FCell: Towards the Tradeoffs in Designing Data Center Network (DCN) Architectures

DAWEI LI AND JIE WU

CIS, TEMPLE UNIVERSITY

ICCCN 2015

Agenda

Introduction

Unified Performance Model

FCell: A Novel DCN Architecture

Comparisons of DCN Architecture

Simulation

Conclusions

My Research on Network Connections

Interconnection Networks (1988-1998)
- Direct networks (no switch)
- Multistage networks (with 2X2 switches)

MANETs (1999-2005)
- Topology control (to control density of neighbors)
- Maintaining “long-distance” links based on small world

DTNs (2005-now)
- Mobility control (for contact distribution and location)

DCNs (2010-now)
- Unifying connection models using servers/switches

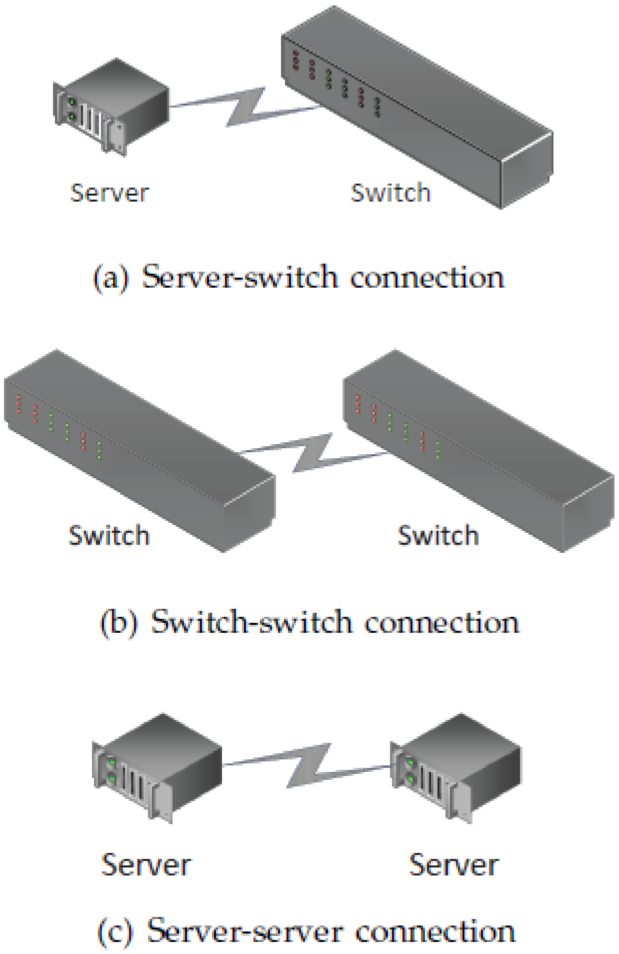
Introduction

Three types of connections:

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

Two classes of DCNs :

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



Introduction

Switch-centric vs. Server-centric

- Server-centric architectures
  - Enjoy the high programmability of servers, but servers usually have larger processing delays than do switches.
- Switch-centric architectures
  - Enjoy the fast switching capability of switches, but switches are less programmable than servers.
- Can we combine the advantages of both categories?

Introduction

Performance vs. Power Consumption

- To provide low end-to-end delays and high bisection bandwidth
  - Large numbers of networking devices are usually used in DCNs.
  - E.g., Fat-Tree: three levels of switches; BCube: three or more levels & extra Network Interface Card (NIC) ports.
- To achieve a low DCN power consumption
  - Other architectures use significantly fewer networking devices.
  - E.g., FiConn, Dpillar, etc.
- Can we achieve high performances and low power consumption at the same time?

Introduction

Scalability vs. Flexibility

- Scalability : networking devices, typically the switches, rely on a small amount of info., which does not increase significantly over the network size, to make efficient routing decisions.
- Flexibility: expanding the network in a fine-grained fashion should not destroy the current architecture
- Can we design both scalable and flexible DCN architectures?

