2024 33rd International Conference on Computer Communications and Networks (ICCCN) | 979-8-3503-8461-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCCN61486.2024.10637644

Enhancing Capacity of Optical Circuit Switching Clos Network in Data Center: Progress and Challenges

Eiji Oki[1], Haruto Taka[1], and Takeru Inoue[2]
[1]Kyoto University, Kyoto, Japan
[2]NTT Network Innovation Laboratories, NTT Corporation, Japan

***Abstract*—Future data centers are anticipated to embrace cutting-edge circuit switching technologies, particularly optical switching, renowned for their heightened transmission capacity and energy efficiency. Optical circuit switching guarantees consistent communication quality by establishing dedicated connections for data transmission and maintaining their integrity. Data centers favor Clos-network structures due to their scalability. This paper examines the advancements in designing Clos networks to boost switching capacity while maintaining internal blocking quality. We analyze the characteristics of Clos network design models, providing essentials. We extensively discuss their performances regarding their switching capacities and computation times. Drawing from the review of research progress, we address the challenges in enhancing the switching capacity of Clos networks for future studies.**

***Index Terms*—Clos network, optical circuit switching, data center, switching capacity, blocking probability**

I. Introduction

A data center network, comprising switches and routers, handles extensive data processing. The upcoming data centers are projected to adopt advanced circuit switching technologies, notably optical switching, known for enhanced transmission capacity and power efficiency [1]–[3]. Optical circuit switching technology ensures stable communication quality by establishing a connection exclusively for data transmission and maintaining it throughout the process. Employed in data centers, a Clos network represents a multi-stage switching architecture featuring hierarchically combined switches [4]–[8].

A typical, traditional type of a Clos network is a three-stage unfolded Clos network [4], which this paper refers to UF-Clos. It has input, intermediate, and output layers, each corresponding to each stage. An optical circuit is established from a transmitter connected to an input-layer switch to a receiver connected to an output-layer switch via an intermediate-layer switch. "unfolded" indicates that a transmitter and receiver pair are connected to different switches.

In a data center, it is convenient that a transmitter and receiver pair are connected to a common switch [6]. For this purpose, a data center likely adopts a folded Clos network, which this paper calls F-Clos. Two-stage F-Clos consists of an input-output layer and an intermediate layer. In F-Clos, a transmitter and receiver pair are connected to a common switch in the input-output layer, from which the name *folded* comes.

Blocking is a critical concern in switching networks that utilize circuit-switching technology. In such networks, when a request fails to reach its destination due to network congestion, we refer to it as blocking. Kabaciński [6] introduced a strict-sense non-blocking (SNB) condition for several types of Clos networks. F-Clos has constraints that can suffer from a limitation regarding the number of ports in an input-output switch to increase the switching capacity when we design F-Clos with the SNB condition.

Mano *et al.* [9], [10] invented a twisted-folded Clos network, which we call TF-Clos, to relax the limitation by introducing the idea of *twisting* regarding connecting links between input-output layer switches and intermediate switches. The authors provided a TF-Clos design model with SNB.

While ensuring the strict-sense non-blocking (SNB) condition in the TF-Clos design eliminates blocking, it may still impose constraints on available switching capacity. A network architecture that allows for some level of blocking is anticipated to offer greater flexibility in design. Taka *et al.* [11]–[13] investigated how TF-Clos can be designed to guarantee an admissible blocking probability (ABP), thereby enhancing the switching capacity. ABP refers to 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.

This paper reviews the research progress on designing Clos networks to enhance the switching capacity while ensuring the internal blocking quality. We analyze the features of Clos-network design models by explaining their fundamentals. We extensively discuss their performances regarding their switching capacities and computation times. Based on the research progress review, we discuss the research challenges to enhance the switching capacity of a Clos network for future studies.

The rest of the paper is organized as follows. Section II describes Clos-network design models with SNB conditions. Section III presents TF-Clos design models ABP. Section IV discusses the performances of different design models. Section V describes the research challenges toward enhancing the switching capacity of a Clos network. Section VI concludes the paper.

II. Clos-network design models with strict-sense non-blocking (SNB) condition

This section considers designing a Clos network that satisfies the SNB condition to maximize the switching capacity using identical $N \times N$ switches whose number is a. N denotes the number of ports on each side of a switch. The switching capacity defines the number of terminals with transmitters/receivers accommodated in the network, where the number of transmitters is equal to that of receivers. Under the SNB condition, any available transmitter must be connected to any available receiver through the network. Clos-network design models with the SNB condition introduced in Section II do not impose any connection routing policy from a transmitter to a receiver. Tables I lists symbols frequently used throughout this paper.

TABLE I
List of symbols frequently used throughout this paper.

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.

This section presents three types of Clos-network design models with the SNB condition: F-Clos and TF-CLos. Mano *et al.* [9], [10] introduced these models.

A. Folded Clos-network (F-Clos) design model

F-Clos consists of input-output and intermediate layers, as shown in Fig 1. F-Clos allows a transmitter and receiver pair to be accommodated in a common switch in the input-output layer. This is an advantage of F-Clos, compared to UF-Clos. The input layer consists of k switches, $S_i \in [1, k]$. We call a switch in the input-output layer an input-output switch. The intermediate layer consists of m switches, $S_i \in [k+1, k+m]$. We call a switch in the intermediate layer an intermediate switch.

$k+m$ switches construct F-Clos as follows. An input-output switch, $S_i \in [1, k]$, accommodates $2n$ links incoming from n transmitters, $t_j^i, i \in [1, k], j \in [1, n]$, and outgoing to n receivers, $r_j^i, i \in [1, k], j \in [1, n]$, on one side and $2vm$ links incoming from and outgoing to m intermediate switches, $S_i \in [k+1, k+m]$, on the other side; v links between an input-output switch and an intermediate switch exist. An input-output switch has $n \times vm$ and $vm \times n$ switching functions. An intermediate switch has a $vk \times vk$ switching function. A circuit connection is established from an originated terminal with a transmitter to a destined terminal with a receiver via an input-output switch, an intermediate switch, and an input-output switch.

