Time Synchronization For An Optically Groomed Data Center Network

Tanjila Ahmed

Friday Group Meeting

Feb 24, 2017

UCDAVIS
UNIVERSITY OF CALIFORNIA

Time Synchronization Mechanisms for an Optically Groomed Data Center Network

Ganesh C. Sankaran and Krishna M. Sivalingam,

Department of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai, INDIA 2HCL Technologies Ltd, Chennai, INDIA

Performance Computing and Communications Conference (IPCCC), 2016 IEEE 35th International, 9-11 Dec. 2016

UCDAVIS

Agenda

- Objective
- Highlights of the Architecture
- Features of the Communication Scheme
- Time Synchronization Schemes
  - ➢ Continuous Time Approach
  - ➢ Discrete-time based approach
- Observation
- Performance Evaluation
- Comparison of Time Synchronization Schemes
- Conclusion

UCDAVIS

Objective

- Optically groomed data center network(OGDCN)
- Power efficient hybrid optical-packet switched network.
- Time synchronization aspect of transmission scheduling in OGDCN.
- Two schemes - Continuous and Discrete time(slotted) based time synchronization.
- Performance evaluation and comparison.

G. C. Sankaran and K. M. Sivalingam, "Optical traffic grooming based data center networks: Node architecture and comparison," *IEEE Journal on Selected Areas in Communications – Series on Green Communica- tions and Networking*, vol. 34, pp. 1618–1630, May 2016.