Introduction

Contributions

- Unified performance model
  - Path length (and hence, diameter)
  - Power consumption
- A range of DCN architectures
  - Based on different trade-offs
- A new DCN architecture: Fcell
  - Situated in the middle of the trade-off spectrum: dual-centric

Unified Performance Model

- Unified Path Length Definition:

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

$n_{P,w}$: # of switches in a path
$n_{P,v}$: # of servers in a path (excluding *s* and *d*)
d_w : processing delay on a switch
d_v : processing delay on a server

- Unified Diameter in a DCN:

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

Unified Performance Model

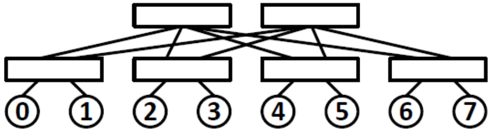
- DCN power consumption per server:

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

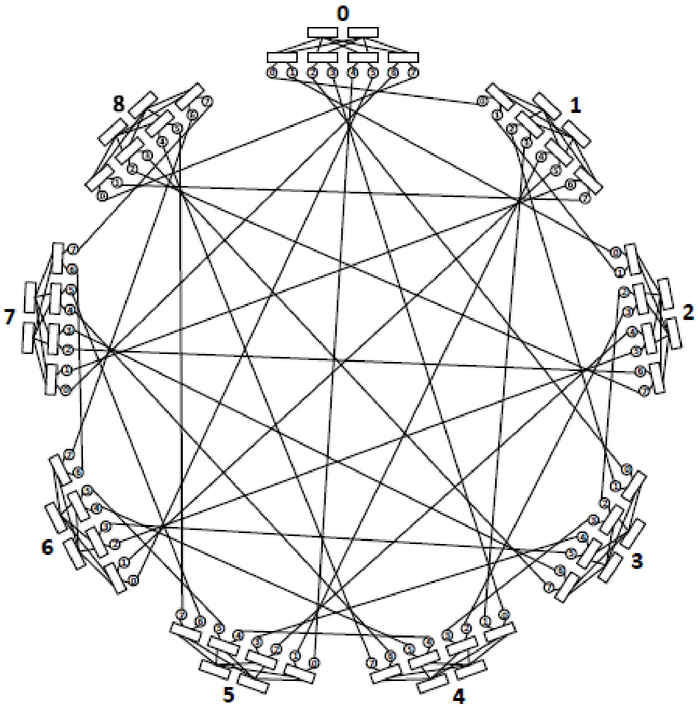
p_w : power consumption of a switch
N_w : # of switches in a DCN
N_v : # of servers in a DCN
n_{nic} : average # of NIC ports each server uses
p_{nic} : power consumption of a NIC port
α : whether the server is involved in packet relaying
p_{fwd} : power consumption of a server’s packet forwarding

FCell: A Novel DCN Architecture

- Intra-cluster
  - The switches and servers form a simple instance of the folded Clos topology
- Inter-cluster
  - Each of the servers in a cluster is directly connected to another server in each of the other clusters
- 2 NIC ports and switches with n ports
  - n/2 level-2 switches and n l-1 switches
  - (n/2)n servers in each cluster
  - Total (n/2)n+1 clusters



(a) The interconnections in one cluster.



(b) Final interconnections of FCell(4).

FCell: A Novel DCN Architecture

- FCell basic properties:

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

Proof. There are $n^2/2+1$ clusters, each with $3n/2$ switches and $n^2/2$ servers. □

