Content Evacuation in Inter-DC Optical Networks under Post-Disaster Cascading Failures

Zhuotong Li
State Key Laboratory of Information Photonics and Optial Communications
Beijing University of Posts and Telecommunications
Beijing, China
lizhuotong@bupt.edu.cn

Memedhe Ibrahimi
Department of Electronics, Information and Bioengineering
Politecnico di Milano
Milan, Italy
memedhe.ibrahimi@polimi.it

Yongli Zhao
State Key Laboratory of Information Photonics and Optial Communications
Beijing University of Posts and Telecommunications
Beijing, China
yonglizhao@bupt.edu.cn

Biswanath Mukherjee
Department of Computer Science
University of California
Davis, USA
bmukherjee@ucdavis.edu

Jie Zhang
State Key Laboratory of Information Photonics and Optial Communications
Beijing University of Posts and Telecommunications
Beijing, China
lgr24@bupt.edu.cn

Massimo Tornatore
Department of Electronics, Information and Bioengineering
Politecnico di Milano
Milan, Italy
massimo.tornatore@polimi.it

***Abstract*—In the post-pandemic era, global work patterns have been reshaped, and the demand for cloud migration for enterprises and government has increased. As a result, cloud data disaster backup/recovery technology has been gaining more attention. Moving beyond the traditional focus on pre-disaster content backup, our study addresses the challenge of rapidly evacuating content during cascading failures in post-disaster scenarios. Due to the interdependence of i) data centers (DCs), ii) inter-DC optical networks, and iii) power grid networks, disasters can have a domino effect on these infrastructures, with their impact gradually expanding over time and space. In this work, we propose two trajectory models for constructing the spatio-temporal features of the inter-DC optical network under cascading failures, and we propose a trajectory-based content evacuation strategy (TCE). Numerical results show that TCE can reduce content loss by up to 25% compared to baseline content evacuation strategies.**

Keywords—content evacuation, cascading failure, Inter-DC, trajectory model, spatio-temporal feature.

I. INTRODUCTION

In the wake of the COVID-19 pandemic, the world witnessed an unprecedented shift towards remote work, digital transformation, and accelerated adoption of cloud computing technologies. However, as the world experiences a disconcerting surge in the frequency and severity of natural and man-made disasters [1], this surge puts into the spotlight the importance of safeguarding essential infrastructures, such Data Centers (DC) and inter-DC optical networks that provide the high-speed transmission at the foundation of today's cloud applications. In a scenario when incoming disaster provides an early warning, cloud operators can use the remaining time before the disaster strikes to *evacuate content* from high-risk to safe DCs. Then, even after the disaster has hit, the disaster aftermath can affect the inter-DC network beyond the initial disruptions. Today, there are complex interdependencies between inter-DC optical networks and the power grid. Optical network equipment (e.g., switches) are powered by the power grid, while a power grid depends on communication network (e.g., to support the Supervisory Control And Data Acquisition (SCADA) system) for remote monitoring, measurement, and control. Thus, the failure of an element in one network may cascade to the other network and vice-versa, hereafter referred to as *cascading failures*. Further cascading failures might instigate a domino effect across the two infrastructures (optical network and power grid) that can culminate in extensive network and DC outages [2-3]. For example, in 2020, cascading failures that occurred in CenturyLink caused outages and service interruptions for global cloud providers, including Google and Cloudflare [4]. Similarly, an unprecedented heatwave and tropical cyclones sparked massive power outages in Canada, Cuba, and the US in 2022, when many cascading failures resulted in significant DC content loss [5]. Hence, it is crucial to develop strategies to urgently evacuate contents from risky DCs to safe DCs before cascading failures cause irreversible damage.

The propagation of cascading failures is a complex and dynamic phenomenon that can cause the topology and path resources of inter-DC optical networks to change over time, which we call a spatio-temporal evolution problem. However, most previous studies on content evacuation investigated short-term approaches, focusing on the immediate consequences of a disaster [6-7], while our work adopts a longer-term approach covering cascading failures. Ref. [8] delves into the intricacies of the emergency backup challenge between DCs within the context of progressively unfolding disaster scenarios. Nonetheless, this approach only incorporates the availability of DCs, but disregards the resource constraints inherent in optical networks, which is a critical aspect given that the post-disaster domino effect invariably influences changes in the optical network.

