# 多層 CDN における CPA と DDoS の複合攻撃の評価と分析

# Evaluation and Analysis of Joint Attack with CPA and DDoS in Multi-Layer Content Delivery Networks

劉 甲奇　　上山 憲昭
Liu Jiaqi　　Noriaki Kamiyama
立命館大学 情報理工学部 情報理工学科
College of Information Science and Engineering, Ritsumeikan University

## 1. Introduction

In the last decades, more content providers are using Content Delivery Network (CDN) to support many user requests. To keep CDN running efficiently, it's necessary to be well protected against cyberattacks targeted to CDNs, i.e., Distributed Denial of Service (DDoS) attacks and Cache Pollution Attack (CPA) [1]. Although there are many existing works which investigate the impact of each of these two types of attacks, the impact of joint attack against CSes making these two types of attacks simultaneously has not been investigated. To effectively allocate resources to defend against these attacks, we should understand the attacker's strategy, and we have found the best attack strategy assuming just a single cache server (CS) [2]. However, many CDNs consist of multiple CSes with single or multiple layers. In this paper, we evaluate and analyze the impact of CPA and DDoS targeting on CSes of CDN with multiple layers.

## 2. DDoS Attack and CPA against CDN

In DDoS attacks, the attackers send many packets from bots to target CSes to increase the processing load of CSes and increase the response time of content delivery. On the other hand, in CPAs, the attackers send request packets for unpopular contents to target CSes to decrease the cache hit ratio of legitimate users. Moreover, in CPAs, the processing load of target CSes is also increased. Because cache miss increases the response time of content delivery, the aim of CPAs is also increasing the response time of content delivery. Therefore, the purpose of DDoS attack agrees with that of the CPA. Therefore, the interest of attackers is how to combine these two types of attacks and select the attacking targets, i.e., contents [1]. In this paper, for simplicity, we define DDoS attack as the cyberattack sending request packets to invalid contents that will be ignored by CSes and increasing the processing load of CSes. On the other hand, we define CPA as the cyberattack sending request packets to valid content to increase both the cache miss ratio and the processing load of CSes.

## 3. Analytical Model

The impact of both DDoS attack and CPA can be measured by the increase of the response time of CSes because the cache miss results in increasing the response time. We use the M/M/1 queue model and the approximation of cache hit ratio to derive the average response time of CDN CS.

Let $M$ denote the number of contents provided by CDN. Let λ$_i$ denote the Poisson arrival rate of request for content $i$, and we define $1/\mu$ as the mean of the exponentially distributed service time of a CS of CDN. Using the M/M/1 queue model, we can obtain the average service time, i.e., the average time a request spends on a CS, of content $i$ by

$$W = \frac{1}{\mu - \sum_{i=1}^{M} \lambda_i} \tag{1}$$

We use the Che-Approximation [3] to predict the hit ratio $h_i$ of each content $i$ on the CS. Let $C$ denote the capacity of the CS, and the maximum number of contents that can be stored in the CS is $C$. We assume that the request ratio of content $i$ is $q_i$, and from the Che-Approximation, we can obtain the cache hit ratio of content $i$, $h_i$ by

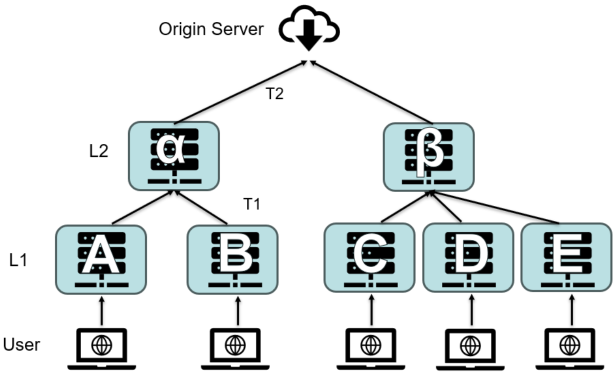

Fig. 1: Multilayer-CDN Model

$$h_i \approx 1 - e^{-q_i t_c} \tag{2}$$

Here, by solving $\sum_{i=1}^{M} h_i = C$, we can obtain $t_c$, the characteristic time of the CS.

