Advancements in Boosting Optical Circuit Switching Capacity in Clos Networks for Data Centers

Eiji Oki
Kyoto University, Japan
IEEE ComSoc Distinguished Lecturer
IEEE Follow

Tutorial talk
IEEE HPSR 2024, July 22-24, 2024, Pisa, Italy

Acknowledgements

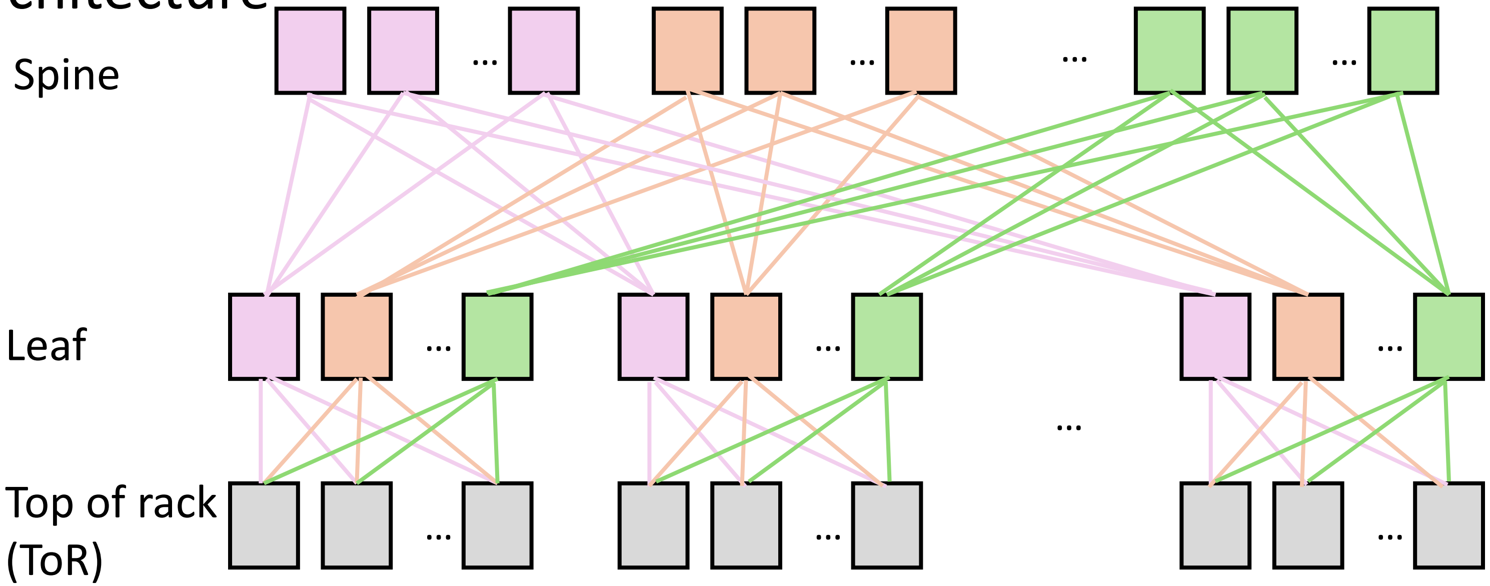
- The presentation slides include research results of my collaborators. I would like to especially thank Haruto Taka and Ryotaro Taniguchi, Kyoto University, and Takeru Inoue and Kazuya Anazawa, NTT, for their contributions.

Outlines

- Background
  - Optical circuit switching in data center
  - Clos network
    - Blocking
    - Strict non-blocking (SNB) condition
- Twisted-folded Clos network (TF-Clos)
  - TF-Clos design with SNB to maximize switching capacity
- Increasing switching capacity
  - Allowing admissible blocking probability
  - TF-Clos design models with guaranteeing admissible blocking probability
- Performance
- Challenges
- Summary

Data center network

- A current data center network, comprising switches and routers, handles extensive data processing
  - Electrical packet switching
- Clos network
  - Multi-stage switching architecture featuring hierarchically combined switches [Clos, BellJ 1953][Jajszczyk, Commag 2003][Kabacinski, Springer 2005]
  - Folded Clos network architecture

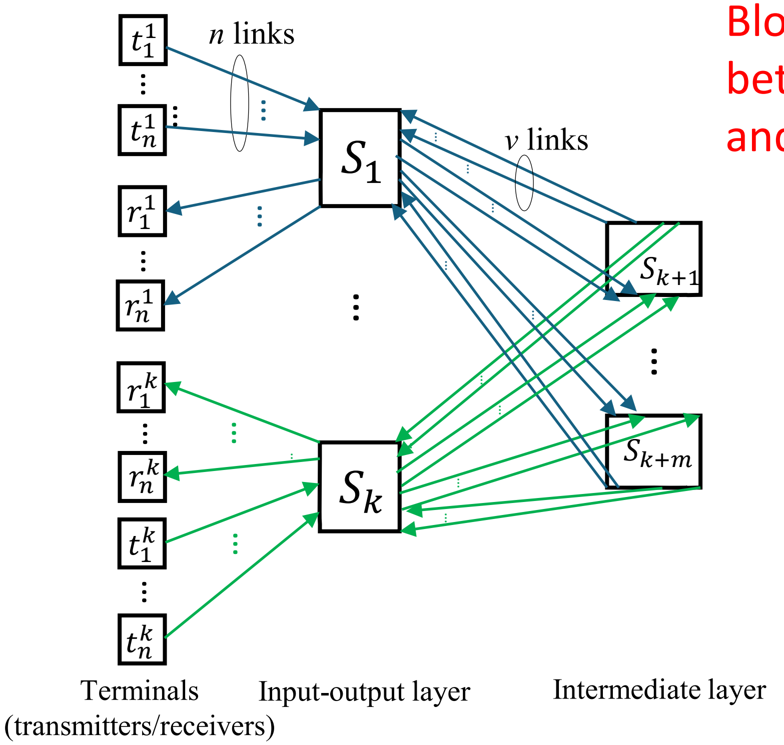


Optical circuit switching in data center

- The slowdown in Moore’s law has worsened the energy efficiency and performance of data center networks
- The future data centers are expected to adopt advanced optical circuit switching technologies for enhanced transmission capacity and power efficiency [Sato, JOCN2024].
- Optical circuit switching technology ensures stable communication quality by:
  - establishing a connection exclusively for data transmission and
  - maintaining it throughout the process.

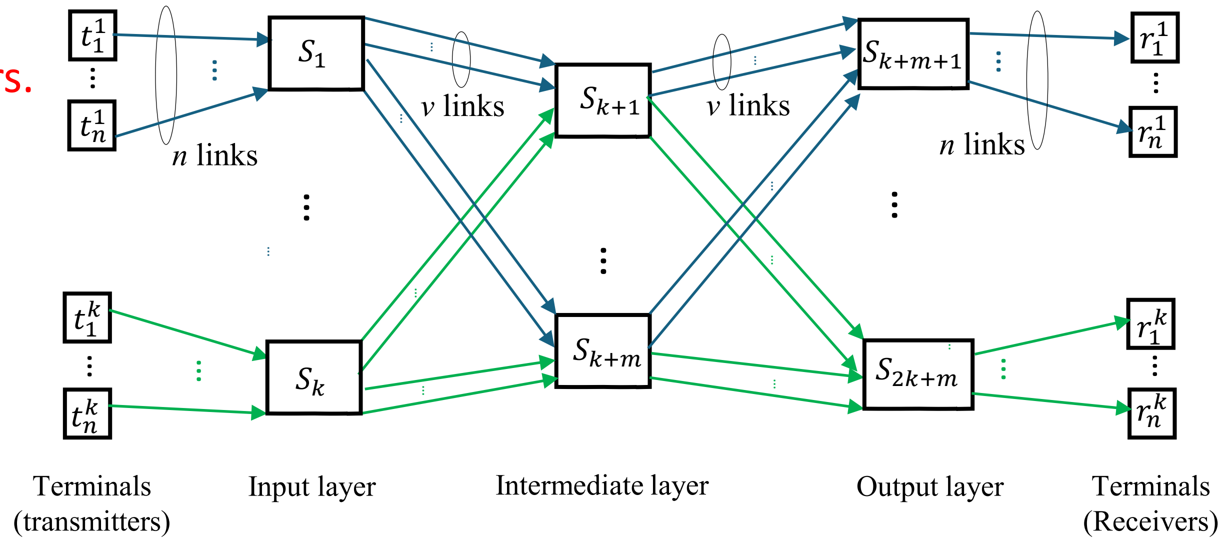
Folded-Clos network (F-Clos) used in datacenter

- F-Clos consists of input-output and intermediate layers.
- A transmitter and receiver pair to be accommodated in a common switch in the input-output layer.
- Datacenters are likely to adopt F-Clos rather than unfolded-Clos (UF-Clos).



Folded-Clos (F-Clos)

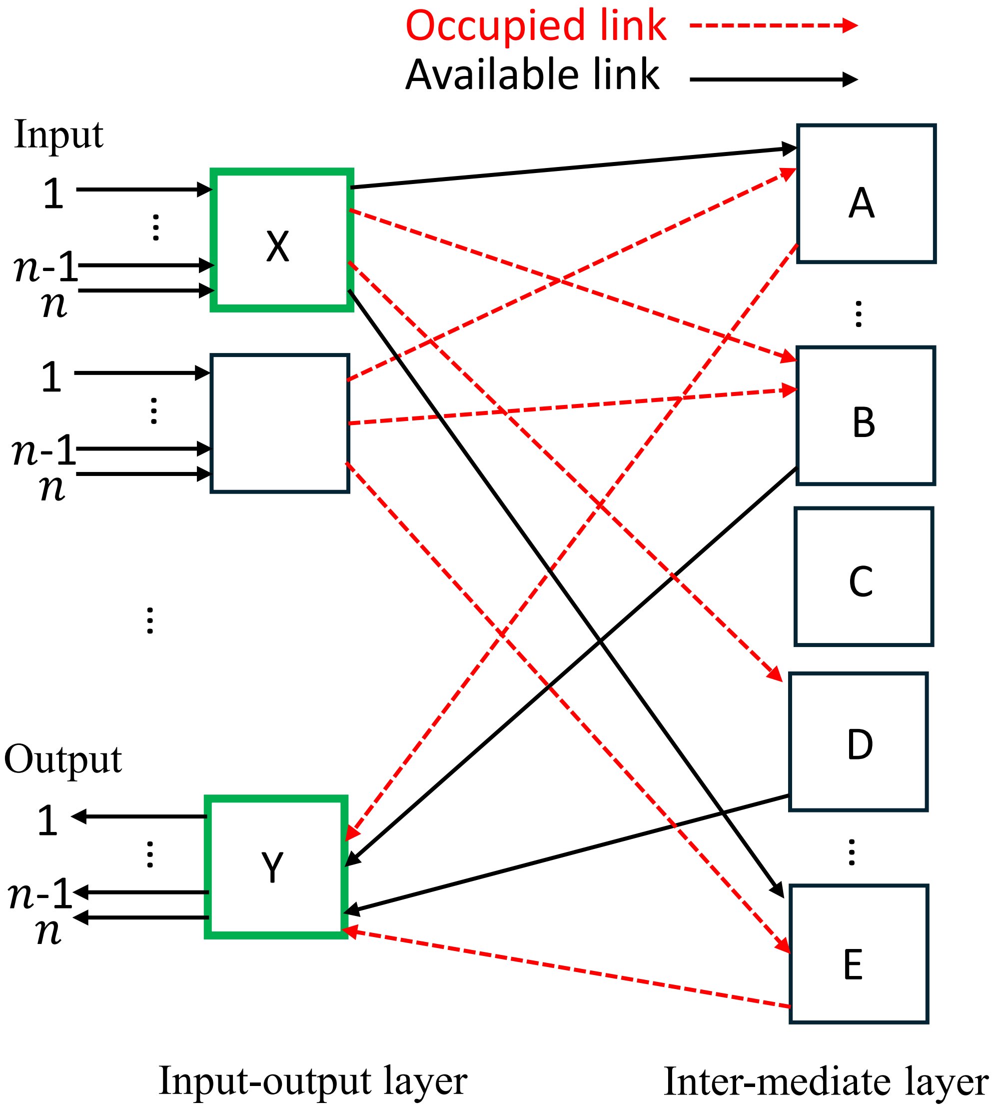
Blocking can occur between input-output and intermediate layers.