In this work, for the first time, we address the problem of content evacuation under cascading failures in inter-DC optical network. We construct new trajectory models that capture the dynamic evolution of cascading failures, and we propose the new ***T**rajectory-based **C**ontent **E**vacuation* strategy (TCE), which considers the spatio-temporal evolution of cascading failures to minimize content loss. Illustrative numerical results demonstrate that, compared to short-term

This work has been supported in part by National Key Research and Development Program of China (2020YFB1805602), Funds for Creative Research Groups of China (62021005), China Scholarship Council, and the US-Japan JUNO3 project: NSF Grant no. 2210384.

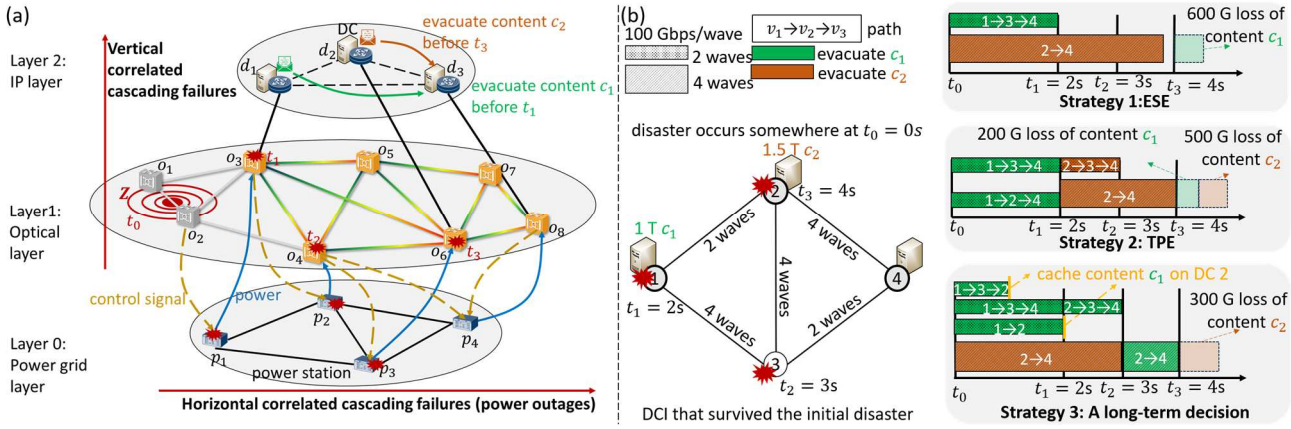

Fig. 1. (a) Horizontal and vertical cascading failures after a disaster; (b) different evacuation strategies.

decision strategies, TCE can reduce the content loss ratio by 25% when evacuating content from risky DCs to safe DCs.

II. Content Evacuation under Cascading Failures

Figure 1(a) illustrates how a disaster can cause a series of cascading failures in a DCs/power-grid/optical-network interdependent system. DCs predominantly rely on Inter-DC optical networks to facilitate high-speed data content transmission. Inter-DC optical networks depend on power stations to supply the necessary electrical power, while power stations, in turn, rely on optical networks for the seamless transmission of control signals. In the event of a disaster $\boldsymbol{Z}$ striking the area shown in Fig. 1(a) at time t_0, immediate damage occurs to nodes o_1 and o_2 in the optical network. Consequently, the lack of control signals from o_1 leads to the cascaded failure of p_1 in the power grid. As p_1 provides power to o_3, a power outage eventually causes o_3 to go offline. Subsequently, a cascading failure affects the upper IP layer due to the lack of available lightpaths. Consequently, DC d_1 in the IP layer needs to evacuate its content before reaching the critical time point denoted as t_1. In a similar chain of events, p_2, o_4, p_3, and p_4 will fail in a domino effect. Eventually, the content hosted in DC d_2 needs to be evacuated before t_3. Hence, contents c_1 (hosted in DC d_1) and c_2 (hosted in DC d_2) will have different evacuation deadlines; therefore, traditional content evacuation strategies are not suitable in case of cascading failures. As cascading failures propagate, path resources in the optical network will also become unavailable over time (e.g., o_4 will be offline at t_2), which is a spatio-temporal evolution scenario.