Highlights of the Architecture

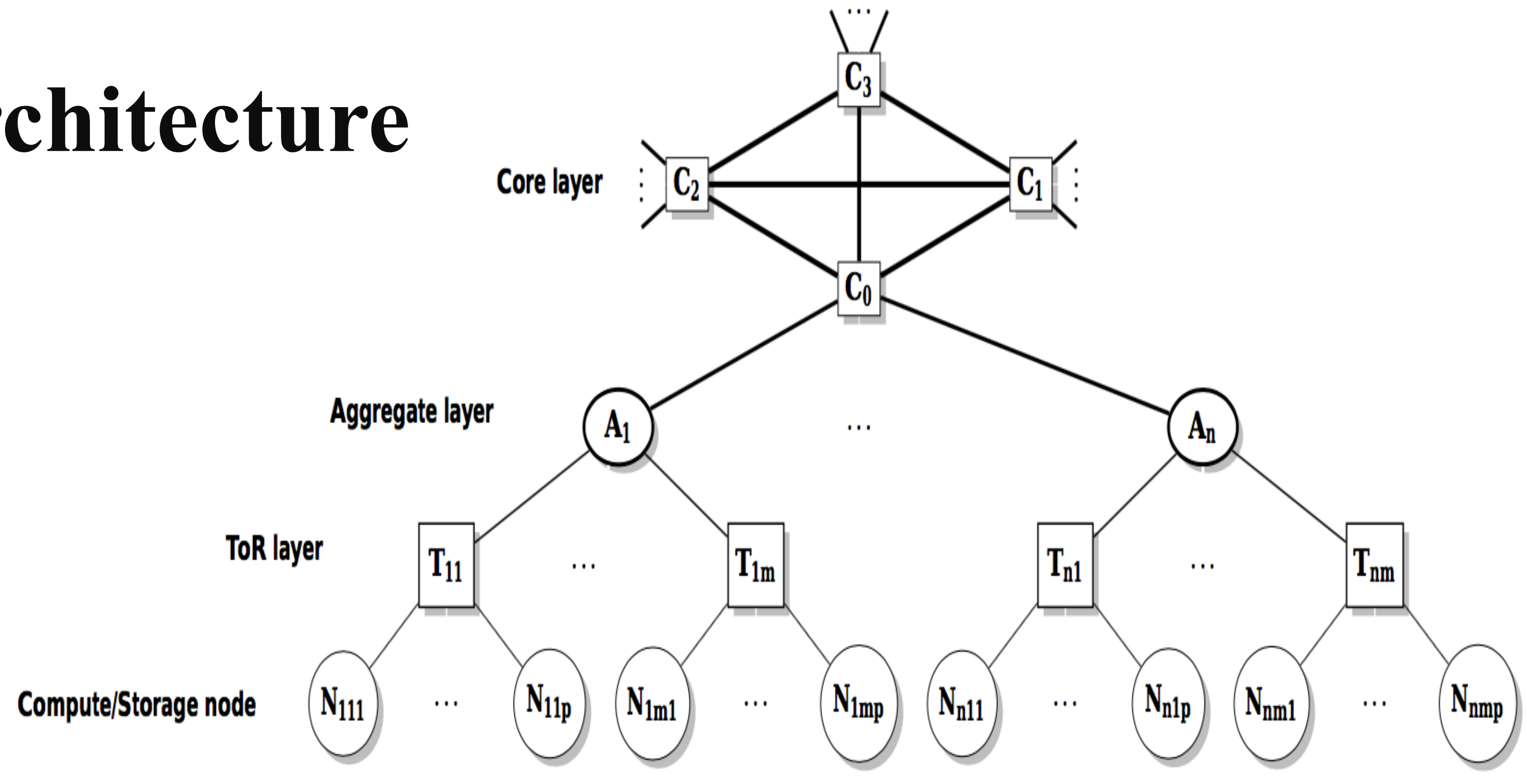


Fig. 1: Three-tier Data Center Network (DCN) topology.

- Interconnected core.
- Each core connected to N aggregate switch.
- Each aggregate switch connected to M TOR switches.
- Each TOR connected to p compute and storage nodes(CSN).
- CSNs are equipped with tunable transceivers only.

UCDAVIS

Features of the Communication Scheme

- Source-destination pair must tune to a predefined wavelength.
- A centralized controller computes the transmission schedule.
- To compute a schedule, all network resources must be available for transmission duration(no network path establishment required).
- This includes transmitter, receiver and the specific wavelength on all link segments along the path.
- Controller uses separate control wavelength to carry control messages(packet switched).
- CSN nodes communicates with each other using WDM and TDM.

UCDAVIS

Features of the Communication Scheme

ation pairs.