Unfolded-Clos (UF-Clos)

Blocking in switching network

- Blocking in switching network
  - A request cannot reach its destination due to switching network usage.
  - Example
    - Consider a request from switch *X* to switch *Y* in the input-output layer.
    - No available route from switch *X* to switch *Y*.



Non-blocking types in switching fabric

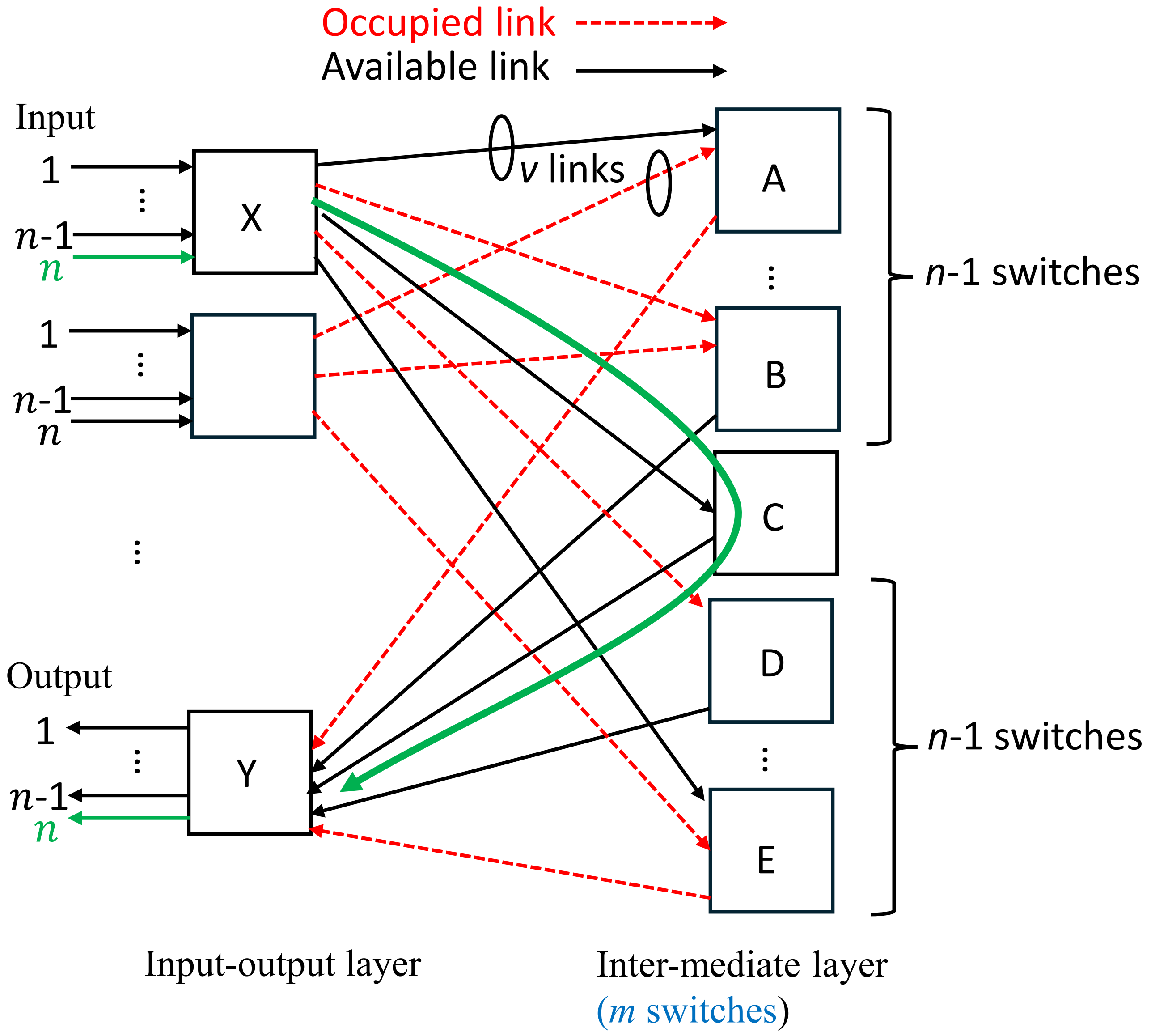
- Strict-sense non-blocking
  - A switching fabric can always connect each idle input port to an arbitrary idle output port independent of its current state and no matter how connecting paths were selected for the existing connections.
- Wise-sense non-blocking
  - A switching fabric can always connect each idle input port to an arbitrary idle output port independent of its current state provided that some given path selection algorithm was used for setting up connections.
- Rearrangeably non-blocking
  - A switching fabric can also always connect each idle input port to an arbitrary idle output port; however, it may be necessary to move existing connections to alternate connecting paths.

Strict non-blocking (SNB) condition

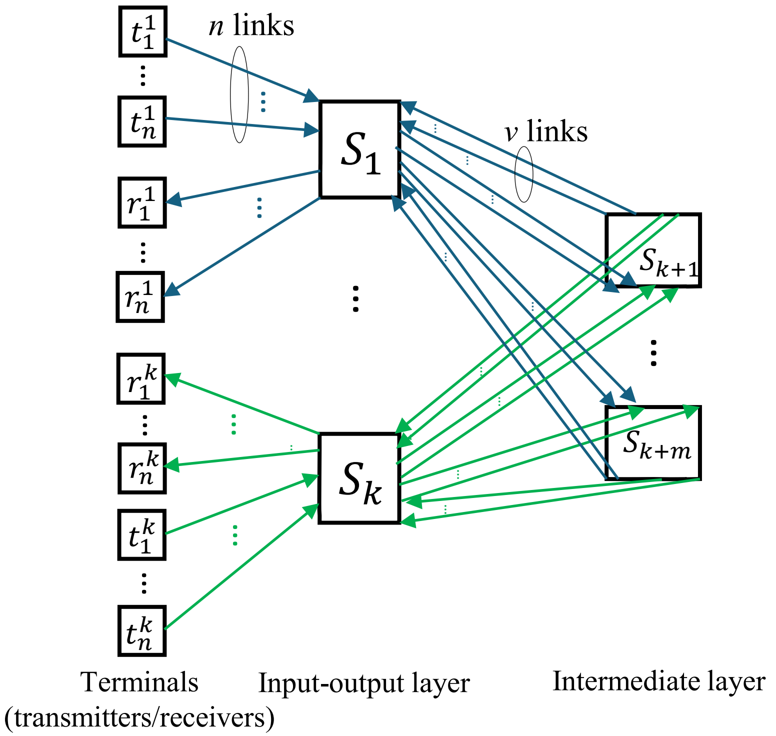
- Strict non-blocking (SNB) condition
  - …
- When $v = 1$
  - $(n-1) + (n-1) + 1 \le m$

  i.e.,
  - $2n - 1 \le m$
- When $v \ge 1$
  - $\left\lfloor \frac{n-1}{v} \right\rfloor + \left\lfloor \frac{n-1}{v} \right\rfloor + 1 \le m$

  i.e.,
  - $2\left\lfloor \frac{n-1}{v} \right\rfloor + 1 \le m$



Design of Folded-Clos (F-Clos) with strict non-blocking (SNB)



Folded-Clos (F-Clos)

$$
\begin{aligned}
\max \quad & nk && \text{(1a)} \\
\text{s.t.} \quad & \boxed{2n \leq N} && \text{(1b)} \\
& \boxed{2vm \leq N} && \text{(1c)} \\
& \boxed{vk \leq N} && \text{(1d)} \\
& \boxed{2\left\lfloor \frac{n-1}{v} \right\rfloor + 1 \leq m} && \text{(1e)} \\
& k + m \leq a && \text{(1f)} \\
& n, k, m, v \in \mathbb{N}, && \text{(1g)}
\end{aligned}
$$

(1a) Maximize switching capacity

(1b)–(1d) Constraint of number of used ports in an $N \times N$ switch

(1e) Strict non-blocking (SNB) condition

(1f) Limitation of number of usable $N \times N$ switches

Given conditions
- Use identical $N \times N$ switches.
- The number of usable $N \times N$ switches, a, is given.

Objective:
Maximize the switching capacity, nk.

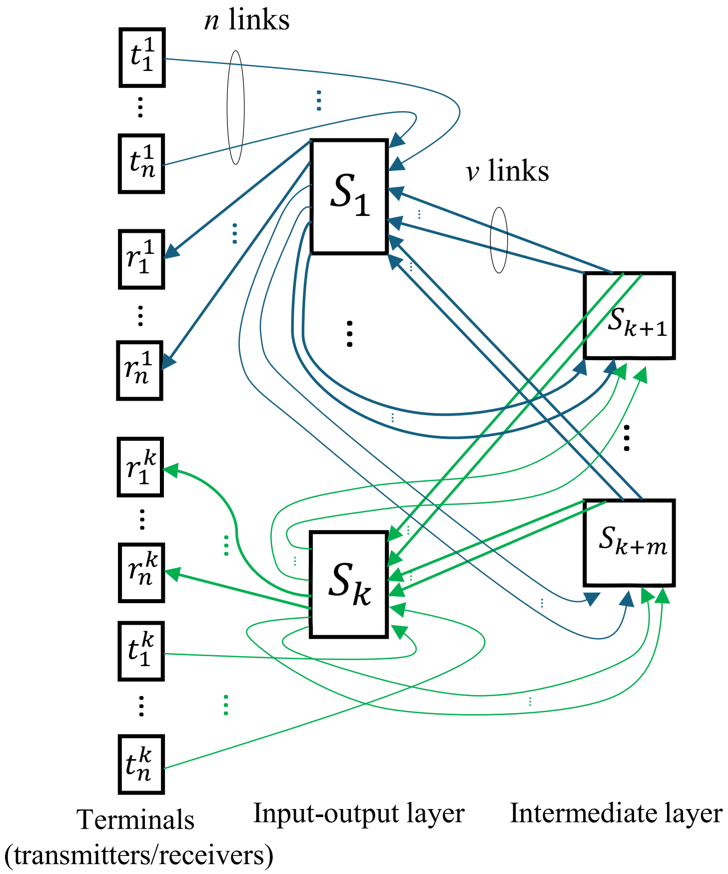
Parameters	Descriptions
N	Number of ports in a switch.
a	Number of usable switches.

Variables	Descriptions
n	Natural. It is the number of ports used in an input-output switch*.
k	Natural. It is the number of input-output switches.
m	Integer. It is the number of intermediate switches.
v	Integer. It is the number of links between an input-output switch and an intermediate switch.

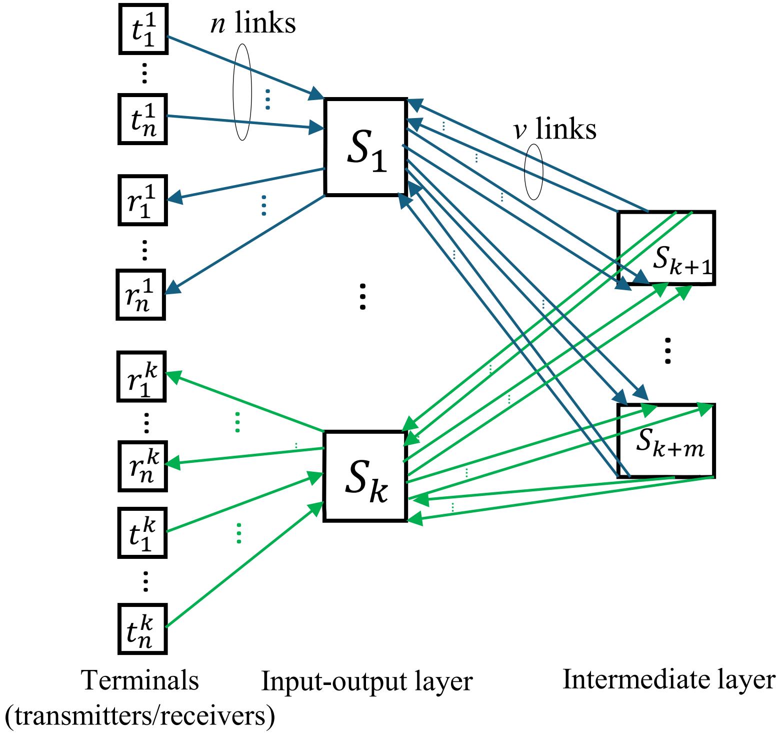
Twisted-folded Clos network (TF-Clos) [T. Mano, GLOBECOM 2019, TNSM 2023]

- F-Clos has a limitation regarding the number of ports in an input-output switch to maximize the switching capacity.
- TF-Clos relaxes the limitation by introducing the idea of *twisting*.