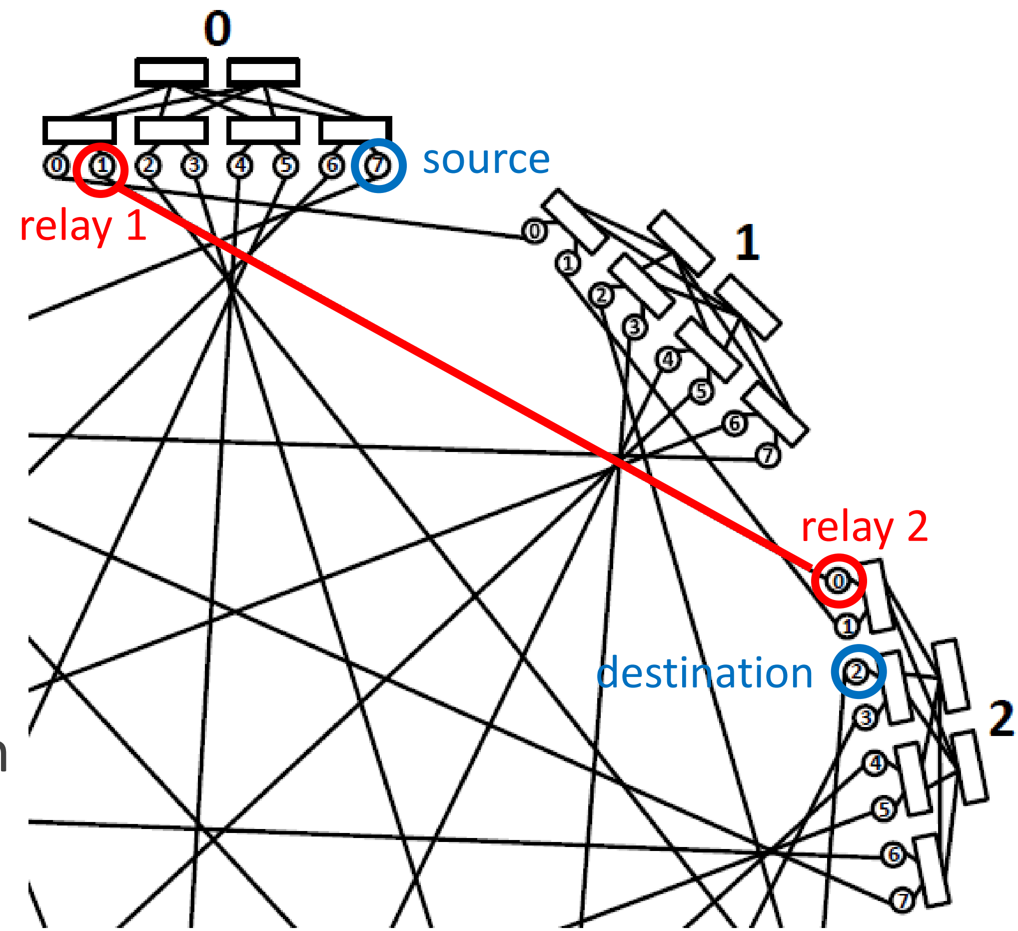
Property 2. *The diameter of an FCell(n) is* $d = 6d_w + 3d_v$.

Property 3. *The bisection bandwidth of an FCell(n) is* $B \approx N_v/4$.

Property 4. *The DCN power consumption per server of an FCell(n) is* $p_V = 3p_w/n + 2p_{nic} + p_{fwd}$.

FCell: A Novel DCN Architecture

- FCell routing schemes
  - Shortest Path Routing:
    - Determines the relay servers
    - Source to relay1 in the source cluster
    - Relay 1 to relay 2
    - Relay 2 to destination in the destination cluster



FCell: A Novel DCN Architecture

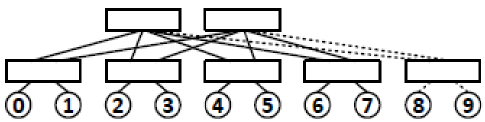
- Detour Routing:
  - Randomly select a relay cluster
  - Conduct shortest path routing from the source cluster to the relay cluster
  - Then, from a relay cluster to the destination cluster