Figure 1(b-left) shows a scenario where cascading failures propagate in a small inter-DC topology. For clarity, the disaster occurrence at time $t_0 = 0s$ and the power grid are omitted from the illustration. If the cascading failures propagate to node 1, 2, and 3 successively, then we need to evacuate different contents c_1 (hosted in DC1) and c_2 (hosted in DC2) to safe DC4 within the corresponding time windows. Figure 1(b-right) depicts three possible evacuation strategies. **Strategy 1** starts to evacuate c_1 and c_2 immediately after the disaster occurrence [9], which is a common approach in studies on the short-term effects of disasters. We refer to this strategy as ***Equal Synchronization Evacuation*** (ESE). **Strategy 2** uses all available resources to evacuate c_1, which is affected first by cascading failures, and then uses available resources to evacuate c_2; we refer to this strategy as a ***Time-Priority Evacuation*** (TPE). **Strategy 3**, proposed by us, is called TCE, which temporarily stores and caches part of c_1 at intermediate DC 2, and then evacuates both contents in DC2 to a safe node, which is a long-term decision, since it has the lowest content loss. The objective of our work is to find the optimal content evacuation strategy for this spatial and temporal problem.

III. Trajectory-Based Content Evacuation Strategy

Three interdependent infrastructures are considered: 1) a power grid network $\boldsymbol{G}_p$; 2) an optical network $\boldsymbol{G}_o$ connected to $\boldsymbol{G}_p$ through directed dependency edges $\boldsymbol{E}_{op}$; 3) a set of DCs $\boldsymbol{V}_d$ connected to $\boldsymbol{G}_o$ through undirected edges $\boldsymbol{E}_{od}$. Given a disaster zone $\boldsymbol{Z}$ that might affect the three infrastructures, and a set of contents $\boldsymbol{C}$ hosted in DCs $v_d \in \boldsymbol{V}_d$ with a corresponding importance metric ε_c for each content c, the objective is to minimize the content loss ratio L:

$$\min L = 1 - \left(\frac{\sum_{c \in \boldsymbol{C}} \varepsilon_c \cdot A_c'}{\sum_{v_d \in \boldsymbol{F}} \sum_{c \in \boldsymbol{C}} \varepsilon_c \cdot A_{v_d,c}} \right) \tag{1}$$

where A_c' is the total amount of content c evacuated to safe DCs, and $A_{v_d,c}$ is the size of content c hosted in DC v_d. $\boldsymbol{F}$ is the set of DCs that will be affected by cascading failures.

To deal with this spatio-temporal scheduling problem caused by cascading failures, we first introduce two concepts: (*i*) **trajectory of a content** in DCs represented as $\delta_{v_d,c}$ to record dynamic changes of content (especially cached content) and (*ii*) **trajectory of available resources** in networks represented as $\boldsymbol{\rho} = [t^b_{lo,\lambda}, t^e_{lo,\lambda}]$ to record dynamic changes of link resources. $\delta_{v_d,c}$ indicates the timestamp of content c hosted in DC v_d, which means the time when content c was hosted/cached in v_d. If content c was hosted in v_d before the disaster occurrs, then $\delta_{v_d,c} = 0$. If it was transferred to v_d for caching and relaying during the evacuation (i.e., v_d is an intermediate DC), then $\delta_{v_d,c}$ is the time when c is fully transferred to v_d. $\boldsymbol{\rho} = [t^b_{lo,\lambda}, t^e_{lo,\lambda}]_n$ represents the available time interval during which

wavelength λ can be used on link l_0. $t^{b}_{l_0,\lambda}$ and $t^{e}_{l_0,\lambda}$ represent the beginning and ending timestamp, respectively. n indicates that the link may have multiple discrete time periods.

