Comparing Cloud Content Delivery Networks for Adaptive Video Streaming

Chen Wang
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598
Email: Chen.Wang1@ibm.com

Andal Jayaseelan
Carnegie Mellon University
Pittsburgh, PA 15213
Email: ajayasee@andrew.cmu.edu

Hyong Kim
Carnegie Mellon University
Pittsburgh, PA 15213
Email: kim@ece.cmu.edu

***Abstract*—Cloud vendors offer content delivery network (CDN) services to compete for the video market. The user experience and the costs of providing the same video streaming service can vary when using different cloud CDNs. We emulate video streaming users in PlanetLab cloud to measure cloud CDNs including Amazon Web Service (AWS) CloudFront, Microsoft Azure Verizon CDN, and Google Cloud CDN. We leverage an approximated Quality of Experience (QoE) as a metric for evaluation. Our study finds that: 1) cloud vendors vary in providing QoE across regions; the video provider should assign a user to the CDN offering the best QoE at his location; 2) the QoE provided by one CDN can change over time; the video provider should adapt the CDN selection according to the real time QoE measurement; 3) cloud CDNs vary in scalability; streaming sessions may crash when there is bursty user demand; video providers should choose among the cloud CDNs that can properly scale; 4) regarding the cost, some cloud CDN is more economical than others given certain cache hit rate; video providers can minimize their costs by forcing free trial users to stream from the cheapest one [1].**

***Keywords*-Cloud, Content Delivery Network, DASH streaming, QoE, Adaptive CDN selection**

I. Introduction

The adoption of the cloud has advanced rapidly over the years [1]. The video providers are the early adopters and they use the cloud services extensively to meet their user demands. Cloud vendors also provide various types of services to compete for the video market. In past three years, popular cloud vendors started to provide Content Delivery Network (CDN) services to cache and distribute content for video providers. As cloud vendors vary in their infrastructure deployment and pricing schemes, the cost and the user experience of using cloud CDN services may vary for the same video streaming service. Thus, it is challenging for video providers to come up with an appropriate CDN selection strategy.

In this paper, we deploy emulated users around the world to study the performance and the cost of using cloud CDNs for video streaming. Our goal is to characterize popular cloud CDNs using a small set of emulated users to obtain insights on cloud CDN performance, so as to inform the CDN selection strategy for video providers. We let emulated users stream videos from popular cloud CDNs including AWS CloudFront, Azure Verizon CDN and Google Cloud CDN. We use the Quality of Experience (QoE) as the key metric to study the performance of cloud CDNs at different geographical locations, during different time periods and at different scales. We also compare the pricing schemes for above cloud vendors. We consider a subscription based model for video streaming services and numerically compare the costs of running the video streaming service in above cloud CDNs. Specifically, the emulated users are deployed on PlanetLab [2] around the world to run dynamic adaptive video streaming (DASH) [3]. A VoD website is cached in all cloud CDNs. The emulated users stream videos from each cloud CDN for 24 hours and their QoE are collected for comparison.

From QoE measurements on emulated users, we obtain the following observations.

Geographical difference: QoE from emulated users show that cloud vendors may have advantages over their competitors in certain regions. By comparing the QoE of the same emulated users on different cloud CDNs, we show that AWS CloudFront provides better QoE than Google Cloud CDN for tested users in North America, Asia, and Australia. Google Cloud CDN provides better QoE than Amazon CloudFront for users in South America and Europe. By studying the number of PoPs for both cloud CDNs, we find that Amazon CloudFront has more PoPs over Google Cloud CDN in North America, Asia and Australia, and vice versa.

Stability: in order to study the stability of cloud CDN performance, we let emulated users probe the CDN hosts periodically while streaming videos from 3 cloud CDNs respectively for 24 hours. To obtain a snapshot of the dynamics, we observe two users' latencies and QoE over 24 hours on 3 cloud CDNs. We find that the fluctuations of latencies over time are random and the fluctuation patterns vary across locations, are different among 3 cloud vendors, and change over time. When observing QoE for these two users, we find that the latency fluctuations provide no clues of the QoE drops. It indicates that the QoE drops are not related to the latency fluctuations. Besides, we do observe two users in two different time zones have QoE drops with the same cloud CDN around the same time. It shows that the cloud CDN they use has global performance issues at that

[1]This work was done while Chen Wang was a Ph.D. student at Carnegie Mellon University and was supported by the FCT under Grant SFRH/BD/51150/2010.

time. It suggests that the video provider should monitor the real time user QoE and adapt the CDN selection accordingly to guarantee good user experience.