FCell: A Novel DCN Architecture

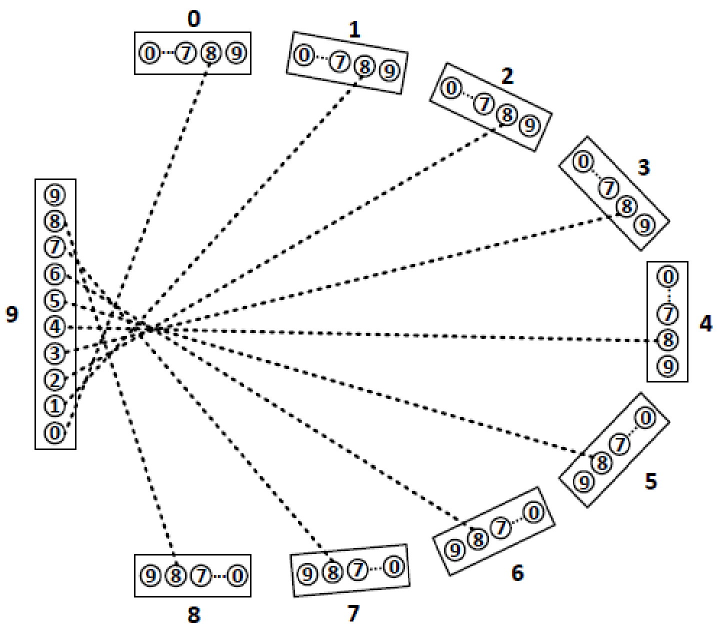
- FCell Scalability and Flexibility
  - FCell has good scalability due to its high degree of regularity.
  - Switches in FCell only need local information for packet forwarding.
  - Servers only need basic configuration parameters of FCell for packet forwarding.

FCell: A Novel DCN Architecture

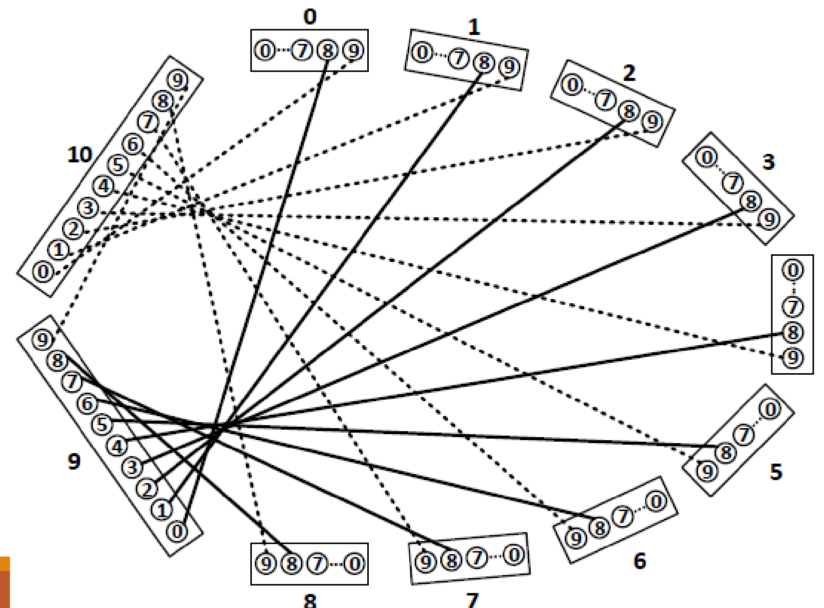
- FCell supports flexibility well, i.e., it allows fine-grained and incremental growth of its network size.



(a) Adding one rack of $n/2$ servers in each cluster.



(b) Adding the first expanded cluster.



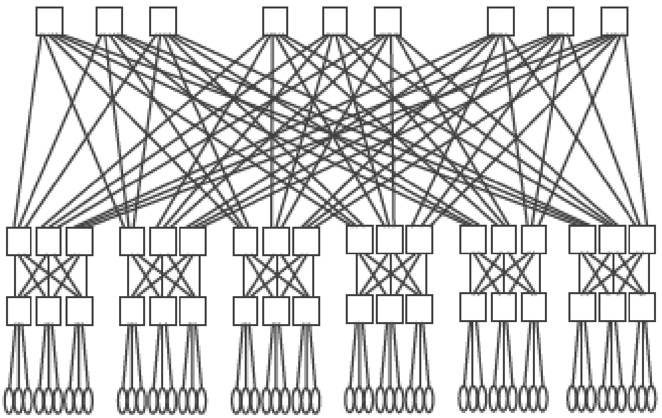
(c) Adding the second expanded cluster.