We describe the F-Clos design model. The F-Clos design problem to maximize the switching capacity while satisfying the SNB condition can be defined by:

$$\max \quad nk \tag{1a}$$
$$\text{s.t.} \quad 2n \leq N \tag{1b}$$
$$2vm \leq N \tag{1c}$$
$$vk \leq N \tag{1d}$$
$$2\lfloor \tfrac{n-1}{v} \rfloor + 1 \leq m \tag{1e}$$
$$k + m \leq a \tag{1f}$$
$$n, k, m, v \in \mathbb{N}, \tag{1g}$$

where $\mathbb{N}$ denotes a set of positive integers. N and a are given parameters and n, k, m, and v are positive integer decision variables. The objective function in (1a) maximizes the switching capacity, nk, which is the number of terminals (transmitters/receivers) accommodated in the network. Equations (1b) and (1c) indicate that both $2n$ and $2vm$ do not exceed N due to the limitation of the number of ports in an input-output switch, respectively. Equation (1d) expresses that vk must not exceed N due to the limitation of the number of ports in an intermediate switch. The SNB condition is given by (1e). Equation (1f) expresses that the used number of switches, $k+m$, does not exceed the number of available switches, a. Equation (1g) gives the range of decision variables.

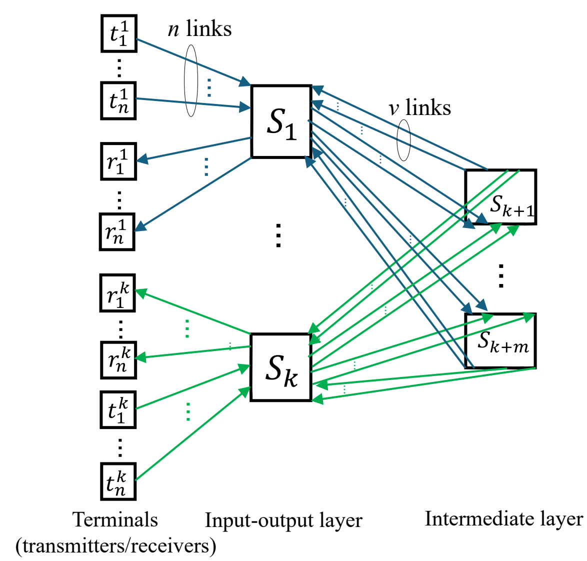


Fig. 1. F-Clos structure with switching capacity nk.

B. Twisted-folded Clos-network (TF-Clos) design model

Studies [9], [10] invented TF-Clos to increase the switching capacity compared with F-Clos while succeeding the advantage of F-Clos, which allows a transmitter and receiver pair to be accommodated in a common input-output switch. TF-Clos consists of input-output and intermediate layers, as shown in Fig 1. In F-Clos, the constraints of (1b) and (1b) can be 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*.

$k+m$ switches construct TF-Clos as follows. An input-output switch, $S_i, i \in [1, m]$, accommodates $n + mv$ links incoming from n transmitters, $t^i_j, i \in [1, k], j \in [1, n]$, and incoming from m intermediate switches, $S_i, i \in [k+1, k+m]$, on one side and $n + mv$ links outgoing to n receivers, $r^i_j, i \in [1, k], j \in [1, n]$, and outgoing to m intermediate switches, $S_i, i \in [k+1, k+m]$, on the other side; v links between an input-output switch and an intermediate switch exist. In the same way as F-Clos, an input-output switch has $n \times vm$ and $vm \times n$ switching functions. An intermediate switch has a $vk \times vk$ switching function. A circuit connection is established from an originated terminal with a transmitter to a destined terminal with a receiver via an input-output switch, an intermediate switch, and an input-output switch.

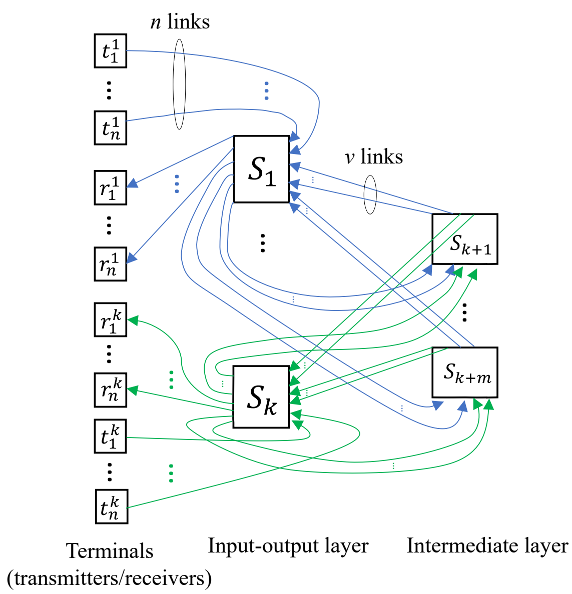


Fig. 2. TF-Clos structure with switching capacity nk.

We describe the TF-Clos design model. The TF-Clos design problem to maximize the switching capacity while satisfying the SNB condition can be defined by:

$$\max \quad nk \tag{2a}$$
$$\text{s.t.} \quad n + vm \leq N \tag{2b}$$
$$vk \leq N \tag{2c}$$
$$2\lfloor \tfrac{n-1}{v} \rfloor + 1 \leq m \tag{2d}$$
$$k + m \leq a \tag{2e}$$
$$n, k, m, v \in \mathbb{N}. \tag{2f}$$

Equation (2b) replaces (1b) and (1b) used in F-Clos, which is the only difference between (1a)–(1g) in F-clos and (2a)–(2f) in TF-clos.

III. TF-Clos design models guaranteeing admissible blocking probability (ABP)

Whereas TF-Clos ensures the SNB condition [9], [10] as detailed in Section II-B, preventing any blocking, it typically results in limited switching capacity. A network structure that allows for some degree of blocking is anticipated to offer more flexibility in design.