hat are separated in time

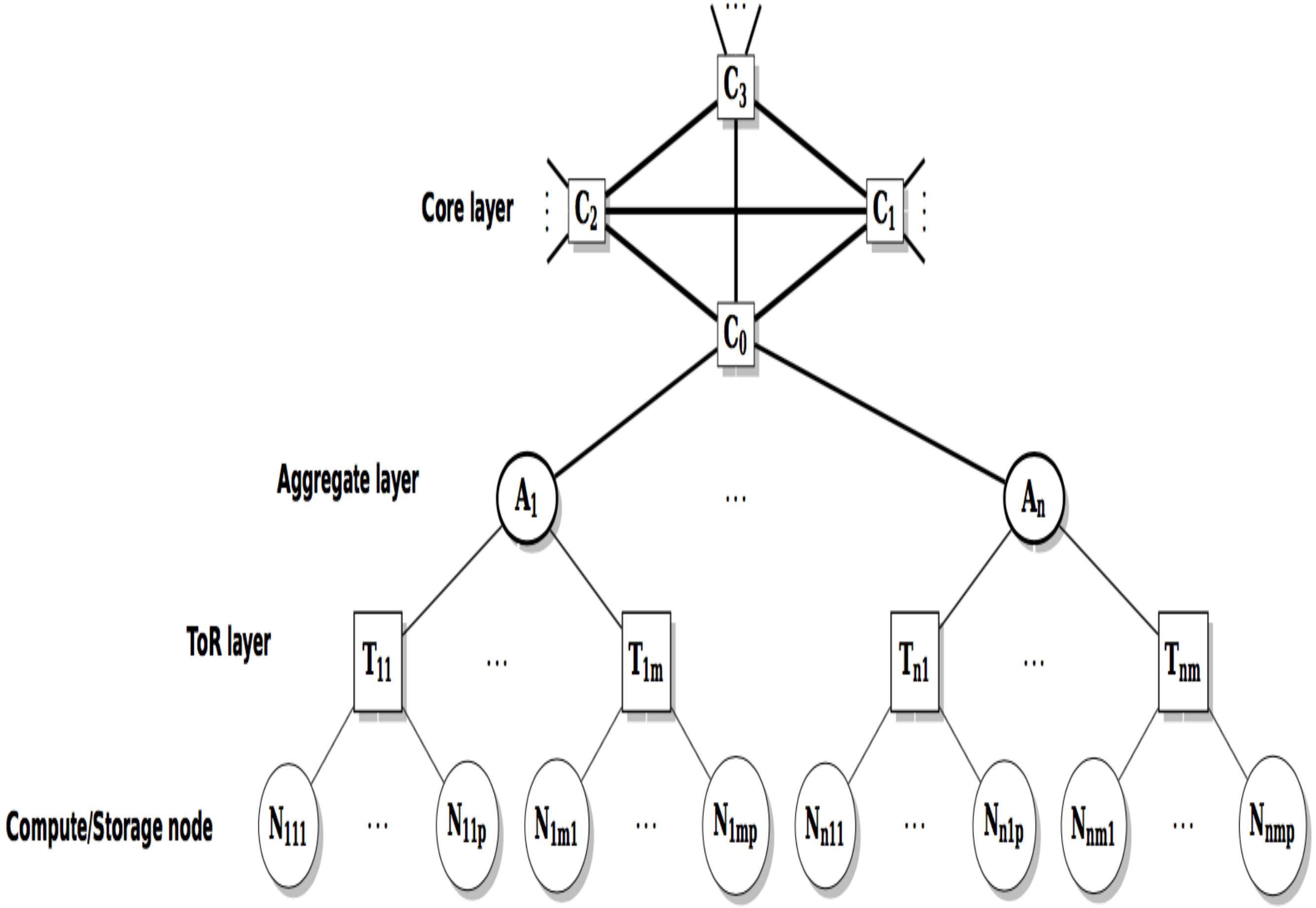


Fig. 1: Three-tier Data Center Network (DCN) topology.

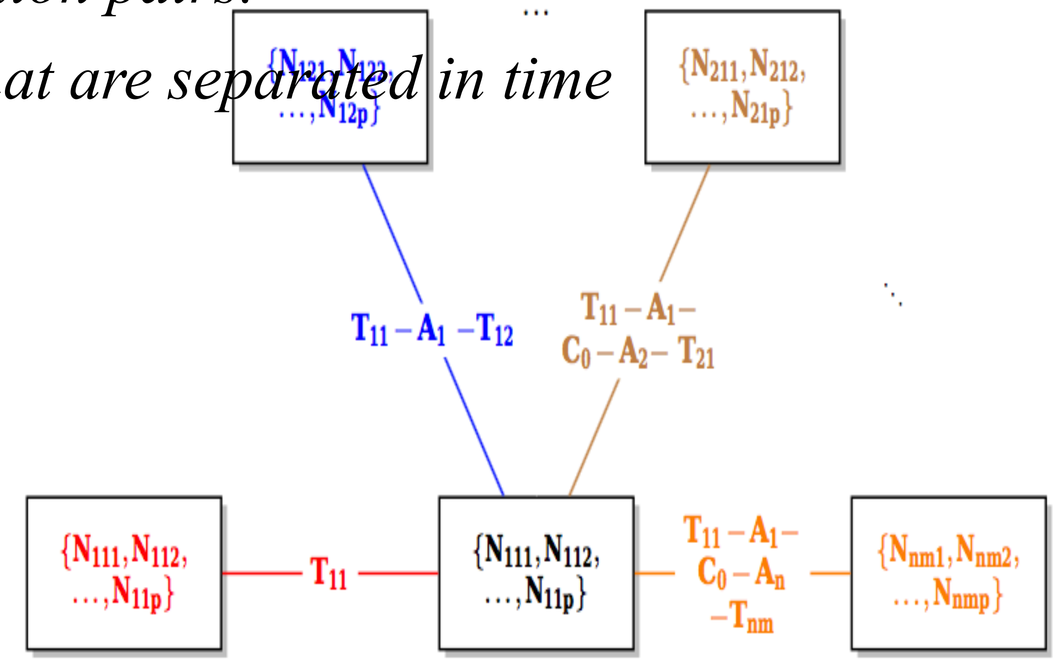


Fig. 2: Shared wavelength circuits for CSNs connected to ToR T_{11}.