Comparisons of DCN Architectures

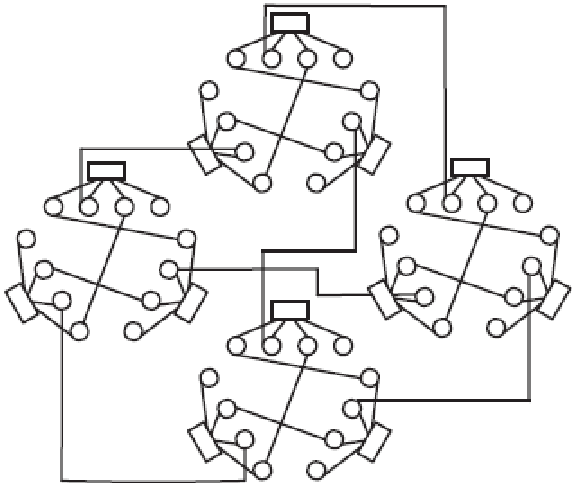
Some existing architectures:

Left: server-centric, Right: switch-centric

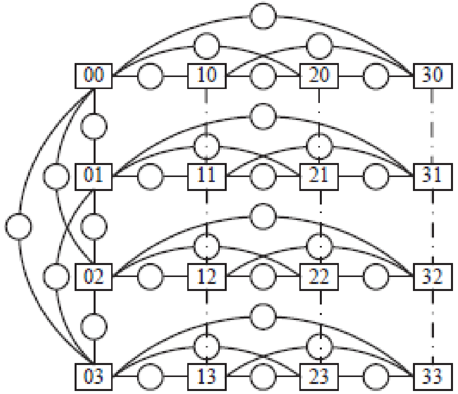
Fat-Tree:



FiConn:



SWCube:



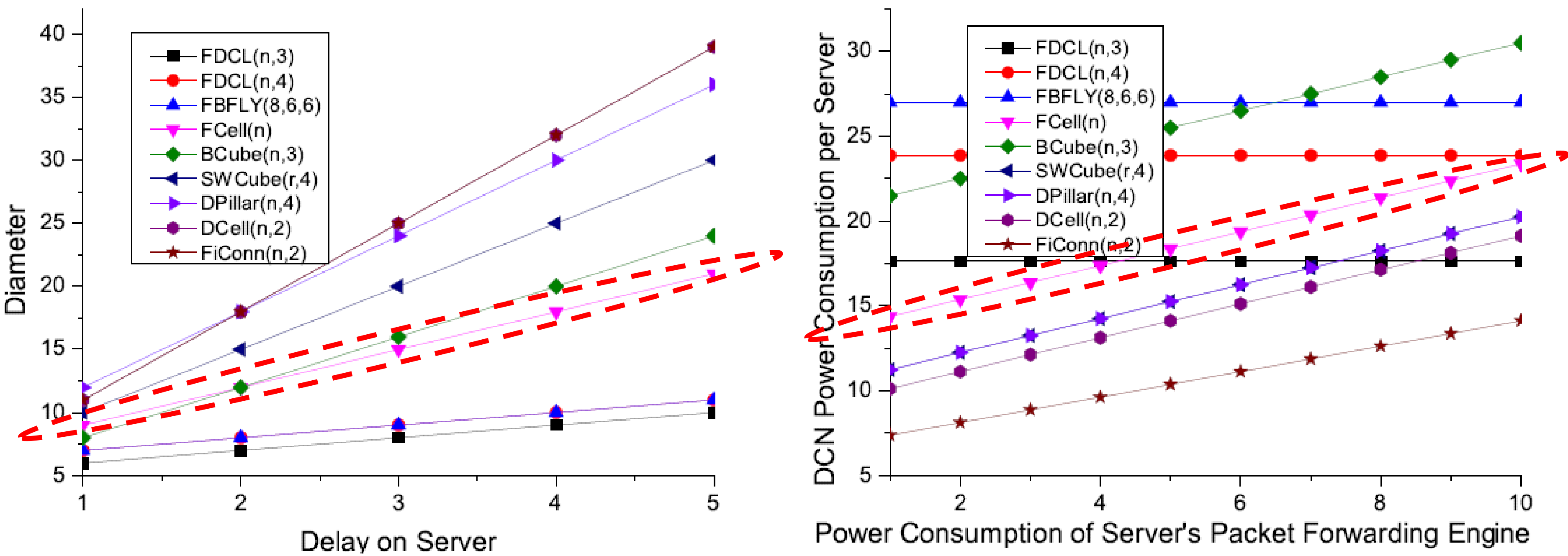
(b) 2D SWCube

Comparisons of DCN Architectures

TABLE I
COMPARISON OF VARIOUS DCN ARCHITECTURES

	$N_v(n{=}24)$	$N_v(n{=}48)$	N_w/N_v	d	B	pV
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	—	$8/24$	$8d_w{+}d_v$	$N_v/3$	$8p_w/24 + p_{nic}$
FBFLY$(8,6,6)$	—	1,572,864	$8/48$	$7d_w{+}d_v$	$N_v/3$	$8p_w/48 + p_{nic}$
FCell(n)	83,232	1,328,256	$3/n$	$6d_w{+}3d_v$	$N_v/4$	$3p_w/n + 2p_{nic} + p_{fwd}$
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	$1/n$	$4d_w{+}7d_v$	$> N_v/16$	$p_w/n + 7p_{nic}/4 + 3p_{fwd}/4$

Comparisons of DCN Architectures



(a) Diameters vs. d_v.

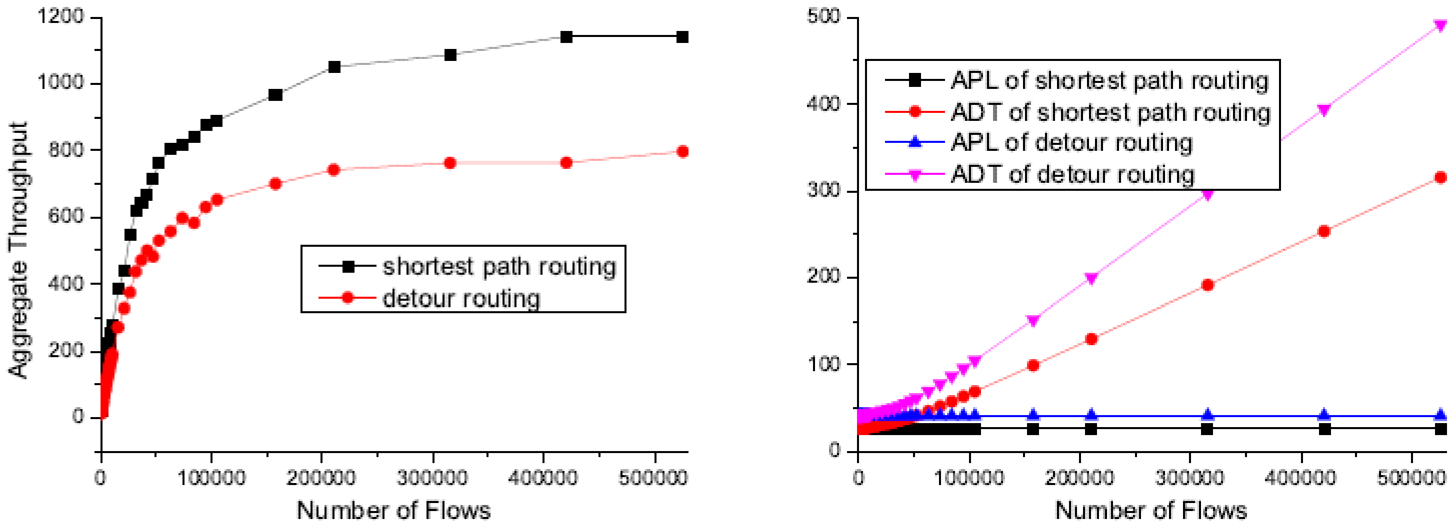
(b) p_V vs. p_{fwd}.