This section presents TF-Clos design models introduced in [11]–[13], each of which guarantees ABP to increase the switching capacity, compared to TF-Clos with SNB. ABP refers to the probability that a connection request from a terminal connected to an input port of an input-output switch is blocked due to internal blocking within the switching network. Section III-A describes a basic design model of TF-Clos with guaranteeing ABP, which handles the request blocking occurrence; a request admission is judged by checking the utilization of links from the input-output layer to the intermediate layer in one step. We called the basic model of TF-Clos with guaranteeing ABP a one-step design model. Section III-B describes an extended version of the one-step design model, called an S-step model, which handles the request admission procedure in a multiple-step manner.

A. One-step design model

The TF-Clos structure of the one-step design model is the same as that with SNB, as shown in Fig. 2, but the difference between both models is how to use n links from an input-output switch to all the intermediate switches. TF-Clos with SNB, which satisfies $2\lfloor \frac{n-1}{v} \rfloor + 1 \leq m$, allows us to use any unoccupied link between both layers while avoiding any blocking. On the other hand, the one-step design model restricts the maximum number of allowable links to use from an input-output switch to all the intermediate switches, which refers to $n^{\text{snb}} (\leq n)$ to guarantee ABP.

We consider that a request generated at a transmitter associated with a terminal is active and inactive at p and $1-p$, respectively. We assume that such a request generation process follows an independent and identically distributed (i.i.d) process.

Let us express the ABP condition, which is the one that the blocking probability in TF-Clos must not exceed ABP denoted by a positive real number of ϵ, in the one-step design model in mathematical form. The ABP condition is expressed by the following two sub-conditions. First, the probability that the number of active requests out of n requests associated with an input-output switch exceeds n^{snb}. We have:

$$\sum_{w=n^{\text{snb}}+1}^{n} \binom{n}{w} p^w (1-p)^{n-w} \leq \epsilon. \tag{3}$$

Second, n^{snb}, m, and v must satisfy:

$$2\lfloor \frac{n^{\mathrm{snb}}-1}{v} \rfloor + 1 \leq m, \tag{4}$$

where n^{snb} replaces n in the SNB condition. The work in [11] proved the ABP condition expressed by (3) and (4). Equation (3) localizes blocking occurrences into the first link from the input-output layer to the intermediate layer. We call the first sub-condition in (3) a localized first-link blocking condition (L-FLB). Equation (4) localizes the strict-sense non-blocking condition for any request that passes (3). We call the second sub-condition in (4) a localized SNB (L-SNB). If a request is not blocked by L-FLB, the request will never be blocked in the second link from the intermediate layer to the input-output layer. In TF-Clos not guaranteeing ABP, request blocking can occur at the first link or the second link. We summarize the features of different TF-Clos design types regarding blocking conditions in Table II.

We note that TF-Clos with SNB is a special case of TF-Clos with guaranteeing ABP when $n^{\mathrm{snb}} = n$, which results in SNB. In this case, (3) is always satisfied regardless of any value of p and (4) is the same as the SNB condition.

We describe how TF-Clos of the one-step design model handles a connection request, as depicted in Fig. (3). Suppose that a connection request arrives at an input-output switch. If the number of occupied links from the input-output switch to all intermediate switches, denoted by χ, does not exceed n^{snb}-1, i.e., $\chi \leq n^{\mathrm{snb}}-1$, the request is accepted, and the connection is established. Otherwise, it is rejected. We note that the one-step design model imposes no connection routing policy from a transmitter to a receiver in the same as TF-Clos with SNB.

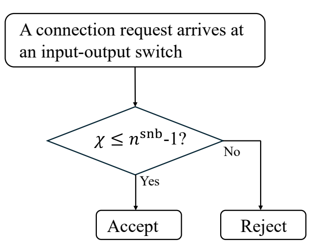


Fig. 3. Connection-request admission procedure in one-step design model. χ denotes the number of occupied links from the input-output switch to all intermediate switches.

We present an optimization problem of the one-step design model for TF-Clos with guaranteeing ABP, given the number of $N \times N$ switches, a, p, and ϵ. The optimization problem of the one-step design model is expressed by:

$$\max \quad nk \tag{5a}$$

$$\text{s.t.} \quad n + vm \leq N \tag{5b}$$

$$vk \leq N \tag{5c}$$

$$\sum_{w=n^{\mathrm{snb}}+1}^{n} \binom{n}{w} p^{w}(1-p)^{n-w} \leq \epsilon \tag{5d}$$

$$2\lfloor \tfrac{n^{\mathrm{snb}}-1}{v} \rfloor + 1 \leq m \tag{5e}$$

$$k + m \leq a \tag{5f}$$

$$n, n^{\mathrm{snb}}, k, m, v \in \mathbb{N}. \tag{5g}$$

The one-step design model of TF-Clos with guaranteeing ABP replaces (2d) with (5d) and (5e). The other constraints are the same between the one-step design model and TF-Clos with SNB.

B. *S-step design model*

The work in [12] developed a TF-Clos design model guaranteeing ABP in a two-step manner by extending the one-step design model. We call the design model a two-step design model. The work adopts a distinctive approach by categorizing intermediate switches into two groups, aiming to relax the constraints of the optimization problem. Such innovation significantly enhances the switching capacity.

The study in [13] further expanded upon the two-step design model, creating a more generalized TF-Clos design model guaranteeing ABP that allows for varying the number of steps, S for intermediate switches. We call it an S-step design model.