UCDAVIS

Time Synchronization Schemes

Continuous Time based approach :

A high precision clock is used to interpret transmission start and finish times consistently.

- Considering propagation delay, $D_b > P_b > S_b$
- $P_b - S_b = \alpha_{ui} + \beta_{ij}$
- $D_b - P_b = \gamma_{jv}$
- $D_b - S_b = \delta_{uv}$
- $\delta_{uv} = \alpha_{ui} + \beta_{ij} + \gamma_{jv}$

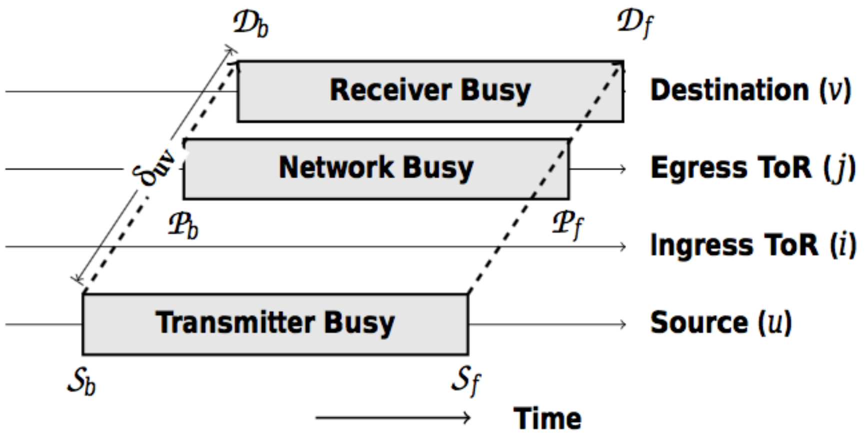


Fig. 4: Impact of propagation delay on resource reservation across the network. Here, $\delta_{uv} = \alpha_{ui} + \beta_{ij} + \gamma_{jv}$.

Continuous Time Approach

- Propagation delays are continuous values, so accurate clocks are required
- Network time protocol/precision time protocol is used

Link Utilization :

$$U_c = 1 - \frac{W}{\bar{L}+W}$$

Maximum wastage ratio

W= rounding-off error in bits. e.g. for 10 Gbps data rate 1 nanosecond accuracy, W= *10 bits*.

$\bar{L}$ = average packet length.

Example: W=10 bits, packet length range 64-8192 bytes(avg.= 4128 bytes), Maximum utilization is 99.9%

UCDAVIS

Discrete Time Approach

Time is divided into slots of constant duration(S).

Slot start times must be consistently interpreted across the network.

- Propagation delay= $\boldsymbol{\delta^*}$ (constant)
- Actual propagation delay=δ_{uv}
- Time slot duration = S
- Effective time slot duration = S'
- r_0, r_1= receiver start time to receive for slot 0,1
- t_0, t_1= transmitter start time to transmit slot 0,1