Figure 2 provides an example of these two trajectories. The disaster occurs at $t_0 = 0s$, and cascading failures will destroy the corresponding nodes 1, 2, and 3 in order at $t_1 = 1s$, $t_2 = 2s$, and $t_3 = 3s$. Assume that c_1 cannot be completely evacuated before t_1. If we choose to transfer 50GB to node 2 for relay and caching during the evacuation of c_1 and evacuate the remaining 100GB to node 4 using λ_1, the transmission of these two processes will be completed in 0.5s and 1s, respectively. Then DC2 has trajectories of contents: $\delta_{v_1,c_1} = 0.5$ and $\delta_{v_1,c_2} = 0$. The trajectories of available resources on λ_1 on links $l_{1\to2}$ and $l_{3\to4}$ will be $\boldsymbol{\rho}_{l_{1\to2},\lambda_1} = [0.5, 1]$ and $\boldsymbol{\rho}_{l_{3\to4},\lambda_1} = [1, 2]$, which represent the period of time they are available before failures. Based on these two trajectories, for each wavelength λ on each content evacuation path p, the size of content that can be evacuated within the available time window is:

$$A^{c}_{p,\lambda} = B \cdot \left(T^{c}_{p,\lambda} - \max\{\delta_{v_d,c}, t^{b}_{p,\lambda}\} \right) \tag{2}$$

where $T^{c}_{p,\lambda}$ is the final timestamp when content c was transmitted on wavelength λ of path p:

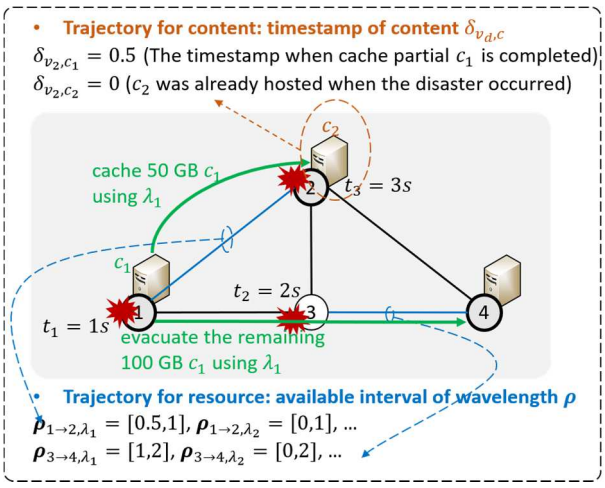

Fig. 2. Trajectory model for evacuation strategy.

$$T^{c}_{p,\lambda} = \min\left\{ t^{e}_{p,\lambda}, \max\left\{\delta_{v_d,c}, t^{b}_{p,\lambda}\right\} + \frac{A_{v_d,c}}{B} \right\} \tag{3}$$

Based on our proposed trajectory models, we design a heuristic content evacuation strategy, called TCE. The pseudocode of the TCE strategy includes three steps:

Step 1 (line1-line3): According to the disaster zone and three dependent networks, TCE gathers information on the network nodes that will be affected by cascading failures (risky nodes) and the DCs that will not be affected (safe DCs). It initializes evacuation deadlines of all risky contents and the available time interval of all network components, which is the trajectory of available resources $\boldsymbol{\rho}$.

Step 2 (line4-line11): This step is for searching evacuation path and spatio-temporal scheduling for content. TCE calculates k shortest paths for each safe DC node and risky DC node pair. For each path, TCE preferentially uses the wavelength with the longest available duration for content evacuation. Based on the trajectory constrains, the size of the content that can be evacuated within the corresponding time window can be calculated multiple times according to Eq. (2) and Eq. (3). Then, the available interval $\boldsymbol{\rho}$ of each link's resource is updated.