Figure 4 presents the TF-Clos structure of the S-step design model consisting of k switches in the input-output layer and m switches in the intermediate layer. $S_i, i \in [1, k]$, denotes an input-output switch, and $S_i, i \in [k+1, k+m]$, denotes an intermediate switch. In the S-step design model, intermediate switches are divided into a maximum of S groups, where S, any positive integer, represents the maximum allowable number of steps, i.e., the maximum allowable number of groups. Each divided subset is called the sth-step group, where $s \in [1, S]$ The number of intermediate switches denotes $m_s, s\check{e}in[1, S]$, where $m = \sum_{s=1}^{S} m_s$. An intermediate switch in the sth-step group denotes $S_i, i \in [k + \sum_{l=i}^{s-1} m_l + 1, k + \sum_{l=i}^{s} m_l]$. There are v_s links from an input-output switch to an intermediate switch in the sth-step group and v_s links from an intermediate switch in the sth-step group to an input-output switch. Table III lists the symbols specifically used in the S-step model. n_s^{snb}, m_s, $v_s, s \in [1, S]$, and $\epsilon_s, s \in [1, S-1]$ are decision variables.

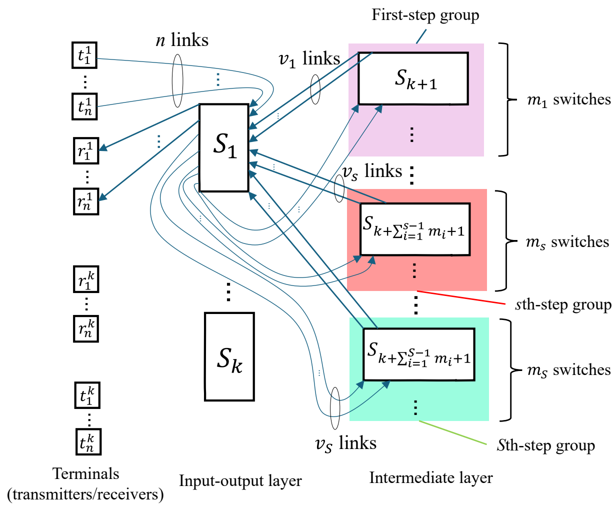
The S-step design model may not need S groups in the intermediate layer, where S is given as the maximum allowable number of groups. The number of groups less than S may guarantee ABP. s_{f} ($\leq S$) denotes the required number of groups for intermediate switches to guarantee ABP; $m_s > 0$ if $1 \leq s \leq s_{\mathrm{f}}$ and $m_s = 0$ otherwise. Designing the TF-Clos structure in the S-step model identifies s_{f}.

We explain how to judge whether the sth step accepts the request, where $s \in [1, s_{\mathrm{f}}]$, as shown in Fig. 5. We extend the idea of the one-step design model to the request-acceptance judgment in the sth step. The s step restricts the maximum number of allowable links to use from an input-output switch to all the intermediate switches in the

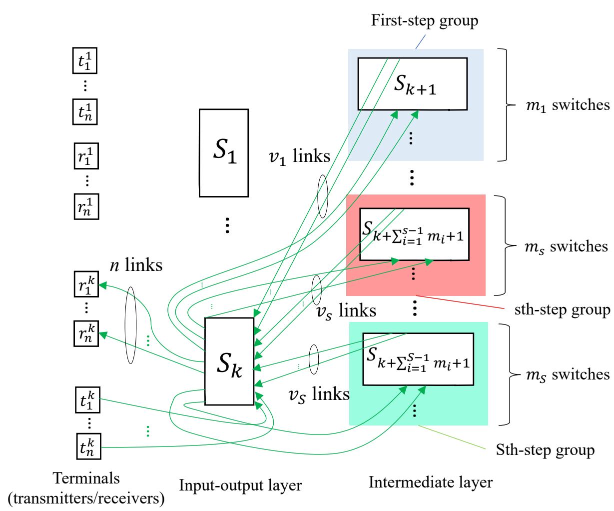
TABLE II
FEATURES OF DIFFERENT TF-CLOS DESIGN TYPES.

Design type	Blocking occurrences		Blocking conditions		
	First link	Second link	SNB, (2d)	L-FLB, (3)	L-SNB, (4)
TF-Clos with SNB	No	No	Satisfied	-	-
TF-Clos with guaranteeing ABP	Yes	No	Not satisfied	Satisfied	Satisfied
TF-Clos without guaranteeing ABP	Yes	Yes	Not satisfied	Not satisfied	Not satisfied

First link: link from input-output layer to intermediate layer
Second link: link from intermediate layer to input-output layer
L-FLB: Localized first-link blocking condition
L-SNB: Localized SNB condition



(a) Structure of terminals, intermediate switches, and links connected to input-output switch S_1.



(b) Structure of terminals, intermediate switches, and links connected to input-output switch S_k.

Fig. 4. TF-Clos structure with S-step groups for intermediate switches; a distinct color encircles each group. The network comprises k input-output switches and $\sum_{i=1}^{S} m_i$ intermediate switches, resulting in a switching capacity of nk.

TABLE III
LIST OF SYMBOLS SPECIFICALLY USED IN S-STEP MODEL.

Parameters	Descriptions
S	Maximum allowable number of steps
Variables	**Descriptions**
n_s^{snb}	Integer. It is the maximum number of allowable links to use from an input-output switch to all the intermediate switches in the sth-step group, $s \in [1, S]$.
m_s	Integer. It is the number of intermediate switches belonging to the sth-step group, $s \in [1, S]$.
v_s	Integer. It is the number of links between an input-output switch and an intermediate switch belonging to the sth-step group, $s \in [1, S]$, in the intermediate layer.
ϵ_s	Real number. It is the upper bound of blocking probability up to the sth step, $s \in [1, S-1]$.

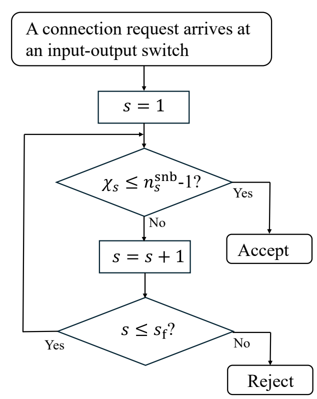


Fig. 5. Connection-request admission procedure in S-step design model. χ_s denotes the number of occupied links from the input-output switch to all intermediate switches in sth-step group.

sth-step group, denoted by $n_s^{\text{snb}} (\le n)$, $s \in [1, s_{\text{f}}]$, to guarantee the admissible blocking probability in the sth step, denoted by ϵ_s, $s \in [1, s_{\text{f}}]$, and $\epsilon_S = \epsilon$. $n_s^{\text{snb}}, s \in [1, s_{\text{f}}]$, is an integer positive decision variable and $\epsilon_s, s \in [1, s_{\text{f}} - 1]$, is a positive real decision variable. n_s^{snb} and ϵ_s are determined by solving the optimization model of the S-step model as described below.