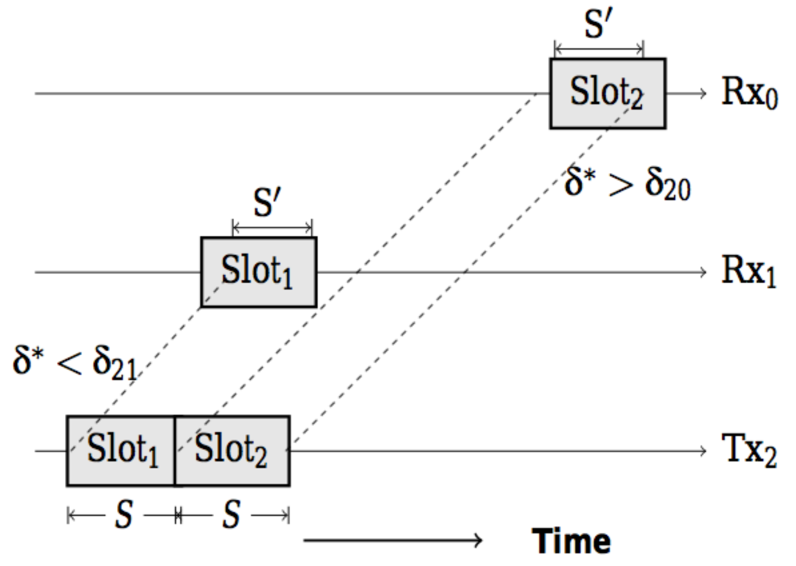


Fig. 5: Effective duration S' less than actual time slot duration S with Discrete time (DT) choice.

Discrete Time Approach

Case 1: $\boldsymbol{\delta}^*<\delta_{uv}$

avoid overlap, the effective transmission duration within the slot

$(r0 + \delta_{21} - \delta^*, r_1)$; S'<S

$\boldsymbol{\delta}^*>\delta_{uv}$ negative time offset

$(r_0 , r_1 + \delta_{20} - \delta^*)$; S'>S

Case 2:Ideal case

(r_0,r_1)

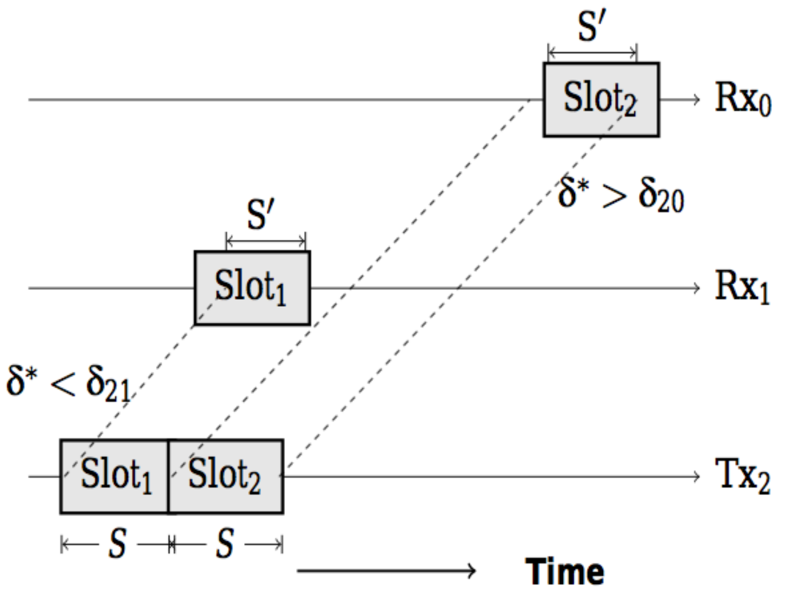


Fig. 5: Effective duration S' less than actual time slot duration S with Discrete time (DT) choice.

Discrete Time Approach

Delay Compensation : A constant propagation delay is considered.

No Delay Compensation : No constant propagation delay is considered.

Link Utilization : $U_d = \frac{\bar{L}}{S}\left[1 - \frac{\sigma}{S}\right]$

$\bar{L}$ = average data transfer length

σ = standard deviation of the propagation delay

S = size of the time slot

Example: slot length= 8192 bytes, avg. data length= 4128 byte, σ=0 (delay compensation case). Utilization cannot exceed 50%, for data transfer length = slot length , this network can be fully utilized.

Observations

Continuous time scheme investigated in DCN

Utilization of CT synchronization depends on per transfer wastage W.

W depends on the accuracy of the clock being used.

Utilization of DT synchronization depends on propagation delay deviation, average transfer length.