Step 3 (line12-line19): If the content cannot be completely evacuated before the deadline, TCE will assign as intermediate DC v_d^{interm} for caching content, the available DC with the smallest ratio $\theta_{v_d^{\text{interm}},c}$ between the size of the hosted content and the time point affected by the cascading failure

Algorithm 1: Trajectory-based Content Evacuation (TCE)

Input: $\boldsymbol{G}_p$, $\boldsymbol{G}_o$, $\boldsymbol{V}_d$, $\boldsymbol{E}_{op}$, $\boldsymbol{E}_{od}$, $\boldsymbol{Z}$, $\boldsymbol{C}$;
Output: A_c';

// Step 1: Disaster and cascading failure mapping
1: Disaster $\boldsymbol{Z}$ occurs at t_0;
2: Get $\boldsymbol{F}$ according to $\boldsymbol{E}_{op}$, $\boldsymbol{E}_{od}$, and initialize all $\boldsymbol{\rho}$ and $\delta_{v_d,c}$ according to failure model and the propagation time of each hop of cascading failures τ;
3: Get the set $\boldsymbol{D}_z$ of risky DCs within $\boldsymbol{Z}$ and safe DC set $\boldsymbol{D}_z'$ which won't be affected
// Step 2: Evacuation path selection and spatio-temporal scheduling
4: **for** each $v_d \in \boldsymbol{D}_z$ **do**
5: calculate k shortest paths as $\boldsymbol{P}_c$ for all node pair $< v_d, v_d' >$, $v_d' \in \boldsymbol{D}_z'$ and sort c hosted in v_d in increasing order of $\delta_{v_d,c}$; $\boldsymbol{D}_z = \boldsymbol{D}_z / v_d$;
6: **for** each c hosted in v_d **do**
7: **for** each $p_c \in \boldsymbol{P}_c$ **do**
8: sort all available wavelengths as a set $\boldsymbol{\Lambda}^{c}_{p_c}$ in descending order of available duration;
9: **for** each $\lambda \in \boldsymbol{\Lambda}^{c}_{p_c}$ **do**
10: **while** $A_{v_d,c} > 0$ **do**
11: calculate $T^{c}_{p_c,\lambda}$ and $A^{c}_{p_c,\lambda}$ according to Eq. (2)(3); update all $\boldsymbol{\rho}$ of each link;
$A_c' \leftarrow A_c' + A^{c}_{p,\lambda}$; $A_{v_d,c} \leftarrow A_{v_d,c} - A^{c}_{p,\lambda}$;
\\ Step 3: Content relay and cache
12: **if** $A_{v_d,c} > 0$ **then**
13: find the DC $v_d^{\text{interm}} \in \boldsymbol{D}_z$ for caching according to $\min_{c \in \boldsymbol{C}} \theta_{v_d^{\text{interm}},c} = A_{v_d^{\text{interm}},c} / t_{v_d^{\text{interm}}}$ and calculate the k shortest path set $\boldsymbol{P}_c'$ for $< v_d, v_d^{\text{interm}} >$;
14: **for** each $p_c' \in \boldsymbol{P}_c'$ **do**
15: sort all available wavelengths as a set $\boldsymbol{\Lambda}^{c}_{p_c'}$ in descending order of available duration;
16: **for** each $\lambda \in \boldsymbol{\Lambda}^{c}_{p_c'}$ **do**
17: **while** $A_{v_d,c} > 0$ **do**
18: calculate $T^{c}_{p_c',\lambda}$ and $A^{c}_{p_c',\lambda}$ according to Eq. (2)(3); update all $\boldsymbol{\rho}$ of each link;
$A_{v_d^{\text{interm}},c} \leftarrow A_{v_d^{\text{interm}},c} + A^{c}_{p_c',\lambda}$;
$A_{v_d,c} \leftarrow A_{v_d,c} - A^{c}_{p_c',\lambda}$;
19: **update** $\delta_{v_d^{\text{interm}},c}$

among the remaining risky DCs. Similarly, TCE schedules the content to be evacuated to the intermediate DC. While updating the available interval of resources, the trajectory of the content in the intermediate DC $\delta_{v_d^{\text{interm}},c}$ is updated as well, i.e., the timestamp is assigned the value of the time when the content is completely transmitted to this intermediate DC.