Simulation

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

- Average Path Length (APL)
- Average Delivery Time (ADT)

Simulations for random traffic:

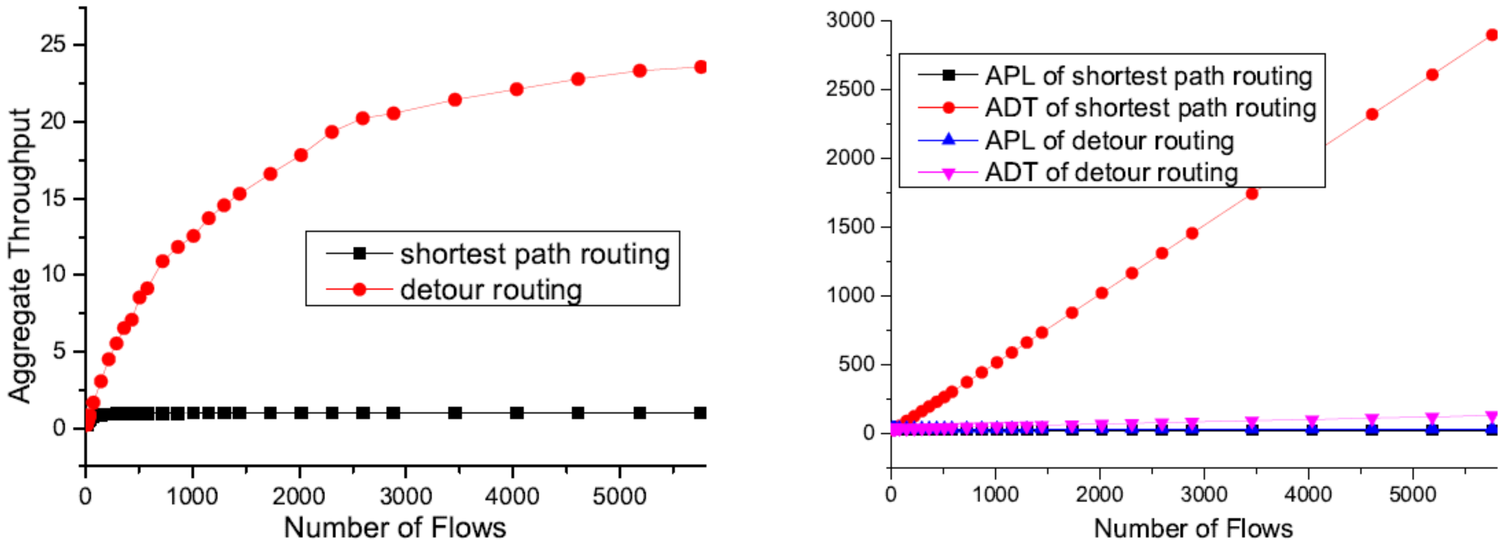


(a) Aggregate throughput. (b) APL and ADT.

Fig. 4. Aggregate throughput, APL and ADT vs. No. of flows (random traffic).

Simulation

Simulations for bursty traffic:



(a) Aggregate throughput. (b) APL and ADT.

Fig. 5. Aggregate throughput, APL and ADT vs. No. of flows (bursty traffic).