Scalability: cloud CDNs can vary in scalability performance. When there are bursty user demand in one location, the scalability of resource provision at CDN edges mostly determine the user QoE. By increasing the number of concurrent video sessions at different locations, we show that 3 cloud vendors vary in the scalability. Some cloud CDN causes more session crashes than its competitors. Besides, one cloud vendor may vary in scalability across regions. In order to guarantee good user experience, video providers should evaluate the scalability of cloud CDNs at different locations in advance. Thus, they can predict the regional user demand and choose cloud CDNs that can properly scale.

Cost: for video service, the videos fetched from the origin server into the CDN incur the traffic of cache fill. The videos transferred out of the CDN incur the CDN egress traffic. Cloud vendors usually charge on both. From the pricing schemes available online, cloud vendors provide similar pricing schemes but the per-GB costs on these two types of traffic vary slightly. The costs also vary across regions. Usually the per GB cost decreases sub-linearly as the total volume of the traffic increases. The volume of the CDN cache fills and the egress traffic can be estimated given assumptions of the user demand and the cache hit rate. In order to compare the costs of using different cloud vendors, we propose a traffic model for adaptive video streaming service to estimate the volume of the CDN egress traffic given the scale of user demand. We also assume the cache hit rate for the video service to estimate the volume of the cache fills. We show that when the cache hit rate is high, one cloud vendor charges significant less than others as it charges less for egress traffic.

II. Related Work

A. Comparison of cloud vendors

CloudCmp [4] compared the major cloud vendors for Infrastructure as a Service (IaaS), including elastic computing, storage and networking. They run benchmark web applications on chosen set of similarly configured instances. They compare the computing service using the finishing time, the cost and the scaling latency. They studied the operation response time etc. to compare the cloud storage. They measured the intra-cloud and wide-area latencies to compare the cloud networking. W. Cai et al. [5] studied the popular cloud vendors for gaming applications. Though gaming applications involve video streaming, the Quality of Experience (QoE) for gaming are very different from the QoE of video streaming service. Besides, both works focus on the infrastructure services such as computing, storage and networking. We focus on cloud CDN service.

B. Comparison of CDN vendors

Prior works compared CDN vendors. Huang et al. [6] primarily characterized the performance of Akamai and Limelight. They compared two CDNs based on the number of cache servers, the internal DNS designs, the geographical locations of data centers, and the DNS/cache servers' response time. They used different mechanisms to calculate each of the parameters. One was to query the DNS server to determine the delays, the up-time, the availability and other characteristics of the edge servers of Akamai and Limelight. A related work in [7] used a tool called $Seattle$ to evaluate YouTube's CDN. They collected the number of IP addresses for YouTube host names and approximated the latencies from users to the actual video cache servers. They also collected network measurements such as the packet loss for evaluation. As cloud and CDN vendors upgraded their infrastructures and services frequently, we believe the performance comparison done in the past has little values for cloud CDN selection today. In addition, none of these works compared user QoE directly.

C. QoE Analysis of video streaming service

Quality of Experience(QoE) is an important metric for video streaming services. Existing studies analyzed the user QoE for various types of video streaming services. Pedro Casas et al. [8] studied the QoE relevant degradation for YouTube users and inferred that the root causes behind QoE degradation were linked to Google CDN's server selection strategies. [9] collected QoE measurement from 379 video service providers worldwide. They clustered QoE issues over the space of client/session attributes to find the attributes that are highly correlated with poor QoEs. Adnan et al. [10] analyzed QoE for a live streaming event in North America and find lower engagements for users with low QoE. Chen et al. [11] proposed a chunk based QoE model for Dynamic Adaptive Streaming over HTTP (DASH) streaming and used end user QoE to monitor the streaming service directly. [12] analyzed the QoE anomalies identified for a video streaming service deployed on Microsoft Azure. They found that 99% QoE anomalies were linked to transit networks.