IV. Numerical Results

We use the USNET (red nodes indicate the optical nodes that host a DC) and IEEE-14BUS dependent infrastructures shown in Fig. 3 and put them in the same coordinate system (0,1) to simulate the geographical location of the networks. The dependencies of the three heterogeneous networks are constructed and shown according to Ref. [10]. Circle disaster zones are generated at a random location with a radius of r. In the coordinate system, a disaster will destroy all heterogeneous network elements at the same location within the range. For each simulation, we repeated disaster zone 50 times at different locations to get an average result. Each link in the optical network has 40 wavelengths. In this work, the available storage space in each DC is not considered. Actually, network operation and maintenance personnel usually discover and intervene after cascading failures occur for a period of time. Therefore, in our simulation, the number of cascading failure hops in the optical network is controlled within different hops. The content size in each DC is randomly generated according to different interval ranges set in different simulations. TPE and ESE are considered as benchmarks in this work. TPE will calculate the content evacuation paths of all risky DCs without distinction after cascading failure mapping, and allocate optical path resources in the order of calculation. ESE will prioritize the calculation and allocation of content evacuation paths for risky DCs with short content evacuation deadlines in the order in which cascading failures occur.

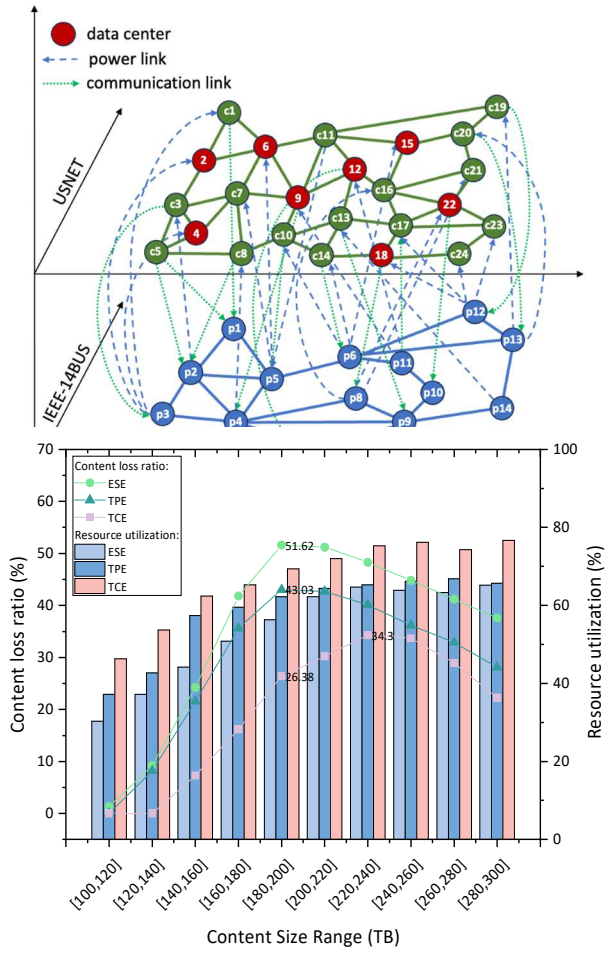

Fig. 4. Content loss ratio and resource utilization under different content size.

Figure 4 shows the content loss ratio L and resource utilization under different initial setting of content sizes. Since cascading failures in power grid usually occurs within seconds to minutes [11], in this simulation we set the time interval between two cascading failures (a hop) as 3s and the cascading failure hops are controlled within 3 hops. We can observe that TCE always outperforms TPE and ESE in terms of content loss ratio, and all three present convex curves with peaks as the content size of each data center increases. Specifically, as the content size increases, the content loss degree of the three strategies will increase, and ESE and TPE reach the peak when each data center accommodates data sizes between [180TB, 200TB], which are 51.62% and 43.03%, respectively. At this time, the content loss of TCE is 26.38%, which was reduced by 25.24% and 16.65% compared to ESE and TPE, respectively. TCE only reaches its peak at [220TB, 240TB], and the peak value is 34.3%, which is 17.32% and 8.73 lower than the peak values of ESE and TPE, respectively. This is due to TCE choosing other intermediate DCs as caching nodes for content that cannot be evacuated within the short time window before the failure. Further increasing the content size decreases the loss rate because the total content that needs to be evacuated is larger. However, all three strategies have reached the maximum content evacuation capacity. In addition, we also report the resource utilization of three strategies. In this work, since the network is investigated in both its spatial and temporal features, we count the resource utilization of each link during its available time (the total statistical time is from the time of the disaster to the occurrence of the last cascading failure). It can be seen that TCE makes the most use of network resources among the three, up to 76%, compared to the baseline approaches that utilize only up to about 65%.