Upon the arrival of a connection request generated by a transmitter at an input port in an input-output switch, it initially attempts to utilize the first-step group for routing to the designated output port. If the acceptance condition in the first step is satisfied, the request is accepted. Otherwise, blocking occurs in the first step, and the request then attempts to use the second-step group. In each step, if the request is not accepted, the number of steps increases as needed to prevent blocking until it reaches s_f. The acceptance condition in the sth step is defined as follows. $\chi_s, s \in [1, s_\mathrm{f}]$, denotes the number of occupied links from the input-output switch to all intermediate switches in the sth-step group. If $\chi_s \leq n_s^{\mathrm{snb}} - 1$ is satisfied, the request is accepted in the sth step, and a connection is established through an intermediate switch in the sth-step group using any available links.

We present an optimization problem of the S-step design model for TF-Clos with guaranteeing ABP, given the number of $N \times N$ switches, a, p, ϵ, and S. The optimization problem of the S-step design model is expressed by:

$$\max \quad nk \tag{6a}$$

$$\text{s.t.} \quad n + \sum_{i=1}^{S} v_i m_i \leq N \tag{6b}$$

$$v_i k \leq N, \forall i \in [1, S] \tag{6c}$$

$$2 \left\lfloor \frac{n_i^{\mathrm{snb}} - 1}{v_i} \right\rfloor + 1 \leq m_i, \forall i \in [1, S] \tag{6d}$$

$$k + \sum_{i=1}^{S} m_i \leq a \tag{6e}$$

$$\sum_{w=n_s^{\mathrm{snb}}+1}^{n-\sum_{j=1}^{s-1} n_j^{\mathrm{snb}}} \binom{n - \sum_{j=1}^{s-1} n_j^{\mathrm{snb}}}{w} \left(p \prod_{i=1}^{s-1} \epsilon_i\right)^w \left(1 - p \prod_{i=1}^{s-1} \epsilon_i\right)^{n-\sum_{j=1}^{s-1} n_j^{\mathrm{snb}} - w} \leq \begin{cases} \epsilon_s, \text{ if } m_{s+1} > 0 \\ \epsilon, \text{ if } m_{s+1} = 0, \end{cases} \quad \forall s \in [1, S-1] \tag{6f}$$

$$\sum_{w=n_S^{\mathrm{snb}}+1}^{n-\sum_{j=1}^{S-1} n_j^{\mathrm{snb}}} \binom{n - \sum_{j=1}^{S-1} n_j^{\mathrm{snb}}}{w} \left(p \prod_{i=1}^{S-1} \epsilon_i\right)^w \left(1 - p \prod_{i=1}^{S-1} \epsilon_i\right)^{n-\sum_{j=1}^{S-1} n_j^{\mathrm{snb}} - w} \leq \epsilon, \text{ if } m_S > 0 \tag{6g}$$

$$m_s = 0, \text{ if } m_{s-1} = 0, \forall s \in [3, S] \tag{6h}$$

$$n, n_1^{\mathrm{snb}}, k, m_1, v_i \in \mathbb{N}, \forall i \in [1, S] \tag{6i}$$

$$n_i^{\mathrm{snb}}, m_i \in \mathbb{N} \cup \{0\}, \forall i \in [2, S] \tag{6j}$$

$$\epsilon \leq \epsilon_1 \leq \cdots \leq \epsilon_{S-1} \leq 1 \tag{6k}$$

$$\epsilon_s \in \mathbb{R}, \forall s \in [1, S] \tag{6l}$$

$$\epsilon_S = \epsilon. \tag{6m}$$

Equation (6a) represents the objective function aimed at maximizing the switching capacity. Equation (6b) denotes the constraint on the number of ports on the input-output switches. Equation (6c) expresses the constraint on the number of ports on the intermediate switches at each step. Equation (6d) illustrates the step-localized SNB condition for each step. Equation (6e) imposes a constraint on the number of available switches. Equations (6f) and (6g) establish the admissible blocking probability conditions to be satisfied at each step. At the sth step, where $s < s_\mathrm{f}$ and $\forall s \in [1, S]$, the left side of (6f) does not exceed ϵ_s. At the s_fth step, the left side of either (6f) or (6g) does not exceed ϵ. The determination of whether s represents the final step s_f is based on whether $m_{s+1} = 0$ holds. Equation (6h) indicates that when the number of intermediate switches reaches zero at a certain step, no further steps are considered in the S-step design model.

IV. Performances

We compare the performances of different design types: UF-Clos with SNB, F-Clos with SNB, TF-CLos with SNB, and TF-Clos with guaranteeing ABP. TF-Clos with guaranteeing ABP is designed by the S-step design model, which covers the one-step model. The S-step design model of TF-Clos uses an exhaustive and binary search combination (EBS) algorithm [13] to obtain an optimal network structure. EBS starts to set the lower bound to the switching capacity of TF-Clos with SNB and the upper bound to a sufficiently large value based on the designer's experience. Using both lower and upper bounds, EBS conducts searching combining the binary search for n and exhaustive search for other decision variables. The design models of UF-Clos, F-Clos, and TF-Clos with SNB employ an exhaustive search algorithm to obtain an optimal structure. We implement the algorithms using C++14. The hardware platform utilized to execute the algorithms was an Intel Xeon Gold 6314U 2.3 GHz 32-core CPU with 128 GB of memory. We use the switch size of $N = 64$ [14] through our evaluation in this paper.