[Mano, TNSM 2023] T. Mano, T. Inoue, K. Mizutani, and O. Akashi, “Redesigning the nonblocking Clos network to increase its capacity,” *IEEE Trans. Netw. Service Manag.*, vol. 20, no. 3, pp. 2558–2574, Sep. 2023.



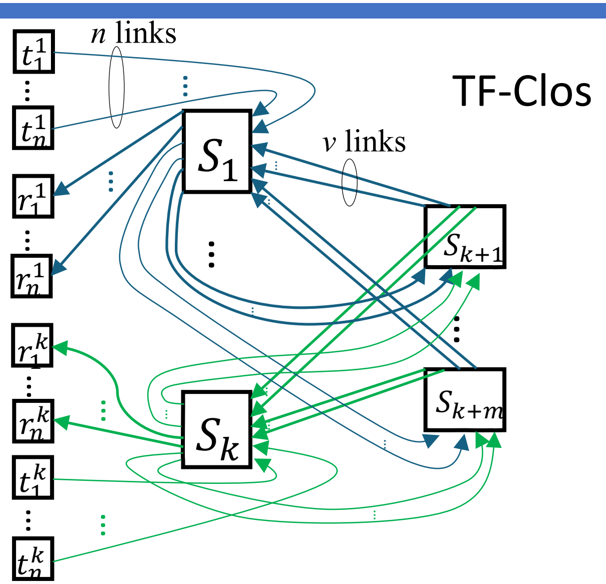
Twisted-folded-Clos (F-Clos)



Folded-Clos (F-Clos)

Checking differences between TF-Clos and F-Clos

- Twisted-folded-Clos (TF-Clos)



$$\begin{aligned}
\max \quad & nk && \text{(2a)} \\
\text{s.t.} \quad & \boxed{\begin{aligned} n + vm &\leq N \\ vk &\leq N \end{aligned}} && \begin{aligned}\text{(2b)} \\ \text{(2c)}\end{aligned} \\
& \boxed{2\left\lfloor \frac{n-1}{v} \right\rfloor + 1 \leq m} && \text{(2d)} \\
& k + m \leq a && \text{(2e)} \\
& n, k, m, v \in \mathbb{N}. && \text{(2f)}
\end{aligned}$$

Maximize switching capacity

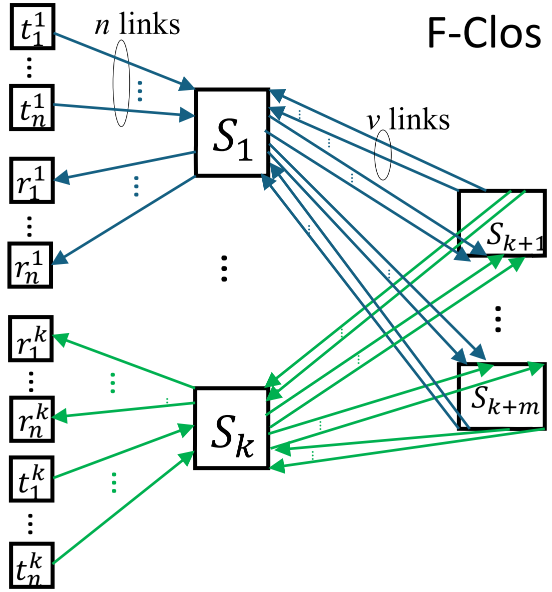
Constraint of number of used ports in an $N \times N$ switch

-> Relaxed by twisting

Strict non-blocking (SNB) condition

Limitation of number of usable $N \times N$ switches

- Folded-Clos (T-Clos)



$$\begin{aligned}
\max \quad & nk && \text{(1a)} \\
\text{s.t.} \quad & \boxed{\begin{aligned} 2n &\leq N \\ 2vm &\leq N \\ vk &\leq N \end{aligned}} && \begin{aligned}\text{(1b)} \\ \text{(1c)} \\ \text{(1d)}\end{aligned} \\
& \boxed{2\left\lfloor \frac{n-1}{v} \right\rfloor + 1 \leq m} && \text{(1e)} \\
& k + m \leq a && \text{(1f)} \\
& n, k, m, v \in \mathbb{N}, && \text{(1g)}
\end{aligned}$$

Maximize switching capacity

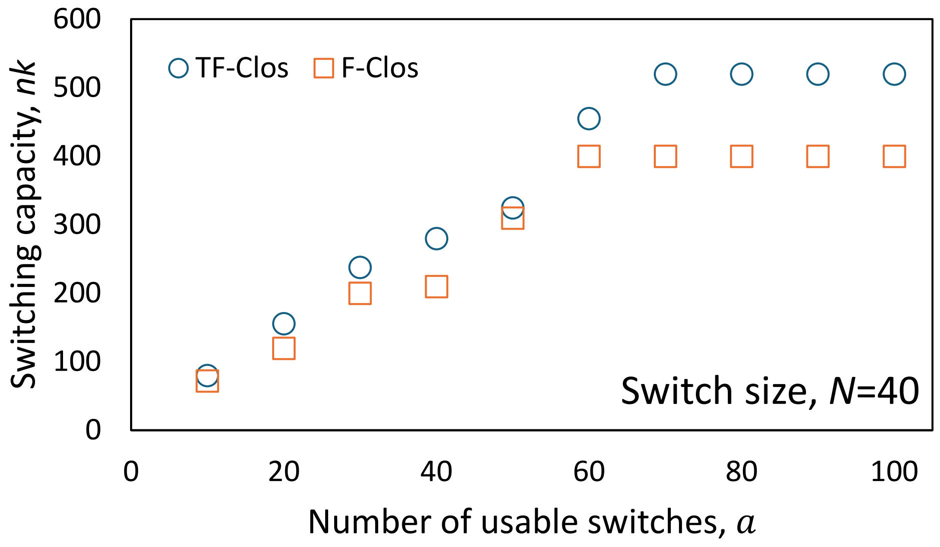
Constraint of number of used ports in an $N \times N$ switch

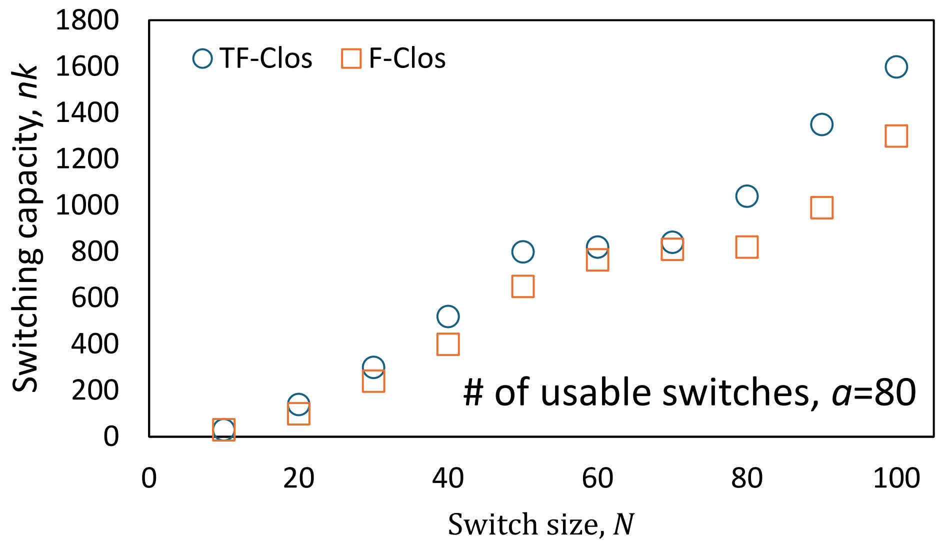
Strict non-blocking (SNB) condition

Limitation of number of usable $N \times N$ switches

Switching capacity of TF-Clos and F-Clos with SNB

- Twisted-folded-Clos (TF-Clos) has a larger switching capacity than folded-Clos (F-Clos).
  - TF-Clos relaxes the port restriction by *twisting*.





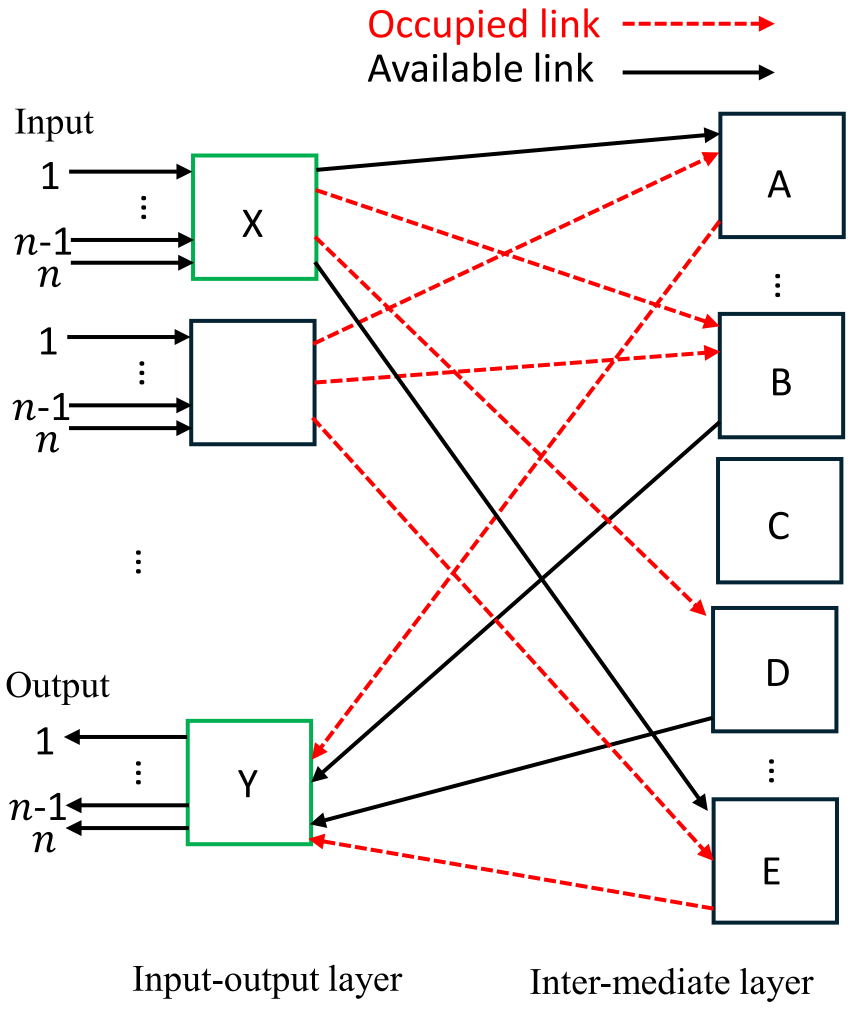
SNB: strict non-blocking

Tolerates blocking to some extent

- While guaranteeing non blocking in TF-Clos has the advantage that any blocking does not occur, the available switching capacity tends to be small.
  - Strict non-blocking (SNB) condition is too strict.
    - SNB condition still limits increasing switching capacity
- A network structure that tolerates blocking to some extent is expected to allow for a more flexible design.
- We want to analyze the blocking probability for a switching network while keeping the network quality.

Difficulty in analyzing blocking probability

- Analyzing exact blocking probability has not been achieved so far.
  - Both links are not available from an input to an output.
    - Two blocking points
  - Occupation behavior of both links are not independent.
- Previous studies approximately analyzed blocking probability [Lee, Bell 1995][Yang, 1998]
  - Introducing the assumption of independent behavior.
  - However, no guarantee the conservativeness.
    - No theoretical upper bound of blocking probability was given.



[Lee, Bell 1995] C. Y. Lee, "Analysis of switching networks," Bell Syst. Tech. J., vol. 34, no. 6, pp. 1287–1315, Nov. 1955.
[Yang, 1998] Y. Yang and N. H. Kessler, "Modelling the blocking behaviour of Clos networks," in Proc. DIMACS Ser. Discrete Math. Theor. Comput. Sci., vol. 42, 1998, pp. 85–102.

Problem statement

- Can we design TF-Clos that guarantees an admissible blocking probability to maximize the switching capacity?
  - An admissible blocking probability
    - is given in advance by a network designer.
    - Is the maximum allowable blocking probability.
  - A blocking probability is defined as:
    - the probability that a connection request generated at a terminal connected to an input port is blocked due to internal blocking in the switching network.