D. Adaptive CDN selection strategies

Multi-CDN video delivery has been adopted by leading video providers including Netflix and Hulu. However, measurement studies [13][14] showed that the CDN adaptation strategies were naive solutions that made users switch CDNs when their streaming bit-rate dropped below a predetermined threshold. The existing strategies do not monitor user QoE directly. Junchen et al. [15] and Chen et al. [16] suggested to monitor end user QoE for adaptive CDN/server selection. Their solutions were based on the assumptions that end user QoE were mostly determined by static factors such as users' location or their Internet service providers (ISPs). From our study, we show that dynamic factors such as the bursty user demand and occasional unknown system-wide issues in CDN can impact user QoE.

III. Measurement System

A. Deployment of video streaming service

We run Dynamic Adaptive Streaming over HTTP (DASH) as the video streaming service for evaluation. DASH is currently the de facto video streaming technology in many commercial VoD systems, e.g. YouTube and Netflix. DASH videos can be hosted on HTTP servers, such as Apache servers. We deploy Apache2 servers in the computing instances in various Clouds to host a VoD website for DASH streaming. We use a 10-minute video clip to generate a video database. The video clip is encoded in 9 bit-rates. We concatenate multiple copies of the clip to generate long videos. We assume the whole video database is stored in an origin server. We set up the cloud CDNs to cache all videos and configure the CDN URLs for these videos.

B. Deployment of global testing users

We emulate 100 testing users on PlanetLab servers. Emulated users run DASH streaming services to evaluate different cloud CDNs. We implement an emulation of DASH player in Python [17] to request chunks from the CDN URLs. Upon receiving each chunk, our DASH client computes the user Quality of Experience for the chunk.

C. Deployment of bursty testing users

In order to test the scalability of resource provisioning at each edge location, we emulate 2 to 2^{10} testing users in each region of AWS. Emulated users run DASH streaming services to evaluate different cloud CDNs. Experiments are run at one scale per day to leave enough gap time between different runs. Thus, we make sure the resource at an edge is not provisioned due to previous runs.

D. QoE measurement

In DASH streaming, a video is encoded in multiple bit-rates and each bit-rate version is split into a series of fixed length segments, called chunks. DASH players detect the network throughput in real time and adaptively select the bit-rate for every chunk. The video bit-rate and the freezing time may change for every chunk. We use a chunk based QoE model proposed in [11] to approximate users' real time QoE for comparison of cloud CDNs. Quality of Experience (QoE) is a subjective measurement of user's satisfaction [18]. The chunk based QoE model in [11] approximates the QoE via the bit-rate of received video chunks and the possible buffering time induced by each chunk to evaluate the subjective user experience at run time, namely evaluations per chunk period.

E. Latency measurement

In order study the relationship between end user QoE and network latencies, we also deploy agents in emulated users to *ping* cloud CDN hosts periodically while streaming. The agent collects one latency measurement per 10 seconds.

IV. Cloud CDNs for Video Streaming

We study the CDN services from three Cloud providers respectively. Microsoft Azure collaborates with Verizon to offer CDN service. Google Cloud CDN uses Google's globally distributed edge PoPs to cache HTTP(S) load balanced content close to users. Amazon CloudFront also manages their own infrastructures. Due to the limited number of emulated users, our results cannot directly draw conclusions on the performance of leading cloud vendors. However, the methodology of QoE based monitoring and the insights obtained are applicable to production video streaming services.

A. Overall user QoE on three cloud CDNs

In order to understand how different cloud CDNs perform for video streaming applications, we run experiments on 100 PlanetLab servers around the world. All PlanetLab users start running the DASH streaming at the same time and each session lasts 1 hour. We use the QoE model in [11] to compute QoE for each chunk received. We measure

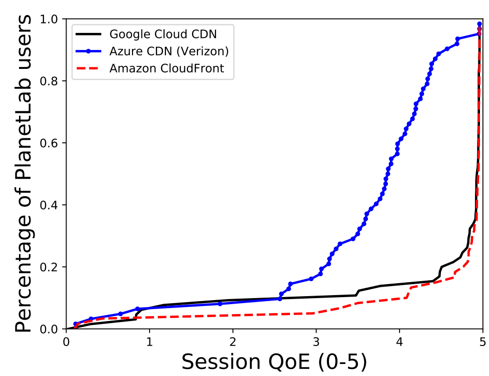


Figure 1: QoE Comparison for all three Cloud CDNs

the session QoE by averaging all chunk QoE monitored in a video session. In Figure 1, we plot the cumulative distribution of all session QoE for 100 users emulated around the world. Google Cloud CDN and Amazon CloudFront provide similar QoE for users. Amazon CloudFront offers a slightly better QoE for testing users. Overall, they offer better QoE than Azure Verizon CDN.