Conclusions

- A **unified path length definition** and a **unified power consumption model** for general DCNs
  - Enabling **fair** and **meaningful** comparisons
- A novel DCN architecture, FCell, which serves as a good example of **a tradeoff design in three aspects**
  - Performance and power, switch-centric and server-centric designs, and scalability and flexibility
- A new class of DCNs, that can be regarded as **dual-centric**, with FCell as an example
  - Two basic routing schemes
  - Performance under different traffic conditions

Future Work

- More in-depth simulation
  - Different flows
  - Different bursty modes
- Simulation of some real applications
- Support for overlay networks

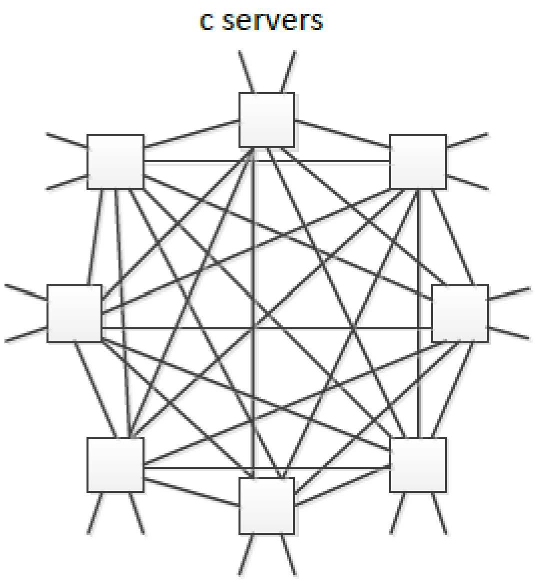
Questions can be sent to:

dawei.li@temple.edu

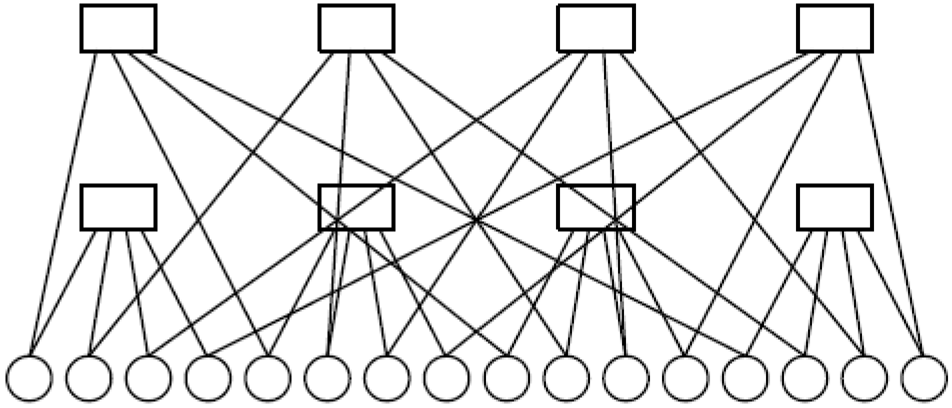
jiewu@temple.edu

Backup slides

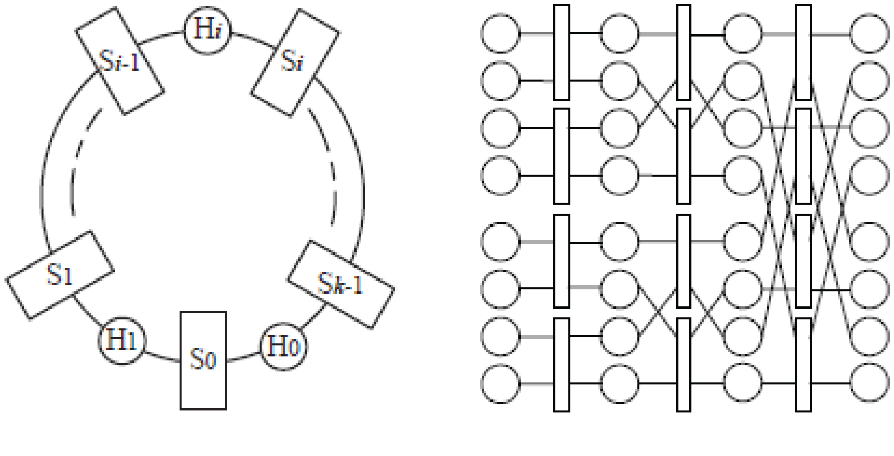
Flattened Butterfly (one-dimensional)



BCube (two-level):



DPillar:



(a) vertical view (b) horizontal view

DCell (two-level):