TF Clos design models to guarantee admissible blocking probability

- One-step design model (basic model) [Taka, NL 2023]
  - This model firstly provided the design model that theoretically guarantees admissible blocking probability.
- Two-step design model [Taka, NL 2024]
  - Extended version of the one-step model.
  - This model ensures an admissible blocking probability in a two-step manner
- *S*-step design model [Taka, JOCN 2024] [Oki, ICTON 2024]
  - Generalizing the number of steps

[Taka, NL 2023] H. Taka, T. Inoue, and E. Oki, "Design of twisted and folded Clos network with guaranteeing admissible blocking probability,"
IEEE Netw. Lett., vol. 5, no. 4, pp. 265–269, Dec. 2023.
[Taka, NL 2024] H. Taka, T. Inoue, and E. Oki, "Twisted and folded Clos-network design model with twostep blocking probability guarantee," *IEEE Netw. Lett.*, vol. 6, no. 1, pp. 60–64, Mar. 2024.
[Taka, NL 2024] H. Taka, T. Inoue, and E. Oki, "Design model of twisted and folded Clos network with multi-step grouped intermediate switches guaranteeing admissible blocking probability," *J. Opt. Commun. Netw.*, vol. 16, no. 3, pp. 328–341, Mar. 2024.

Admissible blocking probability (ABP) guarantee

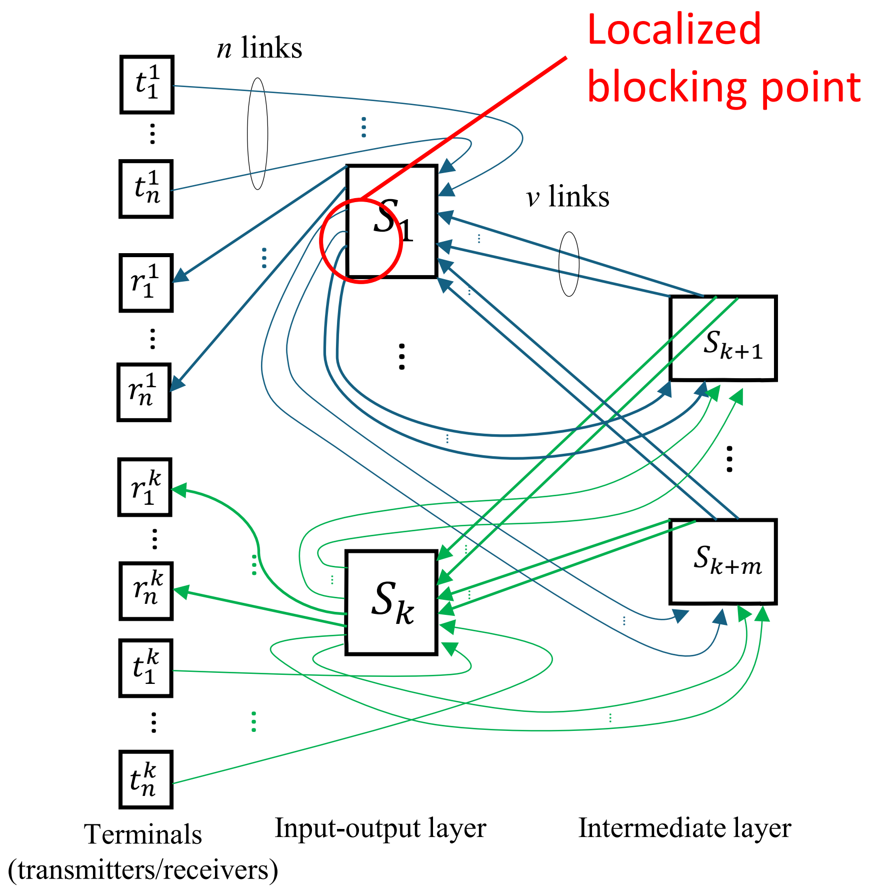
- Given parameters
  - Admissible blocking probability: ϵ
  - Request arrival rate at each port: p
- Strict non-blocking (SNB) condition

$$2\left\lfloor \frac{n-1}{v} \right\rfloor + 1 \leq m$$

- Introducing condition for admissible blocking probability guarantee

$$\sum_{w=n^{\mathrm{snb}}+1}^{n} \binom{n}{w} p^{w}(1-p)^{n-w} \leq \epsilon$$

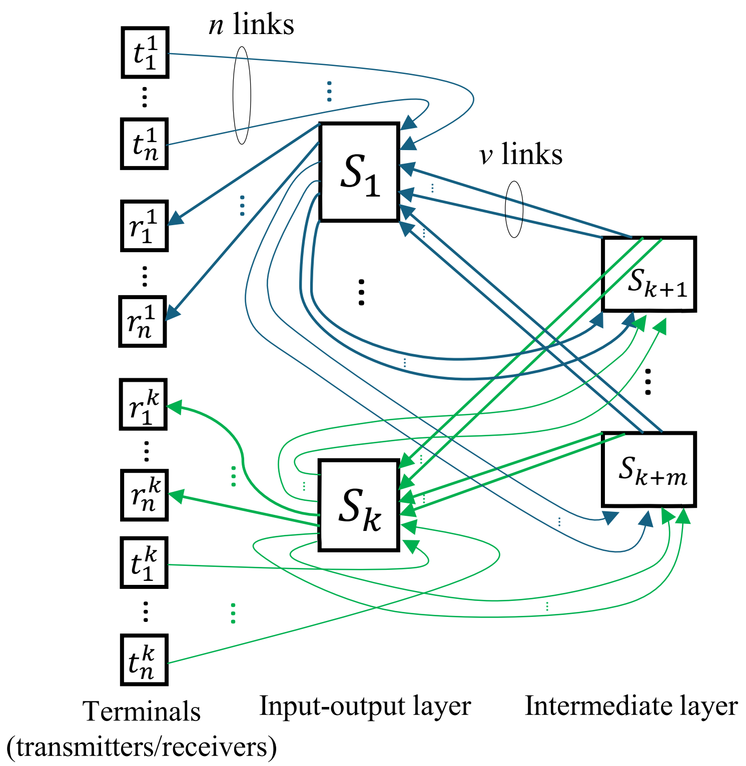
$$2\left\lfloor \frac{n^{\mathrm{snb}}-1}{v} \right\rfloor + 1 \leq m$$



n^{SNB} is a newly introduced decision variable.

n^{SNB} decreases as ϵ increases and p decreases. This relaxes the constraint of TF-Clos design.

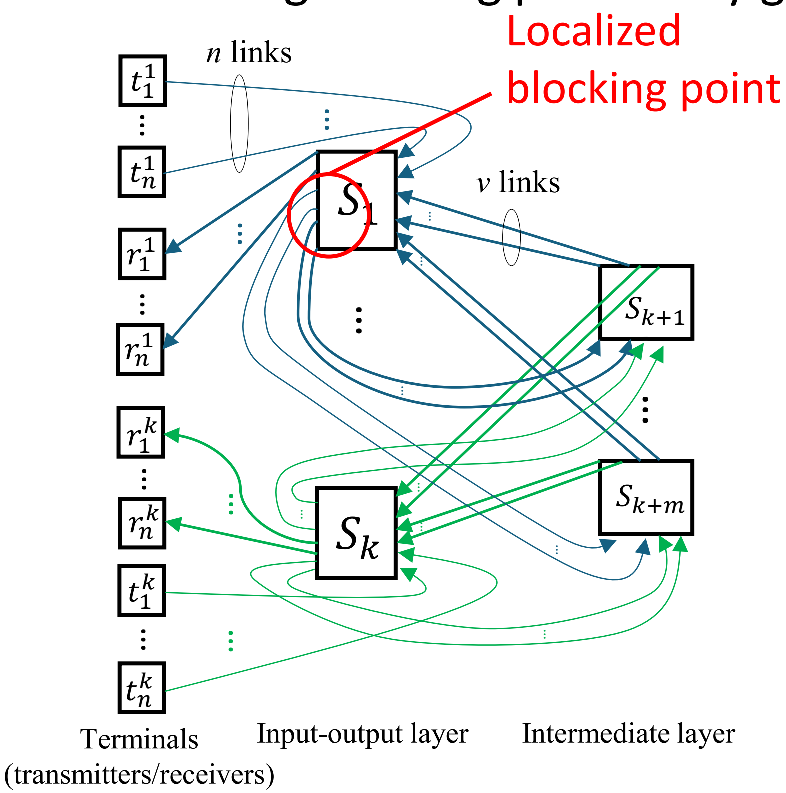
Given parameters and decision variables for TF Clos design to guarantee admissible blocking probability



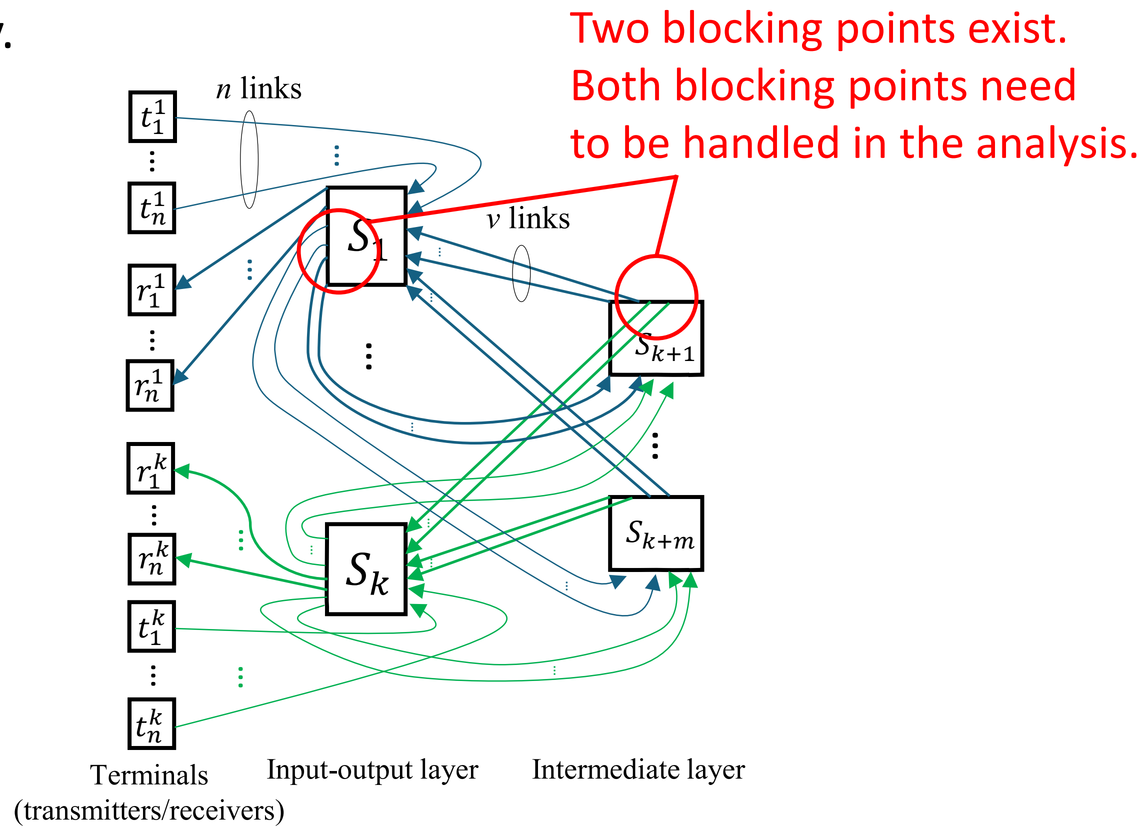
Parameters	Descriptions
N	Number of ports in a switch.
a	Number of usable switches.
p	Probability of request arrival rate.
ϵ	Admissible blocking probability (ABP).
Variables	**Descriptions**
n	Natural. It is the number of ports used in an input-output switch*.
n^{snb}	Integer. It is the number of ports that satisfy the strict-sense non-blocking (SNB) condition.
k	Natural. It is the number of input-output switches.
m	Integer. It is the number of intermediate switches.
v	Integer. It is the number of links between an input-output switch and an intermediate switch.

Comparison between our blocking-localized design model and conventional blocking analytical strategy

- Our design model localizes one blocking points.
  - Achieving blocking probability guarantee strictly.
- The conventional blocking analytical strategy needs to handle two blocking points.
  - Not achieving blocking probability guarantee strictly.



