Over-the-top video streaming, e.g., Netflix and YouTube, has been dominating the global IP traffic in recent years. The traffic will continue to increase due to the introduction of even higher resolution video formats like 4K on the horizon. As users consume video in massive amounts and in an increasing number of ways, ISPs need flexible solutions in place to ensure that they can deliver content quickly and easily regardless of their customer's location or device. More than 50% of over-the-top video traffic are now delivered through content distribution networks (CDNs). In addition to the distributed cache servers provided by the CDN, the edge router can also have a cache so that some videos could be stored in this cache and gets the advantage of the proximity to end-users.

In paper entitled "Joint Optimization of Caching and Routing Strategies in Content Delivery Networks: A Big Data Case", the authors considered a real CDN deployed by an Internet business in China, in which millions of people share their life moments by pictures and videos. They noted **some important issues** that are highly concerned by the CDN operators but are not well studied. The *first* is that end-users can usually access multiple CDN nodes with high quality of experience as, with the development in ISP architecture, the number of nodes used to serve users in the same place increases. The *second* is that the data service price may vary greatly for different regions, which can highly influence the cost of CDNs for large-scale applications. The *third* is the popularity of social contents which can highly deviate from the classic Zipf distribution as, users in regions show different interests and they may change over time. From their investigation, about 60 percent of files are requested once, and files requested more than 4 times can be regarded as popular contents.

The authors investigated the caching and routing strategies of CDNs, in which the contents popularity, prices, are jointly considered. In order to maximize quality of experience and decrease costs of operator, they *proposed* a joint cache management and routing algorithm by using alternating optimization. They **claimed** that the simulation results that their algorithm yield significant performance improvement in contrast with two content placement policies, most popular caching placement policy and random caching placement policy. As the proposed algorithm *outperforms* the current CDN strategy, in which most popular files are cached and end-users are directed to the nearest CDN nodes, by 30% and 12% in terms of latency and data service cost, respectively.

The authors **referred** to various other caching strategies in their Introduction section such as "A survey of information-centric networking," by B. Ahlgren, and others where they **discussed** about the content popularity being exploited to determine which content should be stored. However, in their *Alternating Optimization Algorithm* problem decomposition they mentioned about the work by J. C. Bezdek and R. J. Hathaway on "Some notes on alternating optimization" and compared their proposed alternating optimization policy (AOP) with other representative content placement policies. In contrary to below mentioned references there is **no significant comparison** seen with the previous works to their work for an instance as they did referred to papers where a joint cache management and routing problem are considered such as in "Efficient cache placement in multi-hop wireless networks" by P. Nuggehalli, and others and, in "Joint caching, routing, and channel assignment for collaborative small-cell cellular networks" by A. Khreishah and others.

Their **work is significant** as it would minimize the total latency and cost over the network while some practical constraints are taken into consideration including the storage capacity of caches in CDNs and the maximum number of users that each CDN allows. The *claimed contribution* is dependent on the simulation results as the proposed algorithm outperforms the current CDN strategy, in which most popular files are cached and end-users are directed to the nearest CDN nodes, by 30% and 12% in terms of latency and data service cost, respectively.

They **approached** the problem by comparing their proposed alternating optimization policy (AOP) with other representative content placement policies i.e. most popular caching placement policy (MCP) and random caching placement policy (RCP) where they applied an algorithm of alternating optimization to solve the two problems i.e. minimization of cost (OP2) and latency (OP3). Therefore, they <u>first</u> fix the content placement policy(X) and make all $X_{j,k} = 0$, i.e., all files are not stored in any cache. <u>Next</u>, they substitute X into latency to update routing policy P. In other words, due to the distribution of popularity, they set each cache stores the most popular files while cache still has spare capacity for MCP. For RCP, they set that each cache stores files randomly as long as its capacity is not exceeded.

The proposed solution is **nearly technically sound** as they carried out experimental evaluations and compared the results via simulations though it lacks the clarity as the paper does not contain much assistive tools such as Notation summary table and Controlled vocabularies i.e. standardized and organized arrangements of terms and phrases that provide a consistent way to describe data. The authors should have explained the curves more elaborately in Figures 3-6. It is not mentioned explicitly why the curves go so contrarily when the unit latency of congestion is 1ms and the number of files is 150. The references to other related papers do not assist the readers much as the author have given the summarised conclusions of their observations from the previous work where, they should have worked more on *explaining the experimental comparison*s. The calculations used in the ALTERNATING OPTIMIZATION ALGORITHM is vague albeit the authors did make a reference to the work by J. C. Bezdek and R. J. Hathaway on “Some notes on alternating optimization,”.

The experimental evaluations and result comparisons via simulations **strengthens** the paper. On contrary, Internet video traffic has been rapidly increasing and is further expected to increase with the emerging 5G applications. The paper does not talk about these new technologies as, where the content distribution network (CDN) should be efficiently managed to improve the system efficiency with upcoming technologies. Moreover, the authors do not specify what is the source and format of the files being considered. The **weakness** in the paper lies in the simulation results as those are based on uniform size files where in real world most video files need to be available in different encoding format, and this versioning enlarges caching requirements.

To **improve the technical aspect** of paper, the author should have considered the different formats of data in different conditions. The author should have compared other quality of service (QoS) metrices such as throughput, packet loss and reliability which is missing in the current submitted paper. At the same time, new services are emerging, such as Augmented and Virtual Reality with even tighter bandwidth and latency requirements than typical video streaming. The author should have tried experimenting on these too.

The **overall level of the paper** is good: even if it is quite simple, it is well written, and some important issues that are not well considered by the CDN operators are highlighted. The write-up is lucid, understandable and gripping. References are well mentioned. I have little concerns about the style of this paper. I think that *some attention* by authors should be devoted discussing the application scenario by explaining with realistic examples.