UCDAVIS

Performance Evaluation

- Scheduling algorithm : Earliest first & Bitmap heuristic.
- *Throughput = number of shared wavelength circuits x number of wavelength x data rate*
- *Utilization = Busy Duration / Total duration*
- CH= continuous time synchronization with high accuracy (1ns)
- CL= continuous time synchronization with low accuracy (1000ns)

UCDAVIS

Performance of CT Synchronization

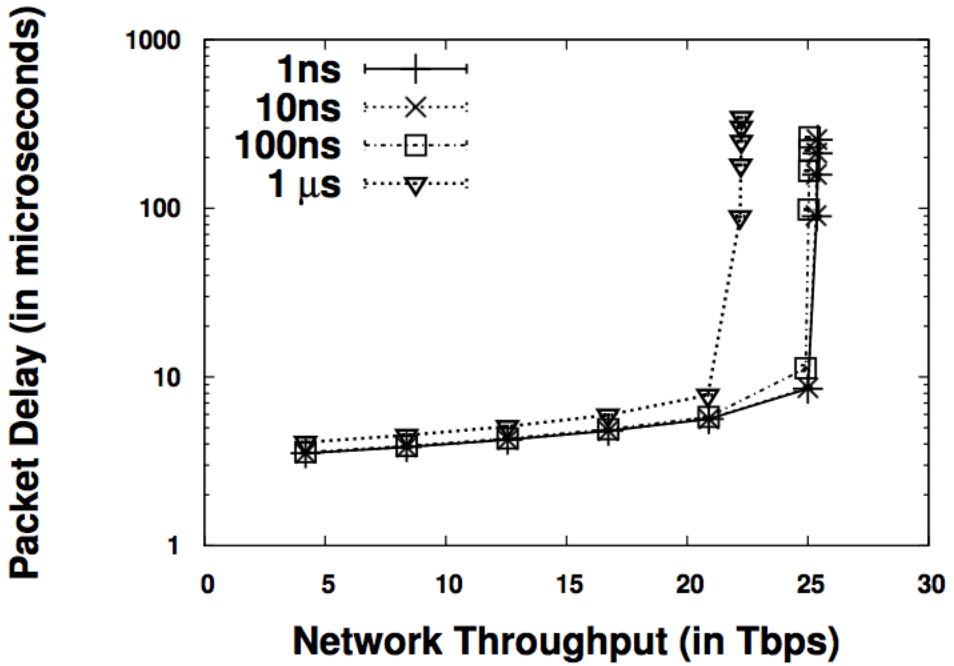


Fig. 6: Throughput and delay of continuous-time (CT) scheme with varying time accuracy.

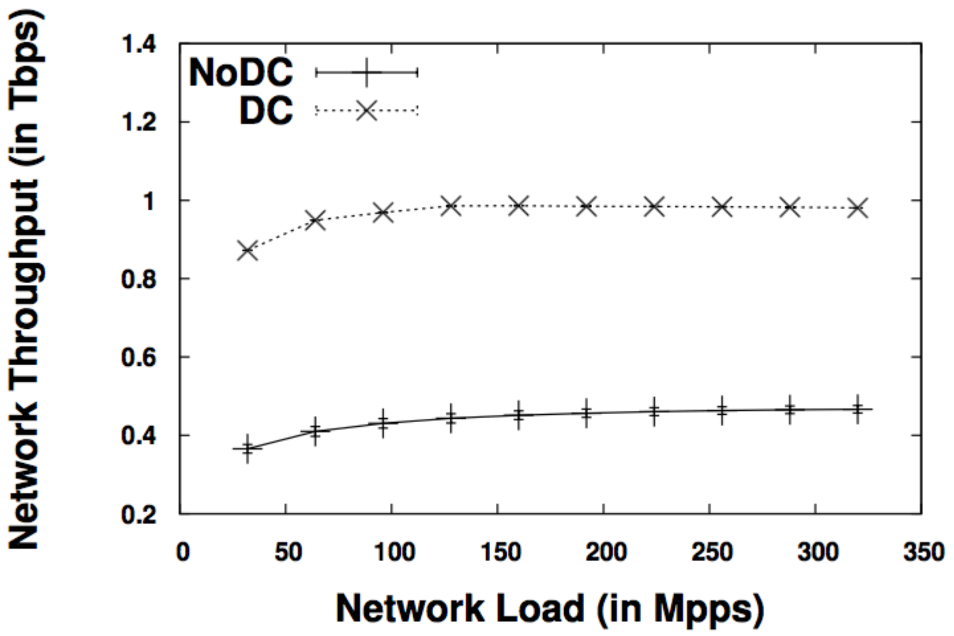
Observations :