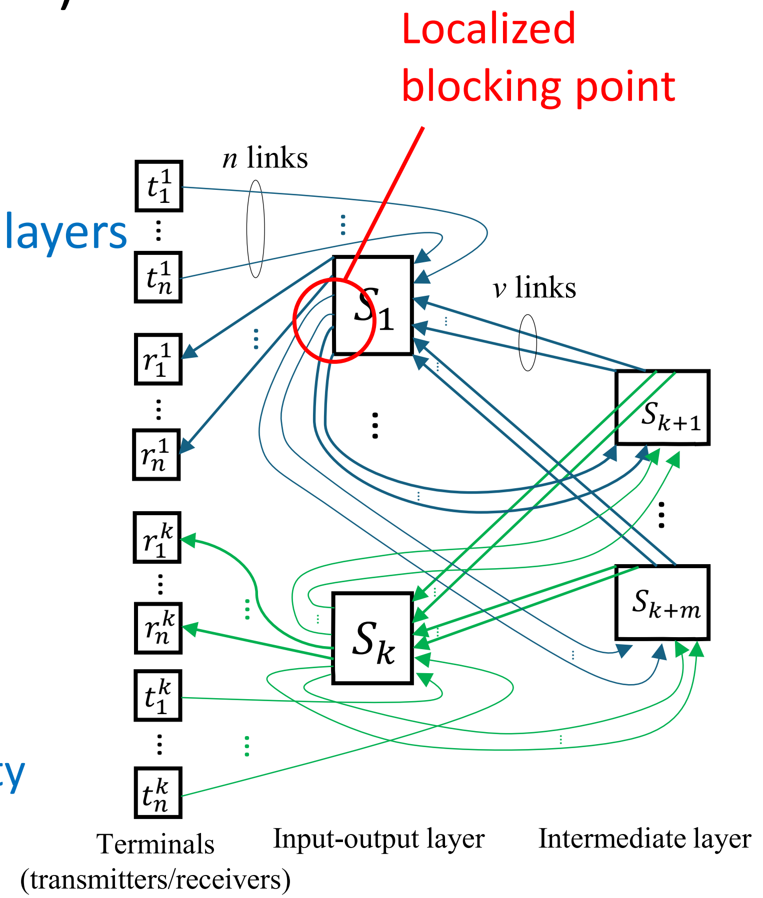
Our blocking-localized design model [Taka, NL 2023]



Conventional blocking analytical strategy

One-step TF-Clos design model guaranteeing admissible blocking probability (ABP)

- The APB guarantee condition replaces the strict non-blocking (SNB) condition.
- If an arriving request violate the APB condition, it is blocked.
  - Blocking point is localized at each input-output switch.
  - SNB condition remains for links between input-output and intermediate layers
  - The admission process is one step.
- This local blocking management guarantees APB.



$$
\begin{aligned}
\max \quad & nk && \text{(5a)} \\
\text{s.t.} \quad & n + vm \le N && \text{(5b)} \\
& vk \le N && \text{(5c)} \\
& \sum_{w=n^{\text{snb}}+1}^{n} \binom{n}{w} p^w (1-p)^{n-w} \le \epsilon && \text{(5d)} \\
& 2\left\lfloor \frac{n^{\text{snb}}-1}{v} \right\rfloor + 1 \le m && \text{(5e)} \\
& k + m \le a && \text{(5f)} \\
& n, n^{\text{snb}}, k, m, v \in \mathbb{N}. && \text{(5g)}
\end{aligned}
$$

Maximize switching capacity (5a)

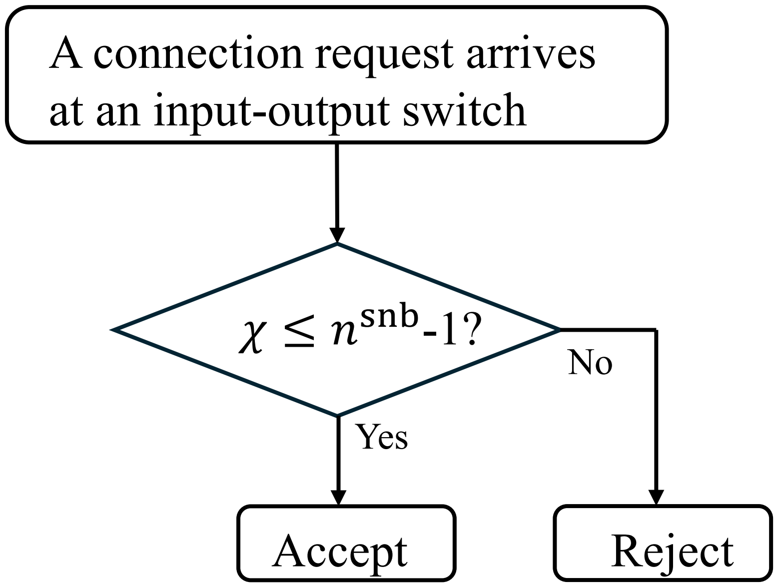
Constraint of number of used ports in an $N \times N$ switch -> Relaxed by twisting (5b)–(5c)

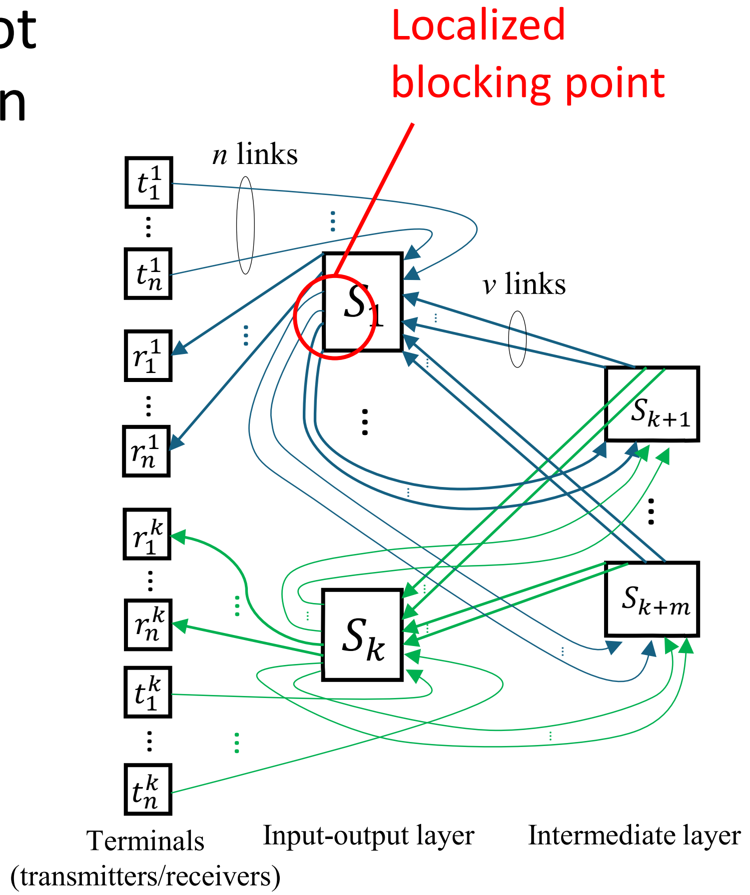
Admissible blocking probability guarantee (5d)–(5e)

Limitation of number of usable $N \times N$ switches (5f)

Request admission in one-step design model

- If the number of occupied links from the input-output switch to all intermediate switches, denoted by χ, does not exceed n^{snb}-1, the request is accepted, and the connection is established, and otherwise rejected.

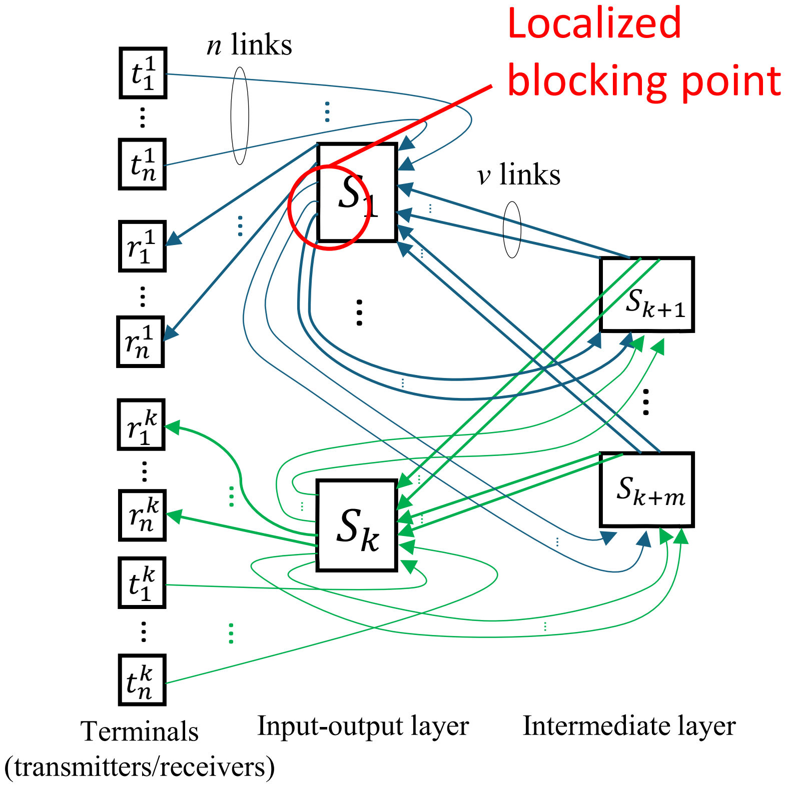




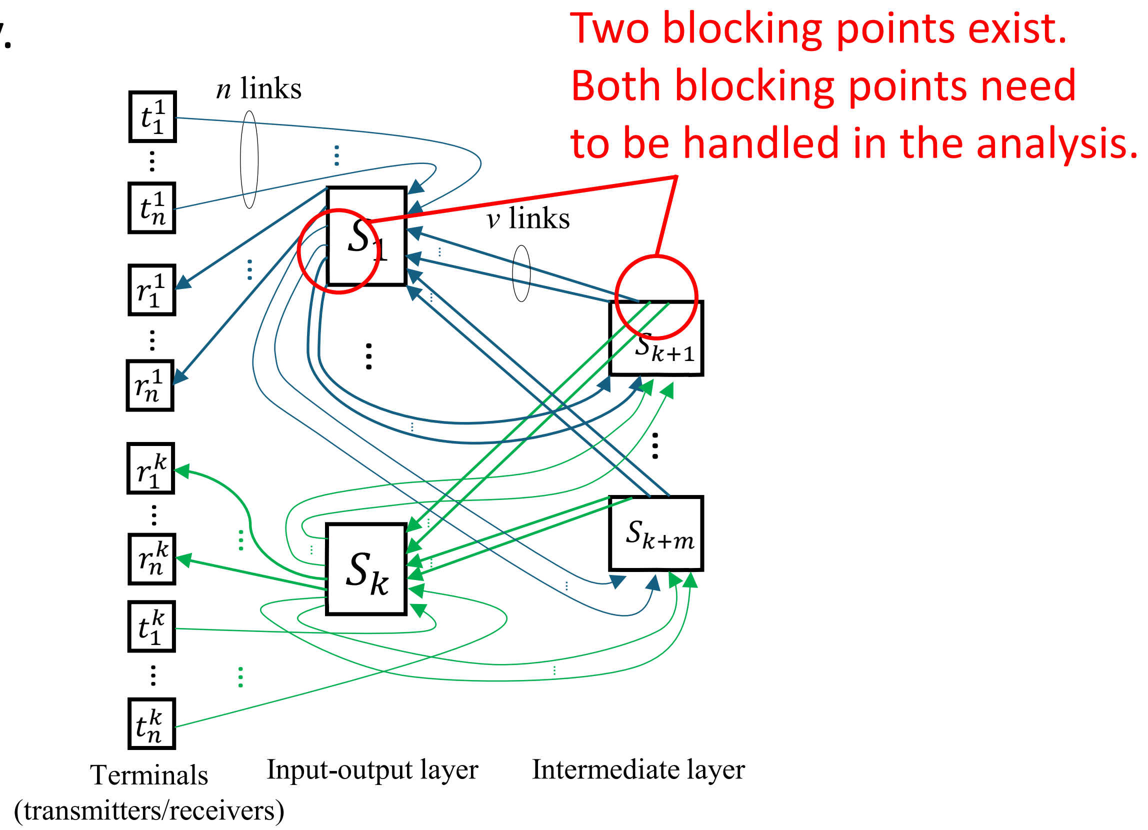
χ: Number of occupied links from the input-output switch to all intermediate switches.