In order to understand why the cloud CDNs vary in providing QoE, we study the number of Point of Presence (PoP) locations of the above cloud CDNs. According to the information released online [19], there are 43 PoPs around the world for Azure Verizon CDN. Amazon CloudFront totally has 82 edge PoP locations and 11 regional edge cache locations in 16 geographical regions. Google has over 90 Internet exchanges and over 100 interconnection facilities around the world. Google Cloud CDN and Amazon CloudFront have more PoP locations around the world than Azure Verizon CDN. We find that the cloud CDNs with more PoPs have better geographical coverages, thus offering better overall user QoE. However, we also notice that the number of PoPs is not the only factor affecting user QoE. Google has more edge locations (>100 PoPs) than Amazon CloudFront (82 PoPs). However, for our testing users in one hour experiment, Amazon CloudFront provide slightly better overall QoE than Google Cloud CDN. To consider other factors affecting the performance of cloud CDNs, we study the stability and the scalability of cloud CDNs in the following sections.

B. Comparison of cloud CDNs across regions

From section IV-A, we observe that Amazon CloudFront overall provides the best QoE among three cloud vendors. However, it is not clear if Amazon CloudFront has advantages over others in different regions. We then group PlanetLab users in 5 regions to obtain the regional performance of cloud CDNs. These regions are North America, South America, Europe, Asia and Australia. To minimize the impact of peak hour traffic in performance, we let each emulated user stream videos from 3 cloud CDN around midnight in local timezone. We then compare the session QoE for all users in each region.

Figure 2 shows the average and the standard deviation of session QoE for users in each region. On average, Amazon CloudFront and Google Cloud CDN provide overall good QoEs (> 4) across regions. Amazon CloudFront has a slight advantage over others in North America, Asia and Australia. Google Cloud CDN provides better QoE than others for users in South America and Europe.

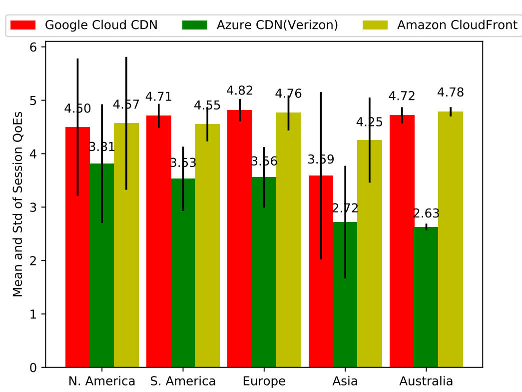


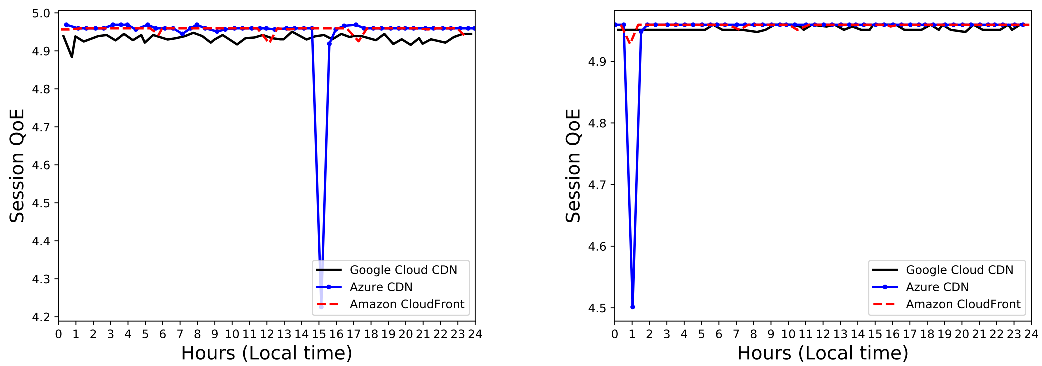
Figure 2: Comparison of QoE on Cloud CDNs across regions

C. Comparison of cloud CDNs over time

From section IV-B, we know cloud vendors have regional advantages over others in terms of QoE. We also wonder if such regional advantages persist over time. We then let 100 PlanetLab nodes run DASH video streaming on 3 cloud CDNs at the same time. We compare chunk QoE measurement for 3 cloud vendors over 24 hours. In order to study factors impacting user QoE, we also let PlanetLab nodes probe the CDN hosts every 1 minute to collect ICMP data.