The multilayer-CDN model is composed of multiple independent CSes with multiple layers. In this paper, we assume CSes of two layer, L1 and L2, as shown in Fig. 1. When a user requests a content, the CS accommodating the requesting user at L1 checks whether the requested content exists or not in its cache storage. If the requested content exists in the cache storage, the CSes of L1 sends the requested content to the user. Otherwise, the request is forwarded to the CS of L2 connecting to the CS of L1. If the requested content does not exist in the CS of L2, the origin server which stores all $M$ contents sends the requested content to the user. We assume that all the CSes update the cached contents based on the LRU (least recently used).

We further assume that there are two CSes, $\alpha$ and $\beta$, at L2, two CSes, $A$ and $B$, at layer 1 connecting CS $\alpha$, and three CSes, $C$, $D$, and $E$, at layer 1 connecting CS $\beta$. Let $T_1$ denote the latency between L1 CSes and L2 CSes, $T_2$ denote the latency between L2 CSes and the origin server, and $W_A$, $W_\alpha$, $W_o$ denote the average time a request spends at CS $A$, CS α, and the origin server, respectively.

For the request sent to CS $A$, when the requested content exists in the CS $A$ at L1, the average response time of the requested content, $r_A$, is $W_A$. When the request is forwarded to the CS $\alpha$ at L2, the average response time of the requested content is $r_\alpha$. When the the request is forwarded to the origin server, the average response time of the requested content is $r_o$. We can obtain $r_\alpha$ and $r_o$ by

$$\begin{aligned} r_\alpha &= W_A + W_\alpha + T_1, & (3)\\ r_o &= W_A + W_\alpha + Wo + T_1 + T_2. & (4) \end{aligned}$$

The average response time of content $i$ when request arrives at CS $A$, $R_A(i)$, is obtained by

$$R_A(i) = h_i^A r_A + (1 - h_i^A) h_i^\alpha r_\alpha + (1 - h_i^A)(1 - h_i) r_o, \quad (5)$$

where $h_i^A$ and $h_i^\alpha$ is the cache hit ratio of content $i$, $h_i$, at CS $A$ and $\alpha$, respectively. By summing $R_A(i)$ among all $i$ in $M$, we can obtain the average response time of request when accessing CS $A$ by

$$R_A = \frac{\sum_{i=1}^{M} R_{a_i}}{M}. \quad (6)$$

In the same way, we can also obtain the average response time when accessing each of other CSes.