Comparison between our blocking-localized design model and conventional blocking analytical strategy

- Our design model localizes one blocking points.
    - Achieving blocking probability guarantee strictly.
- The conventional blocking analytical strategy needs to handle two blocking points.
    - Not achieving blocking probability guarantee strictly.



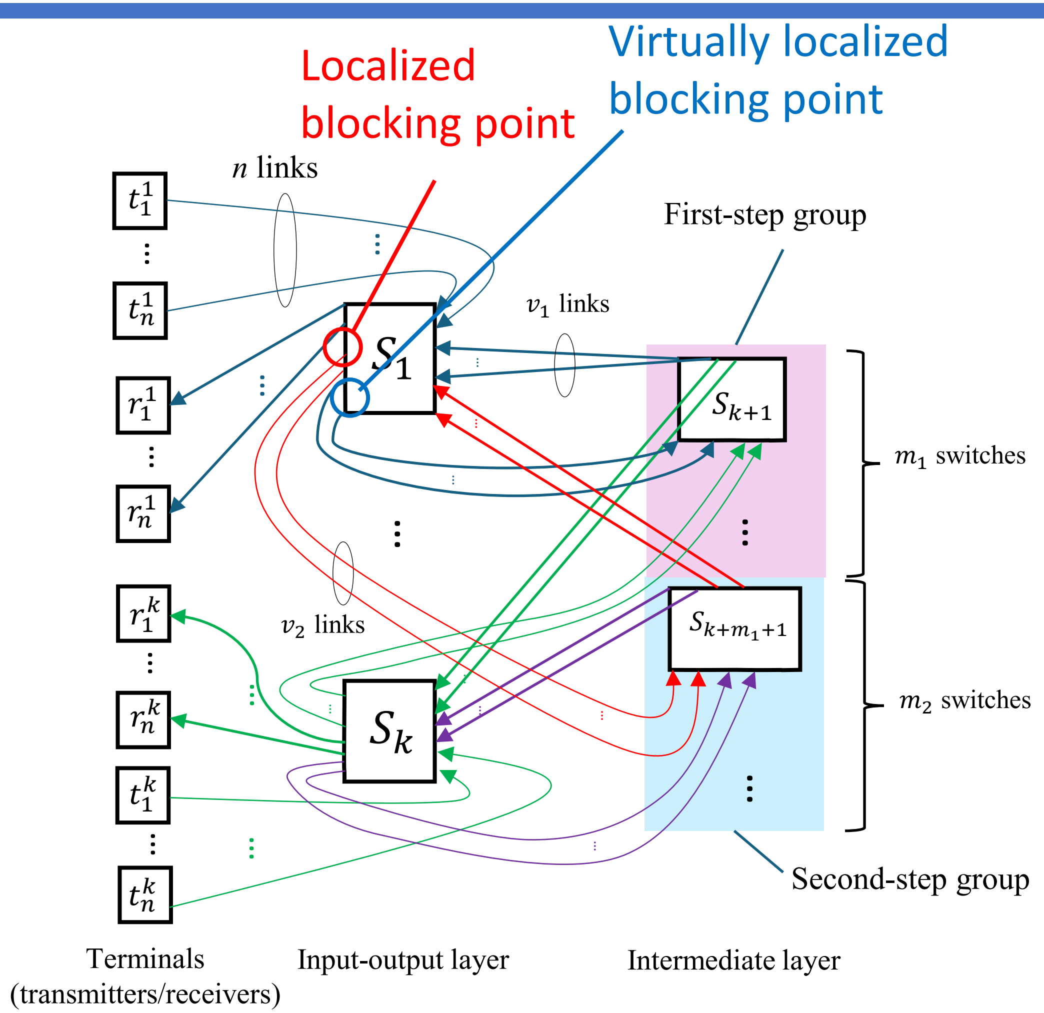
Our blocking-localized design model [Taka, NL 2023]



Conventional blocking analytical strategy

Two-step design model [Taka, NL 2024]

- The admission process is performed by two steps to ensure an admissible blocking probability.
- Categorizes intermediate switches into two groups to alleviate the constraints of the optimization problem.
  - The first group requests are routed first. The second-step group to which requests are routed after they are blocked in the first-step group.
  - In this way, the admissible blocking probability is guaranteed.

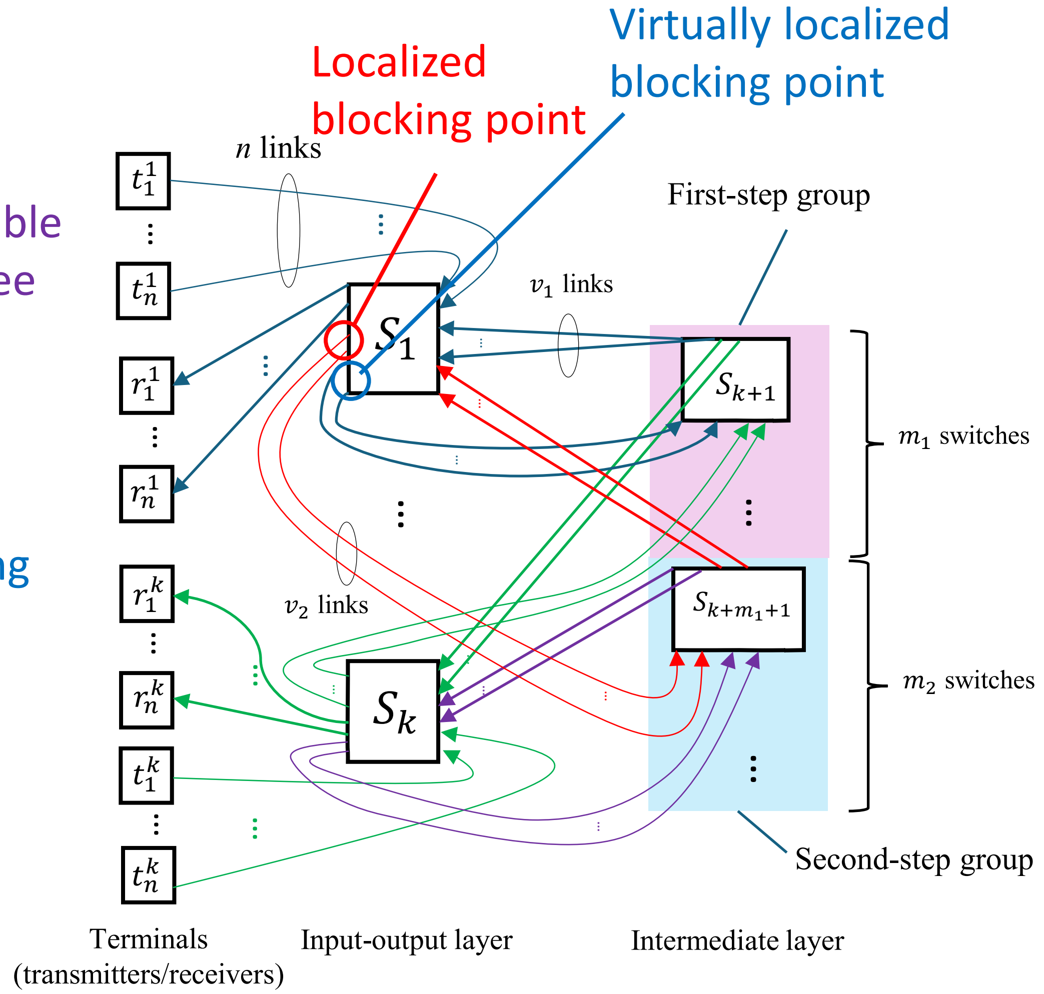


Formulation of two-step design model

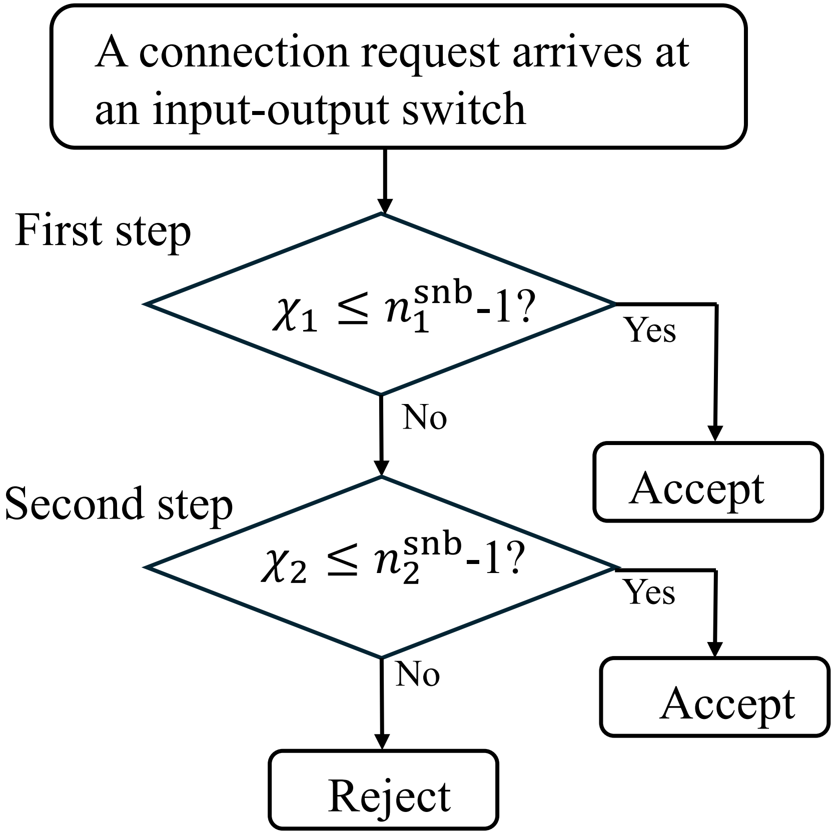
$$
\begin{aligned}
\max \quad & nk && \text{(5a)}\\
\text{s.t.} \quad & n + v_1 m_1 + v_2 m_2 \le N && \text{(5b)}\\
& v_1 k \le N && \text{(5c)}\\
& v_2 k \le N && \text{(5d)}\\
& \sum_{w=n_1^{\mathrm{snb}}+1}^{n} \binom{n}{w} p^w (1-p)^{n-w} \le \begin{cases} \epsilon_1, & \text{if } m_2 > 0 \\ \epsilon, & \text{if } m_2 = 0 \end{cases} && \text{(5e)}\\
& 2\left\lfloor \frac{n_1^{\mathrm{snb}}-1}{v_1} \right\rfloor + 1 \le m_1 && \text{(5f)}\\
& \sum_{w=n_2^{\mathrm{snb}}+1}^{n-n_1^{\mathrm{snb}}} \binom{n-n_1^{\mathrm{snb}}}{w} (\epsilon_1 p)^w \\
& \times (1-\epsilon_1 p)^{n-n_1^{\mathrm{snb}}-w} \le \epsilon, \text{ if } m_2 > 0 && \text{(5g)}\\
& 2\left\lfloor \frac{n_2^{\mathrm{snb}}-1}{v_2} \right\rfloor + 1 \le m_2 && \text{(5h)}\\
& k + m_1 + m_2 \le a && \text{(5i)}\\
& \epsilon \le \epsilon_1 \le 1 && \text{(5j)}\\
& n, n_1^{\mathrm{snb}}, k, v_1, v_2, m_1 \in \mathbb{N} && \text{(5k)}\\
& n_2^{\mathrm{snb}}, m_2 \in \mathbb{N} \cup \{0\} && \text{(5l)}\\
& \epsilon_1 \in \mathbb{R}, && \text{(5m)}
\end{aligned}
$$