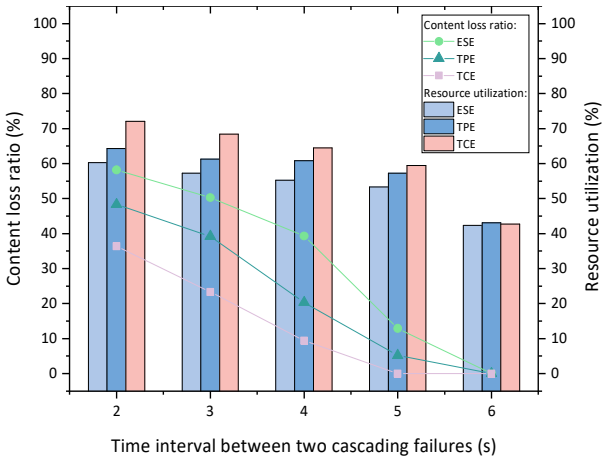

Fig. 5. Content loss ratio and resource utilization under different time interval between two cascading failures.

We also report the content loss ratio and resource utilization under different time intervals between two cascading failures, which is shown in Fig. 5. As the time interval between two cascading failures increases from 2s to 6s (content size range: [200TB, 220TB]), the content loss ratios of the three strategies gradually decrease, while TCE still has the lowest content loss. The content loss rate of TCE

drops to zero when the time interval of cascading failures is 5s, while the content loss rate of the other two strategies reaches zero at 6s. In addition, according to Fig. 5, the resource utilization of the three strategies continues to decrease with increase of the time interval of cascading failures. The reason is that, when the time interval becomes longer, the content of each risky DC has completed evacuation and transmission within the corresponding time window. Therefore, within the statistical time window, there will be a certain period of time when the network link resources are not used after the contents are fully evacuated.

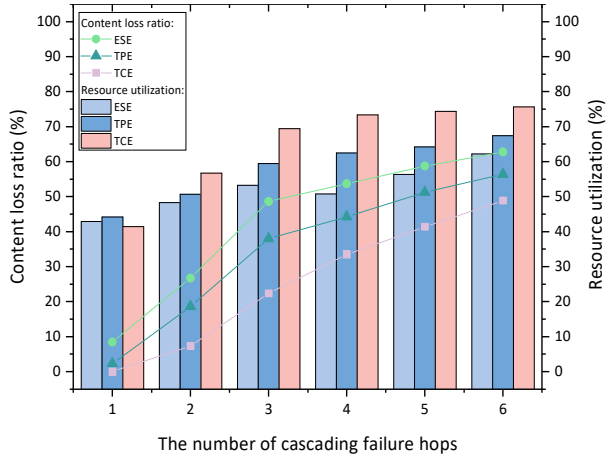

Fig. 6. Content loss ratio and resource utilization under different cascading failure hops.

Figure 6 shows the content loss ratio and resource utilization under different cascading failure hops. Fewer such hops means that the subsequent impact of the disaster will be smaller, so the number of risky DCs will be smaller. As the cascading failure hops increases, more content in risky DCs needs to be evacuated, so the content loss ratios of the three strategies will gradually increase, but TCE still maintains good performance. Similarly, for resource utilization, an increase in the number of hops means that there are more link resources in the network for evacuation of data center content, and there are data flows in and out of the intermediate cache nodes in both directions at the same time. Therefore, resource utilization for the three strategies will continue to increase.

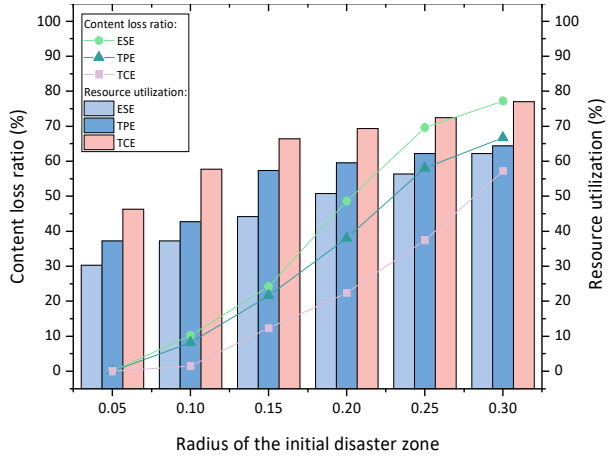

Fig. 7. Content loss ratio and resource utilization under different radius of the initial disaster zone.