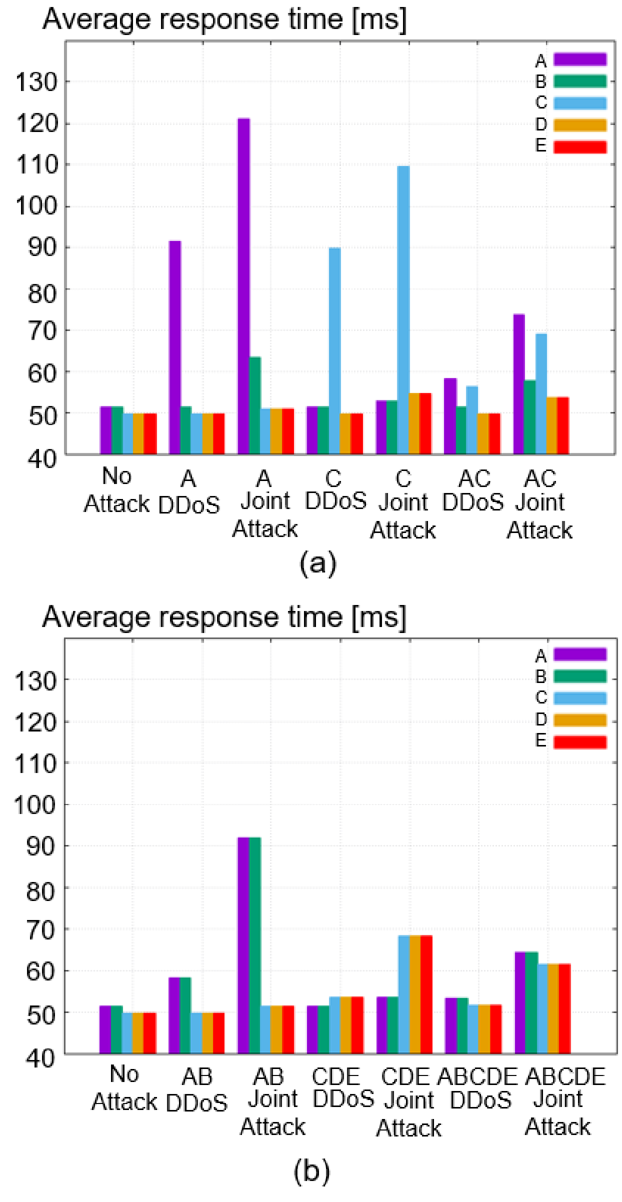

Fig. 2: Average response time of each CS under different attack patterns

## 4. Numerical Evaluation

Through computer simulation, we evaluate the average response time $R$ of each CS under difference attack strategies. We assume 5 contents, $M = 5$, which occupy the same amount of space in the cache and have different request rate, $\lambda_i = 80, 9, 6, 4, 1$ requests per second at all the five CSes of L1. The storage capacity of each CS is $C = 3$. To keep offered load 50% without attacks, we set the average service time as $\mu = 200$ ms for L1 CSes, $\mu = 200$ for CS $\alpha$, $\mu = 300$ for CS $\beta$, and $\mu = 300$ for the origin server. We also set the latency $T_1$ to 50 ms and $T_2$ is 30 ms [4].

We define two attack patterns: DDoS attack and joint attack. In DDoS attack, the attacker sends a large number of requests to increase the average response time of the server. In CPA, the attacker sends requests for unpopular contents to reduce the cache hit rate of popular content, which degrades the performance of CDN. The joint attack with DDoS and CPA will send a large number of requests for unpopular contents, which will spread the effect of the attack throughout the path. To distinguish between the DDoS attack and joint attack, let the DDoS attacker send request packets only to invalid contents, whereas in joint attack, the attacker sends the request packets to the least popular valid content, i.e., content 5.

Note that the total attacking capacity of the attacker is 80 requests per second in all the attack patterns. When the attacker send request packets to multiple CSes, it equally sends packets among the CSes. For example, in the AC DDoS attack, the attacker sends request packets to CS $A$ with 40 requests per second, and it sends request packets to CS $C$ with 40 requests per second. Figure 2 shows the average response time of each CS in each of attack patterns as well as the case without attacks. Compared with case of making DDoS attack only, the joint attack largely increased the response time of CSes with the same attacking capacity. Both the DDoS attack and the joint attack increased the average response time of the target CSes, whereas the joint attack also increased the response time of other L1 CSes connecting to the same L2 CS with the target L1 CS. However, when two or more CSes are attacked at the same time as shown in Fig. 2 (b), the DDoS attack has little effect because the resources are dispersed, but the joint attack still has significant advantages over the DDoS attack.

## 5. Conclusion

Through the multilayer-CDN mode, we used M/M/1 queue model and Che-Approximation to analyze the effectiveness of the joint attack and find the advantage of the joint attack over the common DDoS attack. In some cases, when DDoS attack resources cannot be centrally allocated to a single server, joint attacks will improve the effect of DDoS attacks which also shows the importance of the defense against CPA attacks.

## Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 21H03436 and 21H03437.

## References

[1] M. Ghaznavi, E. Jalalpour, M. A. Salahuddin, R. Boutaba, D. Migault and S. Preda, "Content Delivery Network Security: A Survey," in IEEE Communications Surveys & Tutorials, vol. 23, no. 4, pp. 2166-2190, Fourthquarter 2021
[2] 劉甲奇, 上山憲昭, "CDN のキャッシュを対象とした最適攻撃法 ", 信学会 2023 年ソ大会, B-6-55, 2023 年 9 月
[3] H. Che, et al., "Hierarchical Web Caching Systems: Modeling, Design and Experimental Results," IEEE J. Selected Areas of Commun., vol.20, no.7, Sep. 2002
[4] R. K. Thelagathoti, S. Mastorakis, A. Shah, H. Bedi and S. Shannigrahi, "Named Data Networking for Content Delivery Network Workflows," 2020 IEEE 9th International Conference on Cloud Networking (CloudNet), Piscataway, NJ, USA, 2020