Figure 6 presents the switching capacity depending on arrival rate p in the one-step design model of TF-Clos with $a = 100$ for $\epsilon = 0.001$ and 0.01. The one-step design model is a special case of the S-step design model with $S = 1$. The figure also presents those of UF-Clos, F-clos, and TF-Clos with SNB ($\epsilon = 0$) for reference. The switching capacities of UF-Clos and F-Clos with SNB, which are independent of p, are the same, whereas that of TF-Clos with SNB is larger than them; this is an advantage of the TF-Clos structure, which utilizes the available switching ports efficiently. The switching capacity of the one-step model is larger than that of TF-Clos with SNB; this effect becomes significant as p gets small and ϵ increases. Smaller p and larger ϵ allow TF-Clos to take smaller n^{snb} in (5d), which encourages smaller m in (5e) and larger k (5f). Smaller m promotes larger n in (5b). As a result, the switching capacity of nk in (5a) increases.

Figure 7 observes the switching capacity depending on ϵ in the one-step design model of TF-Clos with $a = 100$ for $p = 0.4, 0.6$, and 0.8. It also presents those of UF-Clos, F-clos, and TF-Clos with SNB for reference. The switching capacity of the one-step model becomes larger than that of TF-Clos with SNB as ϵ increases and p decreases. The switching capacity of the one-step design model increases at the cost of blocking occurrence.

We investigate the switching capacity in the S-step model of TF-Clos with $p = 0.6$ and $\epsilon = 0.000001$ for different values of $a = 30$ and 60, as shown in Fig. 8. The

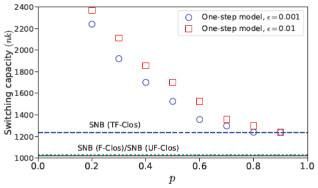


Fig. 6. Switching capacity depending on arrival rate p in one-step model of TF-Clos with $N = 64$ and $a = 100$.

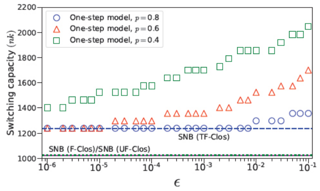


Fig. 7. Switching capacity depending on admissible blocking probability ϵ in one-step model of TF-Clos with $N = 64$ and $a = 100$.

switching capacities of UF-Clos, F-clos, and TF-Clos with SNB ($\epsilon = 0$) are also presented for reference. As the number of steps, S, increases, the switching capacity increases. Increasing S indicates that the types of decision variables, n_s^{snb}, m_s, $v_s, s \in [1, S]$, and $\epsilon_s, s \in [1, S-1]$, increases. This allows us to have the more flexible design of TF-Clos, where a solution region in the optimization problem in (6a)–(6m) becomes large. This results in increasing the switching throughput. We observe that increasing a from 20 to 60 increases the switching capacity.

Figure 9 shows the switching capacity gain in the S-step design model of TF-Clos with $p = 0.6$ and $\epsilon = 0.000001$ for different values of $a = 30$ and 60. The switching capacity gain defines the switching capacity with S divided by that of $s = 1$, i.e., the one-step model. As s increases, the switching capacity increases; the gain is 1.28 with $a = 30$ and 1.11 with $a = 60$. The switching-capacity increasing effect becomes large as a gets smaller. Since smaller a has a tighter bound in the optimization problem in (6a)–(6m), increasing S relaxes the tighter bound with smaller a.

Figure 10 show the computation time depending on the number of steps, S, in the S-step design model of TF-Clos with $p = 0.6$, $a = 60$, and $\epsilon = 0.000001$. As S rises, so does the quantity of decision variables to be explored, consequently leading to increased computation time in each algorithm. When $S = 5$, the computation time reaches 6.3×10^5 [s] (7.3 days). The S-step design model proves useful during the network design phase before providing services, with computation times deemed acceptable spanning several weeks.

Next, let us consider another design scenario in which we minimize the required number of switches, a, to achieve a given targeted switching capacity, denoted by ξ, whereas, so far, we have considered maximizing the switching capacity, under the condition that a is given as a parameter. To obtain the minimum value of a, one may simply consider that one defines a as a decision variable, instead of a given parameter, replaces $\min a$ with $\max nk$ in (6a), adds $\xi \geq nk$ as a constraint in the optimization problem in (6a)–(6m), and directly solve the optimization problem. However, this approach does not work well from a viewpoint of computation time to solve the optimization problem. To find the minimum a, we must explore all possible values of $\xi (\geq nk)$ since an optimal structure does not always exist when $\xi = nk$. We employ a binary search to find such a minimum of a. First, we give the upper and lower bounds of a. For example, the upper bound is set to a sufficiently large value of a based on the designer's experience, and the lower bound to $\lceil \xi / N \rceil$. Next, we conduct a binary search for a. Each iteration of the binary search solves the optimization problem in (6a)–(6m) by using EBS and checks whether the obtained switching capacity is higher than or equal to ξ. Based on the binary search, the minimum value of m is obtained.

Figure 11 presents the required minimum number of switches achieving the targeted switching capacity $\xi = 750$ in the S-step model of TF-Clos with $p = 0.6$ for defferent values of ϵ and S. Similarly, the required numbers of switches of UF-Clos, F-clos, and TF-Clos with SNB ($\epsilon = 0$) are also presented for reference. In the SNB condition ($\epsilon = 0$), the required minimum number of switches of TF-Clos is smaller than those of UF-Clos and F-Clos. As ϵ increases, the required minimum number of switches in the S-step design model decreases with the same S. Larger ϵ leads to taking smaller $n_s^{\text{snb}}, \forall s \in [1, S]$. This relaxes other constraints in the S-step design model and promotes a smaller required minimum number of switches. Furthermore, larger S reduces the required minimum number of switches; increasing S gives more decision variables and makes the solution space of the TF-Clos structure larger. Figure 11 observes that the number of input-output switches and the number of intermediate switches at each group vary depending on ϵ and S.