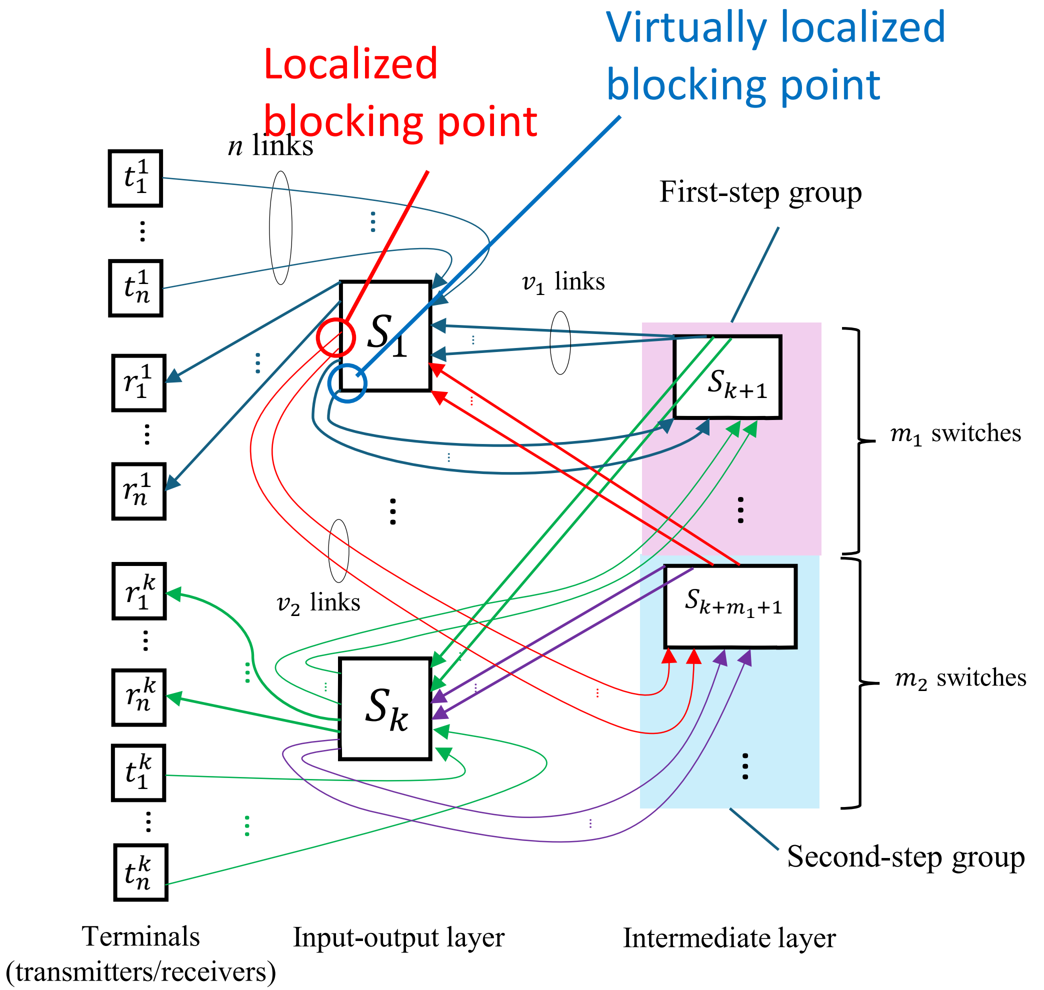
First-step admissible blocking guarantee

Second-step admissible blocking guarantee



Request admission in two-step design model

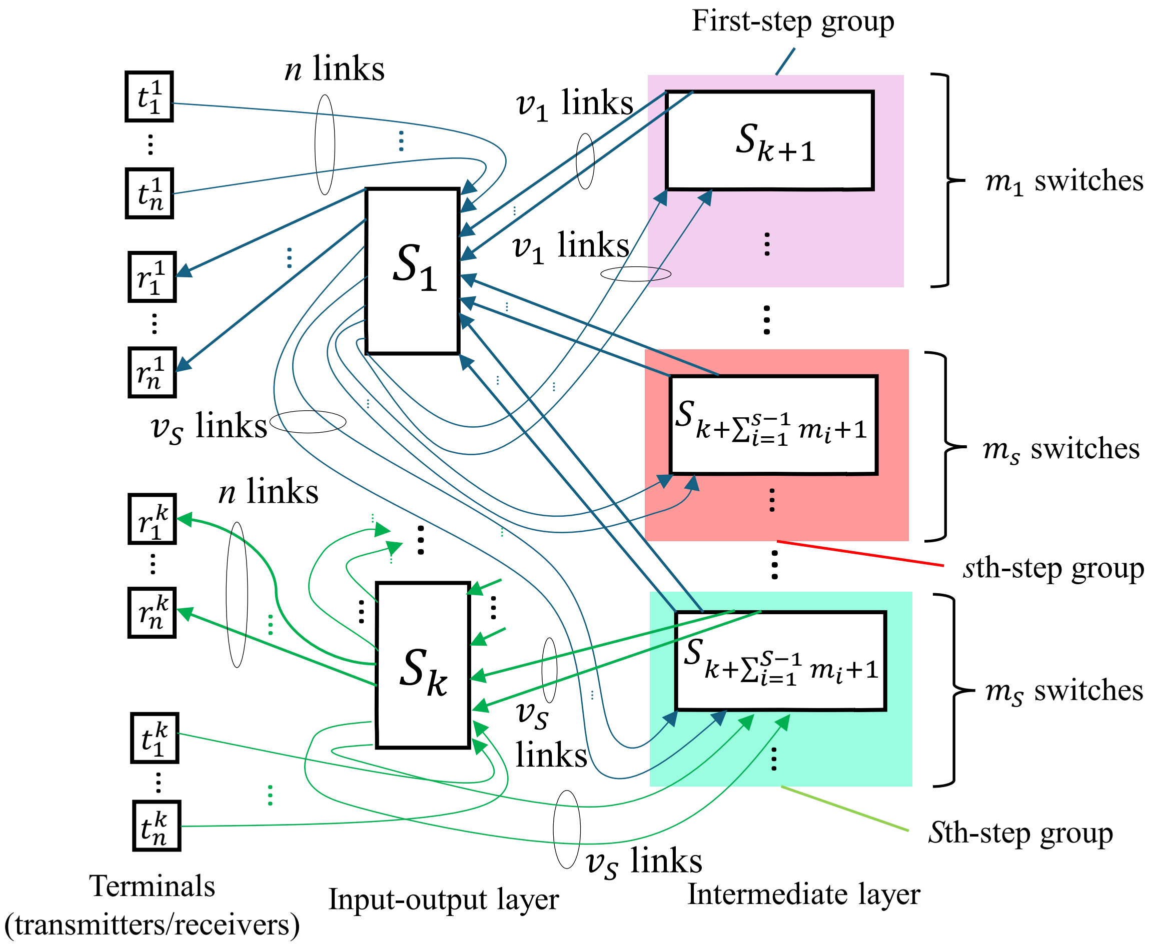




χ_s: Number of occupied links from the input-output switch in the sth step to all intermediate switches.

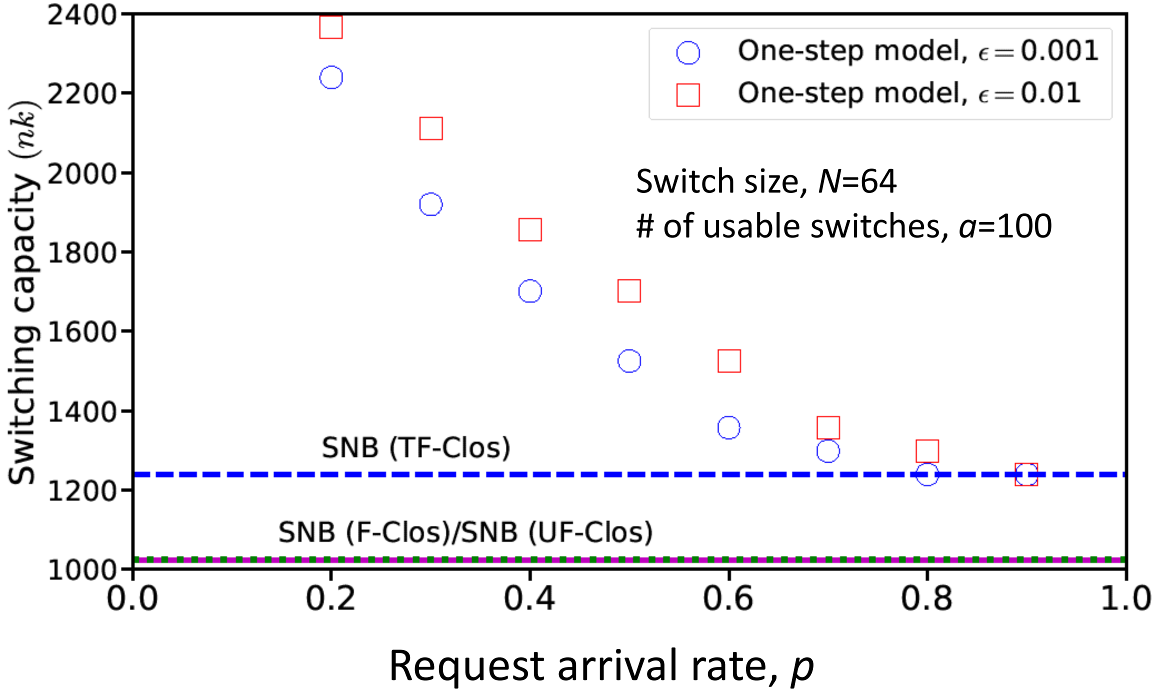
S-step design model [Taka, JOCN 2024] [Oki, ICTON 2024]

- The S-step model expands upon the two-step model, creating a more generalized TF-Clos design model that allows for varying the number of steps for intermediate switches.
- The admission process is performed by S steps to ensure an admissible blocking probability.
- Intermediate switches are partitioned into a maximum of S groups.
- [Oki, ICTON 2024] refines the constraints in the S-step design model.

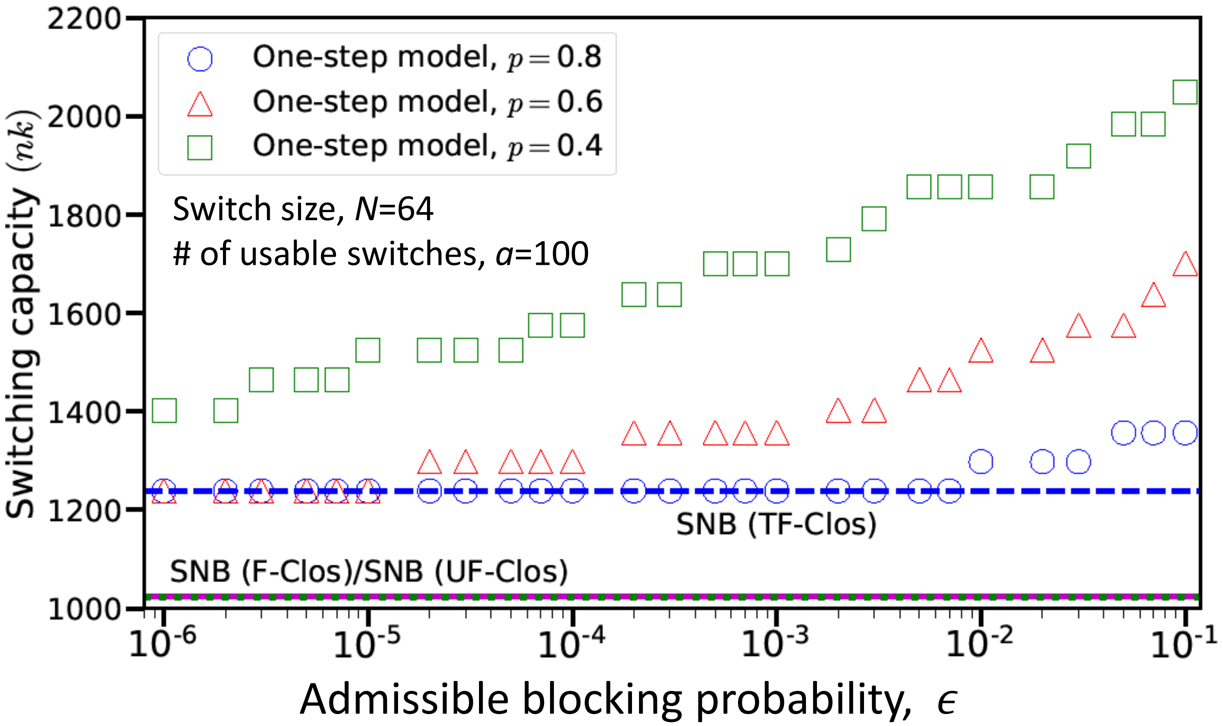


Performance of one-step design model guaranteeing admissible blocking probability (ABP)

- The one-step design model increases the switching capacity as an arrival request rate decreases.
  - One-step model is a special case of *S*-step model with *S*=1.