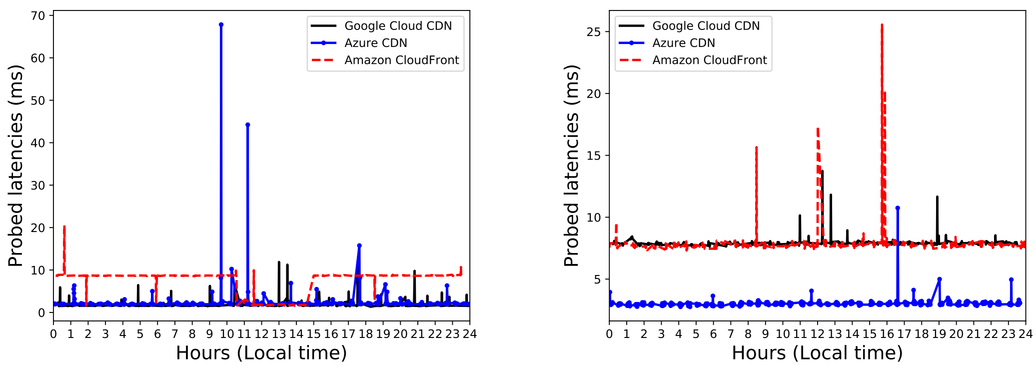
We first randomly pick up two users in Asia/Tokyo and America/New_York time zone to compare their QoE over 24 hours in Figure 3. Interestingly, both users experienced a short period of QoE drop on Microsoft Azure CDN. Considering the Tokyo time zone is 14 hours ahead of New York, these usersÓoE were dropping around the same time. We wonder if the QoE drops were related to networking issues, so we compare latencies probed from two users to the CDN edge servers in Figure 4. From Figure 3 and 4, we notice that the increases in network latencies do not correlate with the drops in QoE. It indicates that the QoE drop is not caused by networking issues and the QoE drops might be caused by some occasional system-wide performance issue on Azure CDN. As such issues usually impact more than one user, monitoring QoE for users around the world can help detect such performance issues on cloud CDNs.

Then, we study how users in the same time zone experience differently on 3 Cloud CDNs. In Figure 5, we plot the mean and the standard deviation of user QoE per hour on 3 cloud CDNs over 24 hours. The results show that Azure CDN has the lowest average and the highest deviation in QoE around 10:00 am. Similarly, Google CDN and Amazon CloudFront have the lowest average and the highest variance in QoE at 12:00 pm and 11:00 am respectively. We find that when a cloud CDN gives low QoE in a region on average , the QoE of users in the region is also highly varied and fluctuated. From above results, we see that there are dynamic factors affecting the QoE on all cloud CDNs. These factors were not captured by the network measurements in Figure 4. As routers usually handle ICMP traffic differently from TCP packets and there might be delays in processing TCP packets in transport layer,the dynamic factors can be the external Internet traffic congestion in peak hours or the CDNs' internal workload changes over time. As the workload on cloud CDNs both change over time and vary across time zones, we compare the QoE for users in different time zones. We aggregate measurements from 24 hours, as shown in Table I. We observe that there are time zones where users had low and unstable QoEs in all 3 Cloud CDNs (Asia/Singapore and America/Indiana). There are time zones where users had high and stable QoEs on all 3 Cloud CDNs (America/Toronto and America/Detroit). We also observe that Azure Verizon CDN provided the best and the most stable QoE for users in two time zones (America/New_York and America/Los_Angeles), although its overall and regional performance were not as good as other cloud CDNs shown in section IV-A. The results show that the variance in workload indeed affects user QoE. Thus, it is reasonable to infer that the variance of workload over time can be one of the dynamic factors affecting user QoE. To provide good