Figure 12 shows the required number of links achieving targeted switching capacity $\xi = 750$ in the S-step design model of TF-Clos with $p = 0.6$. We depict each result in Fig. 12 obtained from the TF-Clos structure in Fig. 11. The required number of links tends to decrease in an increase of ϵ and S, which is the same trend of the minimum required number of stitches excluding an exception case of $S = 4$ with $\epsilon = 0.0001$. This exception indicates that the required number of links in Fig. 11 is not always the minimum required number of links using the minimum number of switches. If the minimum number of switches is given, we can have one or more TF-Clos structures. We discuss this point in Section V-D.

V. Research challenges

A. Selection of suitable number of steps S

In the S-step design model, as S increases, the switching capacity increases in some case, whereas it does not in some case, as shown in Fig. 8. This behavior affected by S is not surely analyzed. 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. An extensive approach to the S-step

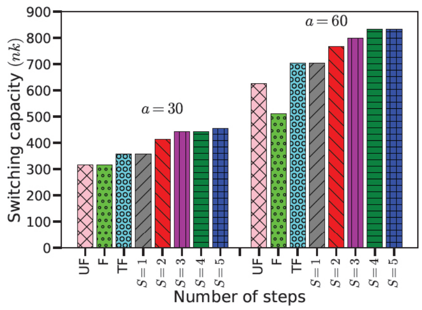


Fig. 8. Switching capacity in S-step model of TF-Clos with $N = 64$, $p = 0.6$, and $\epsilon = 0.000001$.

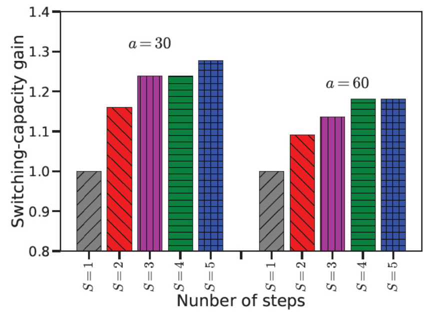


Fig. 9. Switching-capacity gain in S-step model of TF-Clos with $N = 64$, $p = 0.6$, and $\epsilon = 0.000001$.

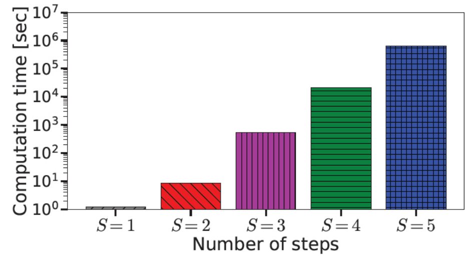


Fig. 10. Computation time depending on number of steps in S-step model of TF-Clos with $N = 64$, $p = 0.6$, $a = 60$, and $\epsilon = 0.000001$.

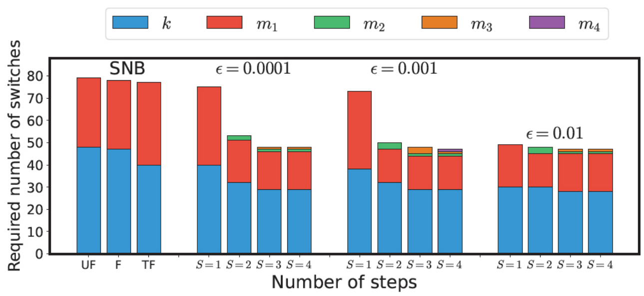


Fig. 11. Required minium number of switches that archives targeted switching capacity $\xi = 750$ in S-step model of TF-Clos with $N = 64$ and $p = 0.6$.

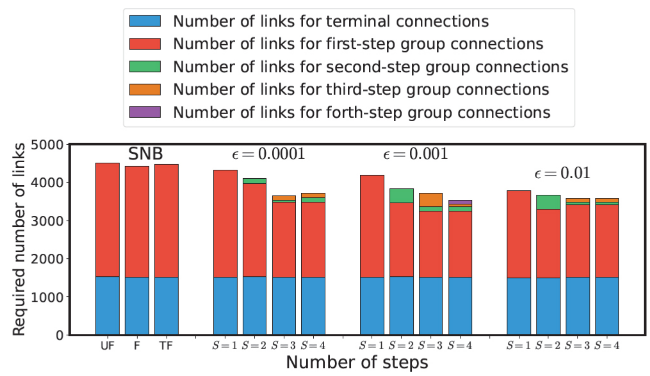


Fig. 12. Required number of links that archives targeted switching capacity $\xi = 750$ in S-step model of TF-Clos with $N = 64$ and $p = 0.6$.

design model formulates an optimization problem that uses S as a decision variable instead of a given parameter. It should also be addressed for further work.

B. 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.

The bottleneck in computation time stems from the processing limitations of a single CPU rather than memory constraints. Employing parallel processing to exploit multiple CPU cores can alleviate the overall computation time when searching for the optimal TF-Clos structure. However, this poses an area of research challenge.

Another approach to tackle this issue involves crafting an approximation algorithm, such as a meta-heuristic employing stochastic elements, to diminish computation time while retaining the advantages of scaling the number of steps. A network designer weighs the trade-off between solution precision and computation time for practical implementation, which warrants further investigation.

C. Handling different traffic rates generated at different terminals

This paper has assumed that all the traffic rate generated at each terminal is the same as p, but the traffic rate generated at each terminal can differ. We address how to handle this case in the TF-Clos design model guaranteeing ABP. Three possible strategies exist.

The first strategy adopts the largest value of traffic rate as p. A network designer can simply determine the TF-Clos structure using the design model provided in this paper as it is, which is an advantage of this strategy, but it is highly conservative.

The second strategy, suggested by [13], adopts the $n \in [1, N]$ largest rates, $p_i, i \in [1, n]$, among all the traffic rates. We compute the probability function of the number of active requests by convoluting the largest n rates, each of which follows an i.i.d process. In the one-step design model, we replace the internal term of the summation in the left-hand side of (3) with the probability function after the convolution. The S-step design model can also adopt a similar with some modifications. This computational arrangement does not alter the position of each transmitter/receiver or the connecting relationship between each transmitter/receiver and each input-output switch. The second strategy is conservative but not higher than the first one; the strategy has advantages.