1. Variation of delay and throughput with the offered traffic load.
2. Higher accuracy increases throughput.
3. No significant difference for 1 ns & 100 ns accuracy.
4. Throughput difference was less than 0.1 % and their delay difference varied from 56 ns (3.526 and 3.582 μs) to 9 μs (254 and 263 μs). This is not significant considering the delay of a DCN.

UCDAVIS

Performance of DT Synchronization

Varying Load:



(a) Throughput

Observation :

1. Delay compensation is able to accommodate more packets & higher throughput at all load.
2. Delay compensation improves performance of DT scheme.

Performance of DT Synchronization

Varying packet lengths:

(i)**DTC:** packets of constant length of 8192 bytes.

(ii)**DTT:** packets following a tri-modal distribution with modes at 64, 1500 and 8192 bytes.

(iii)**DTU:** uniform packet length distribution, with packet lengths ranging from 64 to 8192 bytes.

UCDAVIS

Performance of DT Synchronization

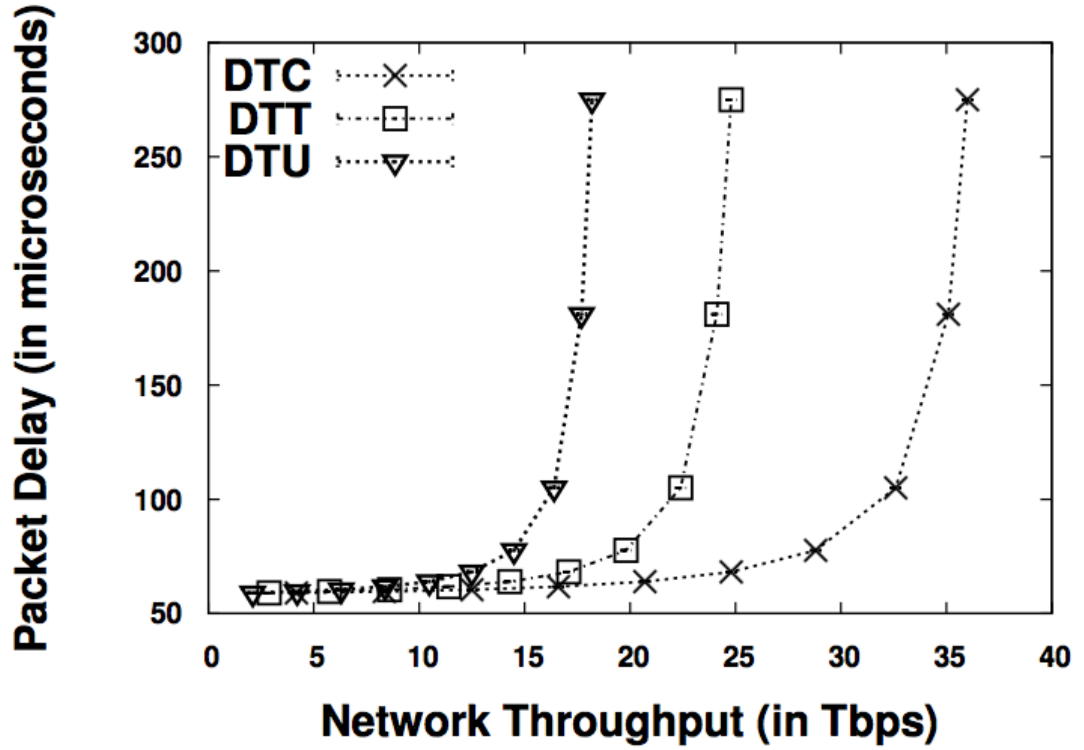


Fig. 8: Throughput and delay for Discrete time with fixed packet length (DTC), with trimodal distribution (DTT) and with uniform distribution (DTU).