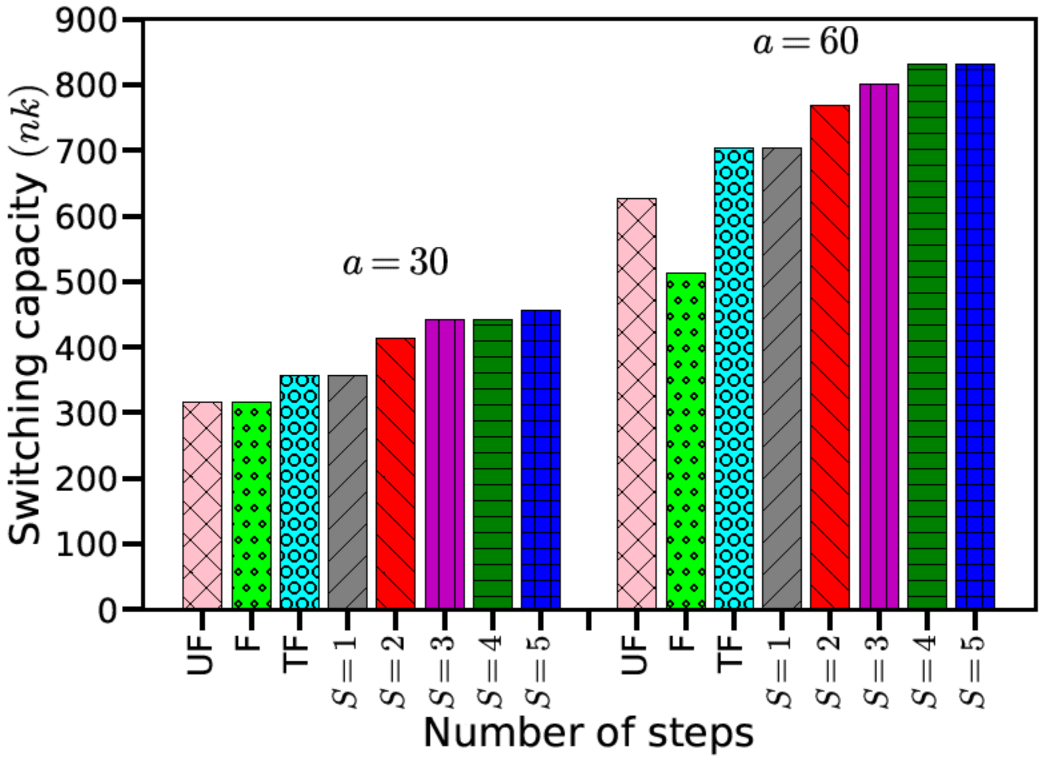
- The one-step design model increases the switching capacity as ϵ increass.

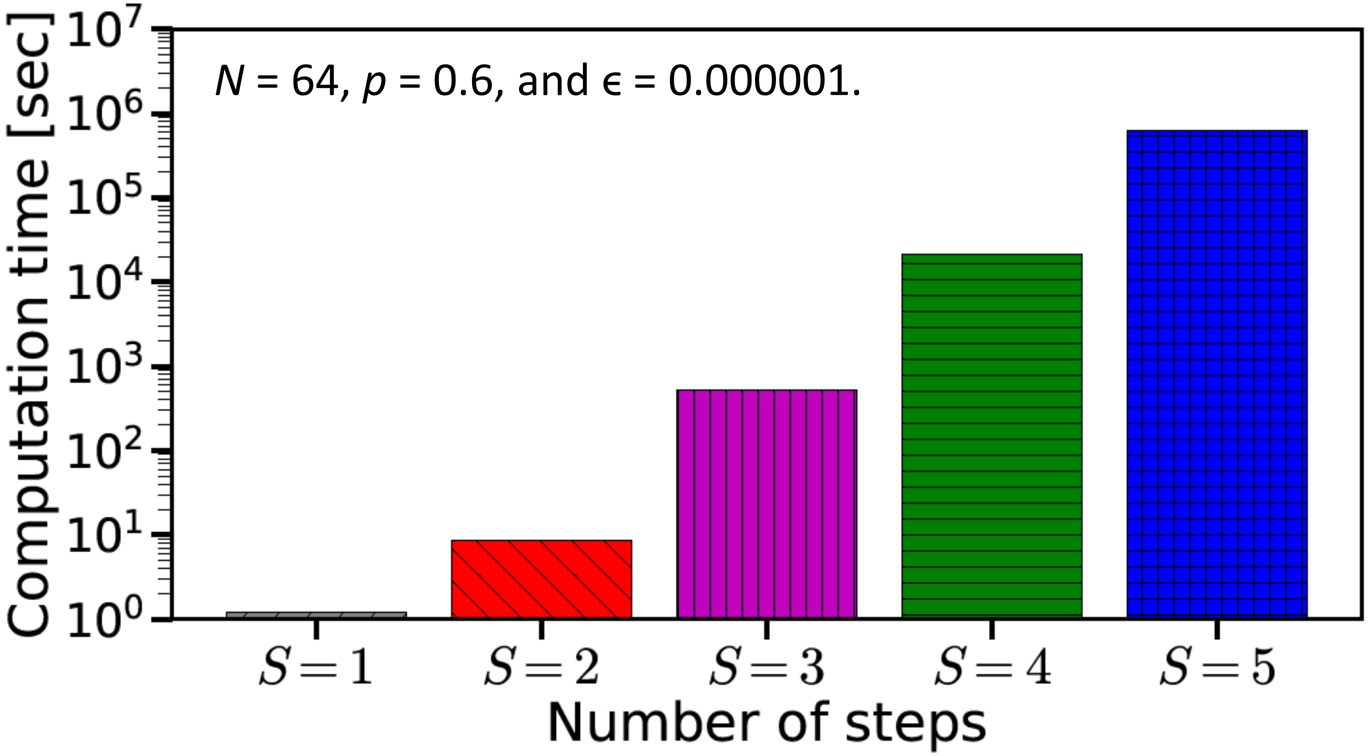


SNB: strict non-blocking
F-Clos: folded Clos, UF-Clos: unfolded Clos

Performance of *S*-step design model guaranteeing admissible blocking probability (ABP)

- The *S*-step design model increases the switching capacity as the number of steps, *S*, increases.
  - The computation time increases in an increase of *S*.





SNB: strict non-blocking

TF: twisted folded Clos, F: folded Clos, UF: unfolded Clos

Effect of refining constraint on the range of $\epsilon_s, s \in [1, S-1]$

- [Oki, ICTON 2024] relaxes the range of ϵ_s used in [Taka, JOCN 2024].
  - [Oki, ICTON 2024] replaces $0 \leq \epsilon_1 \leq \cdots \leq \epsilon_{S-1} \leq \epsilon$ with $0 \leq \epsilon_s \leq 1$
- This constraint refinement increases the switching capacity and reduces the computation time.

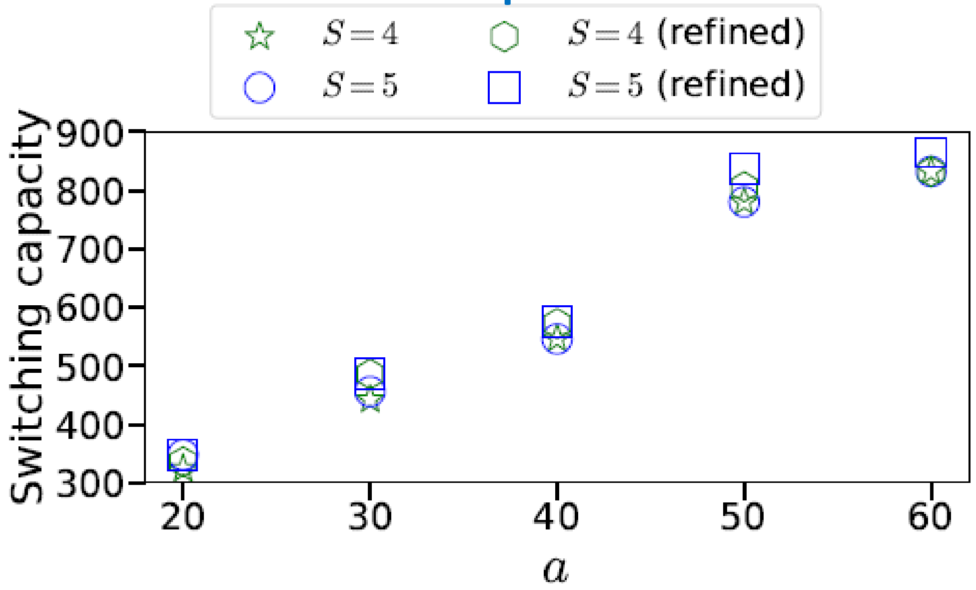


Figure 2: Switching capacity obtained by S-step model [13] and refined model (refined) with $(N, p, \epsilon) = (64, 0.6, 0.000001)$.

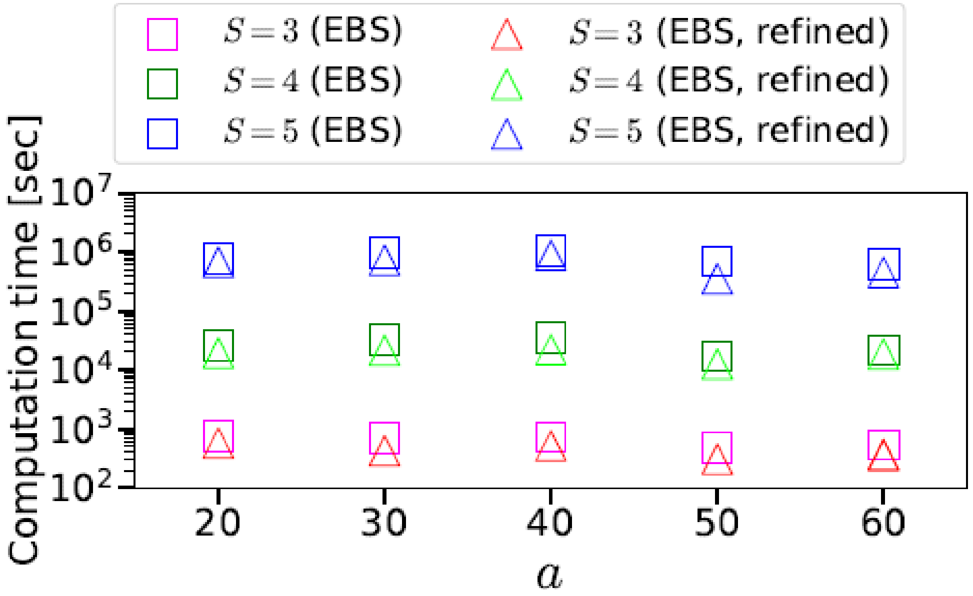


Figure 3: Computation time for S-step model [13] and refined model (refined) with $(N, p, \epsilon) = (64, 0.6, 0.000001)$. The algorithm employed to obtain the solution is EBS.

Challenges

- Reducing computation time
  - A network designer aims to expedite the discovery of a workable solution.
  - As the number of steps or network size expands, locating a viable solution becomes increasingly more work due to longer computation times.
- Selection of suitable number of steps *S*
  - A network designer has difficulty judging with which *S* the switching capacity saturates. On the other hand, it is sure that the computation time increases sharply in an increase of S.
  - A network designer needs to know a suitable value of *S*, considering both increasing behaviors of the switching capacity and the computation time. How to set a suitable value of *S* is left for future work.

Challenges (cont’d)

- Handling different traffic rates generated at different terminals
  - The traffic rate generated at each terminal can differ.
  - We should address how to handle this case in the TF-Clos design model guaranteeing the admissible blocking probability.
- Suppressing the numbers of switches and links
  - Excising works maximizes the switching capacity, provided that the number of switches is given.
  - Another scenario minimizes the required number of switches to achieve a given targeted switching capacity,
- Increasing number of stages
  - The TF-Clos design model can employ this extension idea to increase the number of stages to increase the switching capacity.

Summary

- Background
  - Optical circuit switching in data center
  - Clos network
    - Strict non-blocking (SNB) condition
- Twisted-folded Clos network (TF-Clos)
  - TF-Clos design with SNB to maximize switching capacity
- Increasing switching capacity
  - Allowing admissible blocking probability
  - TF-Clos design models with guaranteeing admissible blocking probability
- Challenges including:
  - Reducing computation time
  - Handling different traffic rates
  - Increasing number of stages