Last, we modified the size of the initial disaster zone to observe the performance of the three content evacuation strategies. As shown in Fig. 7, as the damage area of the initial disaster gradually increases, more cascading failures will occur, resulting in more risky DCs which will be affected by cascading failures requiring the evacuation of their contents. The content loss ratios of the three strategies will gradually increase. There are two reasons for the loss of content: i) more cascading failures make it difficult for more content to be evacuated within its time window; ii) inter-DC has fewer available resources remaining for content evacuation. When the disaster area reaches 0.3, it is difficult for the three strategies to evacuate most of the content, and the content loss degrees of ESE and TPE reach more than 77% and 67%, respectively. TCE has the lowest content loss ratio, around 55%. In addition, as the initial disaster area increases, the resource utilization of the three strategies also gradually increases, indicating that the inter-DC optical network is evacuating as much content as possible during the occurrence of subsequent network cascading failures.

V. Conclusion

In this paper, for the first time, we considered the spatio-temporal problem of content evacuation caused by cascading failures in post-disaster scenarios. During the process of failure propagation and content evacuation, we introduced two trajectory models for content and available resources to formulate the spatio-temporal scheduling problem. Based on these two trajectories, we proposed a long-term decision-making evacuation strategy (TCE), which can selectively select other relay risky DCs to cache content that is too late to evacuate within the time window. TCE can reduce the content loss ratio by up to 25% compared to TPE and ESE when evacuating about 200 TB of content for each DC. In addition, TCE also showed good performance when evaluating for different number of cascading failures, different cascading failure hops, and different radius of the disaster zone.

References

[1] J. Chapagain, et al., "World Disasters Report 2022", International Federation of Red Cross and Red Crescent Societies (IFRC), Geneva, Switzerland, 2023. [Online] Available: https://www.preventionweb.net/publication/world-disasters-report-2022.

[2] M. F. Habib, M. Tornatore, and B. Mukherjee, "Cascading-Failure-Resilient Interconnection for Interdependent Power Grid - Optical Networks," in Proc. OFC, paper M3I.3 (2015).

[3] B. Mukherjee, M. F. Habib and F. Dikbiyik, "Network adaptability from disaster disruptions and cascading failures," in IEEE Communications Magazine, vol. 52, no. 5, pp. 230-238, May 2014.

[4] C. Saha, "Predicting Network Failures with AI Techniques," Electronic Thesis and Dissertation Repository, 9583 (2023).

[5] K. Feng, et al., "Tropical cyclone-blackout-heatwave compound hazard resilience in a changing climate," in Nat. Commun., 13(1): 4421 (2022).

[6] Y. Wang, et al., "Maximizing Optical Inter-DC Emergency Backup Reliability in Unpredictable Disasters," 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 2023.

[7] S. Ferdousi, et al., "Rapid data evacuation for large-scale disasters in optical cloud networks," in Journal of Optical Communications and Networking, vol. 7, no. 12, pp. B163-B172, Dec. 2015.

[8] L. Ma, W. Su, X. Li, B. Wu and X. Jiang, "Heterogeneous data backup against early warning disasters in geo-distributed data center networks," in Journal of Optical Communications and Networking, vol. 10, no. 4, pp. 376-385, April 2018.

[9] L. Ma, et al., "Joint Emergency Data and Service Evacuation in Cloud Data Centers Against Early Warning Disasters," in IEEE TNSM, vol. 19, no. 2, pp. 1306-1320, June 2022.

[10] M. F. Habib, et al., "Cascading-failure-resilient interconnection for interdependent power grid-Optical network," in OSN, 42, 100632 (2021).

[11] B. Schäfer, et al., "Dynamically induced cascading failures in power grids," in Nat. Commun., 9(1), 1975 (2018).