The third strategy relaxes the constraint regarding the connecting relationship between each transmitter/receiver and each input-output switch. Instead of the $n \in [1, N]$ largest rates adopted in the second strategy, the third strategy allows us to shuffle the connecting relationship so that each probability function associated with n traffic rates can be balanced to minimize n^{snb}. This strategy is still conservative and can increase the switching capacity compared to the second one at the expense of network deployment and computational time complexities. This direction should be addressed in future work.

D. Suppressing amount of switches and links

Section IV has discussed the performances of two design scenarios of TF-Clos. One maximizes the switching capacity, nk, provided that the number of switches, a, is given. The other scenario minimizes the required number of switches, a, to achieve a given targeted switching capacity, ξ. This section focuses on the second scenario.

If a targeted switching capacity, ξ, is given, a network designer wants to design TF-Clos with a suppressing amount of switches and links. For this purpose, two possible strategies exist. First, if the number of switches is a more critical concern than the number of links, a network designer minimizes the required number of switches and obtains the least number of links. As we discussed in Fig. 11 in Section IV, the required number of links is not always the minimum required number of links using the obtained required minimum number of switches. Therefore, we use the obtained required minimum number of switches, denoted by $a_{\min}$, as a given parameter to find the TF-Clos structure to minimize the number of links. In the S-step design model expressed by (6a)–(6m), we replace $\max nk$ in (6a) with $\min \sum_{i=1}^{S} v_s$, replace $k+\sum_{i=1}^{S} m_i \leq a$ in (6e) with $k+\sum_{i=1}^{S} m_i \leq a_{\min}$, and add a constraint of $nk = \xi$. Second, suppose that the number of switches is not a more critical concern than the number of links and it is necessary to consider the weighted sum of the number of switches and links. In that case, we must redefine an optimization problem in (6a)–(6m). Addressing two design strategies to suppress the amount of switches and links is left for future work.

E. Increasing number of stages

This paper have focused on the structure of TF-Clos, where an established connection passes through three switches, i.e., an input-output switch, an intermediate switch, and an input-output switch, from a transmitter to a receiver.

Study [15] provided variable types of F-Clos. It presented extensions of F-Clos from two to three stages to expand the switching capacity. Three types of three-stage F-Clos structures exist; the study proves that the three types are isomorphic if each switch size can be flexibly adjusted.

The TF-Clos design model can employ this extension idea to increase the number of stages provided by [15] to increase the switching capacity. However, we impose the switch-size restriction of our used switches, such as $N \times N$.

VI. Conclusions

With future data centers poised to adopt cutting-edge circuit switching technologies, such as optical switching, the potential for heightened transmission capacity and energy efficiency is promising. Through an examination of Clos-network structures, renowned for their scalability, this paper has reviewed research progress to boost switching capacity while maintaining internal blocking quality. Analysis of Clos network design models has yielded valuable insights into their characteristics and performance metrics, including switching capacities and computation times. By addressing the research challenges identified in this review, future studies can concentrate on further enhancing the switching capacity of Clos networks, thereby contributing to the evolution and optimization of data center networking infrastructure.

References

[1] K.-I. Sato, “Optical switching will innovate intra data center networks [invited tutorial],” *J. Opt. Commun. Netw.*, vol. 16, no. 1, pp. 1–23, Jan 2024.

[2] M. Taubenblatt, P. Maniotis, and A. Tantawi, “Optics enabled networks and architectures for data center cost and power efficiency,” *J. Opt. Commun. Netw.*, vol. 14, no. 1, pp. A41–A49, Jan. 2022.

[3] Y. Shen, W. Wang, T. Liu, and J. Zhang, “Energy-efficient scaling of active electrical/optical switches in hybrid packet/circuit switched data center networks,” in *Int. Conf. Opt. Commun. Netw.*, 2021, pp. 1–3.

[4] C. Clos, “A study of non-blocking switching networks,” *Bell Syst. Tech. J.*, vol. 32, no. 2, pp. 406–424, 1953.

[5] A. Jajszczyk, “Nonblocking, repackable, and rearrangeable Clos networks: fifty years of the theory evolution,” *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 28–33, 2003.

[6] W. Kabaciński, *Nonblocking Electronic and Photonic Switching Fabrics.* Springer, 2005.

[7] E. Oki, Z. Jing, R. Rojas-Cessa, and H. Chao, “Concurrent round-robin-based dispatching schemes for Clos-network switches,” *IEEE/ACM Trans. Netw.*, vol. 10, no. 6, pp. 830–844, 2002.

[8] E. Oki, N. Yamanaka, K. Nakai, and N. Matsuura, “Multi-stage switching system using optical WDM grouped links based on dynamic bandwidth sharing,” *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 56–63, 2003.

[9] T. Mano, T. Inoue, K. Mizutani, and O. Akashi, “Increasing capacity of the Clos structure for optical switching networks,” in *2019 IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.

[10] ——, “Redesigning the nonblocking Clos network to increase its capacity,” *IEEE Trans. Netw. Service Manag.*, vol. 20, no. 3, pp. 2558–2574, Sep. 2023.

[11] 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.

[12] ——, “Twisted and folded Clos-network design model with two-step blocking probability guarantee,” *IEEE Netw. Lett.*, vol. 6, no. 1, pp. 60–64, Mar. 2024.

[13] ——, “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.

[14] B. G. Lee and N. Dupuis, “Silicon photonic switch fabrics: Technology and architecture,” *J. Lightwave Technol.*, vol. 37, no. 1, pp. 6–20, 2019.

[15] A. Jajszczyk, “On combinatorial properties of broadband time-division switching networks,” *Comput. Netw. ISDN Sys.*, vol. 20, no. 1, pp. 377–382, 1990.