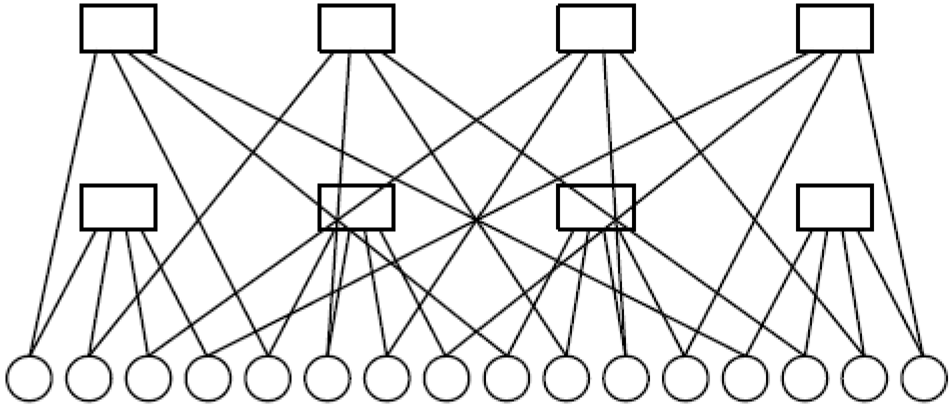
dawei.li@temple.edu

Backup slides

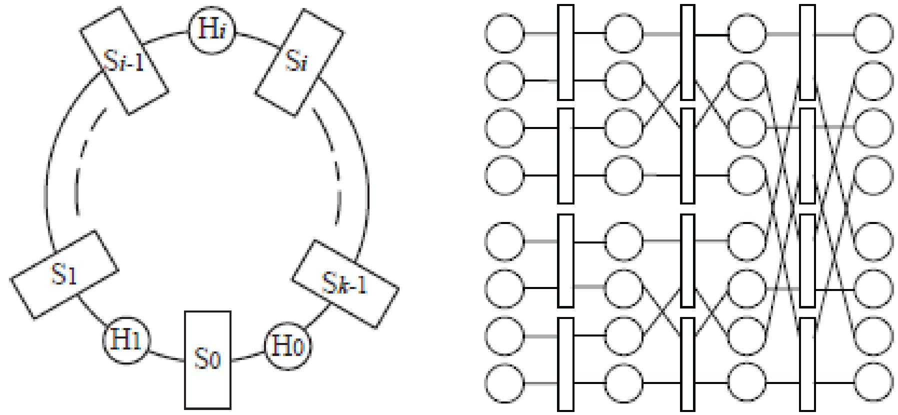
Flattened Butterfly (one dimensional)



BCube (two level):



DPillar:



(a) vertical view (b) horizontal view

DCell (two level):