Observations :

1. Variation of delay and throughput with the varying packet length.
2. DTC uses constant packet length of 8192 bytes and the throughput saturates at 36 Tbps.
3. DTT uses 5634 bytes of avg. packet length and throughput saturates at 24.76 Tbps.
4. DTU uses 4128 bytes of avg. packet length and throughput saturates at 18.14 Tbps.
5. Packet delays are independent of the packet lengths.

UCDAVIS

Comparison of Time Synchronization Schemes

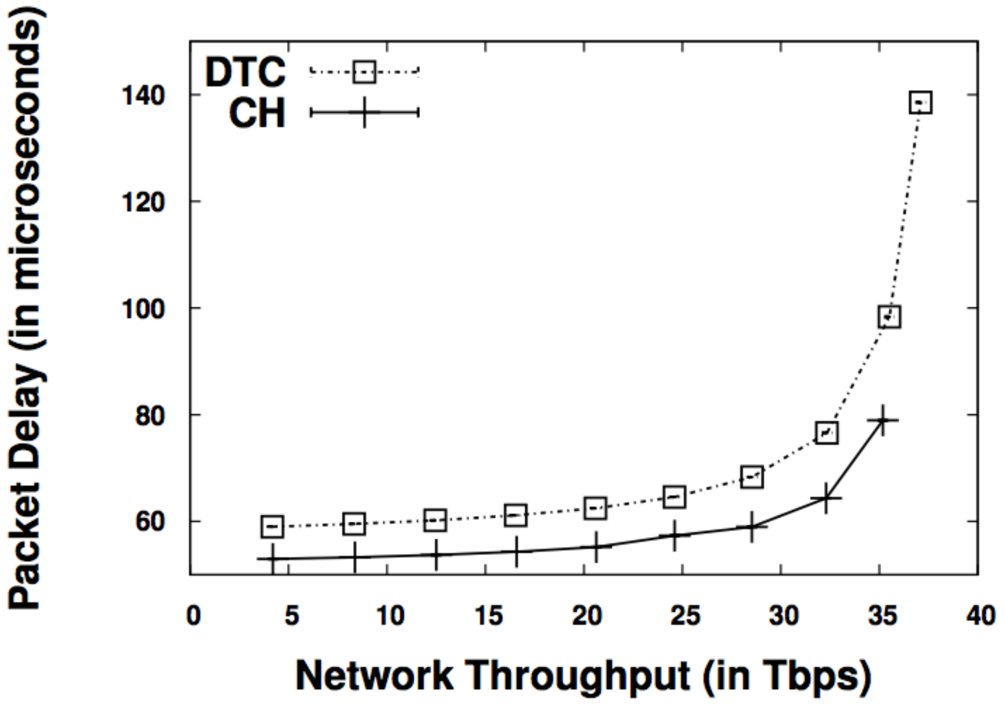


Fig. 9: Throughput and delay for continuous time (CH) and discrete time with fixed packet length (DTC) varying network load.

Observations :

1. Packet delay increases steeply when load is higher than 80%
2. Throughput of DTC achieve 37 Tbps.
3. CH limits at full load as the controller bandwidth(800 Gbps) saturates.
4. Delay obtained by CH is 80 us, DTC is 140 us.
5. CT performs better without any packet length restriction.

Conclusion

- Optically groomed DCNs offer optically transparent end to end path between any source and destination, the network can readily support higher data rates and hence throughput.
- DT synchronization is affected by propagation delay variance and packet length distribution.
- CT synchronization requires reasonably good accuracy (100 ns) to avoid any significant performance impact.

CT synchronization appears to be promising for OGDCN specifically and to all DCNs in general.

UCDAVIS

Thanks!

UCDAVIS