(a) User in Asia/Tokyo (b) User in America/New_York

Figure 3: User QoE on 3 cloud CDNs over 24 hours



(a) User in Asia/Tokyo (b) User in America/New_York

Figure 4: Latency between the user and 3 cloud CDN edge hosts over 24 hours

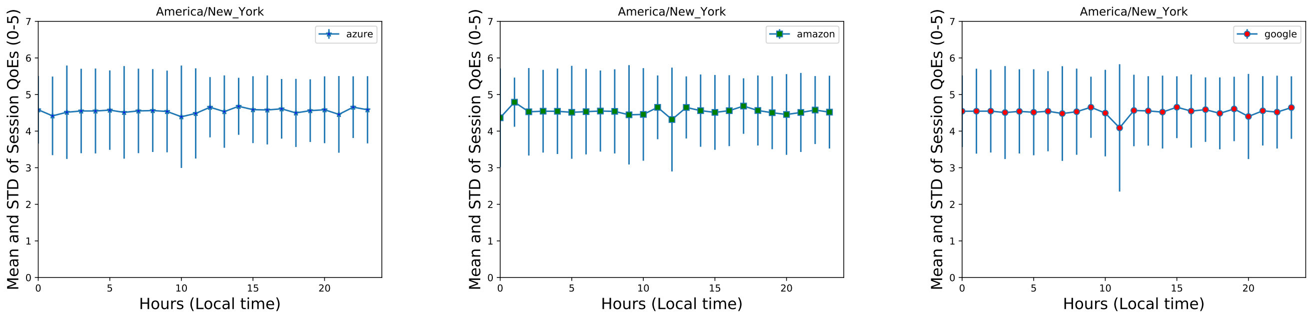
QoE for users, we believe it is necessary for VoD provider to monitor QoE over time, capture the impacts of those dynamic factors, and adapt the CDN selection accordingly.

D. Comparison of cloud CDNs under bursty user demand

Another important factor that may affect user QoE is how the cloud CDN scales the resources at edges. To study the scalability, we vary the number of users at one location and observe how users' QoE decreases as the number of co-located users increases. As PlanetLab servers have limited capacity and have only 1 to 2 servers at each location, we emulate a large number of users on Amazon EC2 $t2.small$ instances in each AWS zone. $t2.small$ instance has limited outbound bandwidth. However, from testing we find that each instance is able to support 64 clients streaming videos at the same time without session crashes (They stream in low bitrates, $1Mbps$) [2]. We vary the number of co-located users from 2 to 2^{10} to stress the CDN edge server at different scales. All users in the same AWS zone start streaming at the same time as if a burst of video requests arrive. All users stream videos for 1 hour on each of 3 cloud CDNs, one after the other. We measure user QoE at different scales to show if the cloud CDNs can properly scale up resources. To minimize the impact of caching at the edges, we run only one scale of co-located users per day to leave enough idle period between two runs on the same cloud CDN.

Figure 6 shows the number of session crashes over the number of co-located users on three cloud CDNs in three regions. There are no session crashes for all cloud CDNs when the number of co-located users is < 256, so the figure only shows the scale of 256, 512 and 1024. First, we observe that different cloud vendors vary much in the scalability. In North America, Google Cloud CDN has the highest session crashes at each scale. Azure Verizon CDN has the fewest session crashes at the scale of 256 and 512 while Amazon CloudFront has the fewest session crashes at the scale of 1024. It is worth to notice that the fewest crashes on Amazon CloudFront is not due to the deployment of testing users in AWS EC2. Amazon CloudFront also has the highest session crashes in Europe at the scale of 1024. Second, we notice that the scalability of the cloud

[2] The video session crashes if one chunk request times out three times. The chunk request timeout period is set as the default socket timeout period, which is 60 seconds.



(a) Azure Verizon CDN (b) Amazon CloudFront (c) Google Cloud CDN

Figure 5: Performance of Cloud CDNs over 24 hours perceived by all user in America/New_York

Time Zone	Azure CDN		Amazon CloudFront		Google Cloud CDN	
	Mean	Std	Mean	Std	Mean	Std
Australia/Melbourne	4.7497	0.4260	4.9301	0.0425	4.8643	0.02
Asia/Singapore	**1.6019**	N/A	**1.4347**	N/A	**1.6994**	N/A
Asia/Tokyo	**4.9343**	0.1282	4.9567	0.0097	4.9311	0.0104
Asia/Shanghai	4.4048	0.6879	4.4510	0.5761	4.5905	0.4114
Pacific/Auckland	4.6190	0.4749	4.8181	0.1495	4.6109	0.0291
Europe/Prague	4.9284	0.1872	4.9574	0.0045	4.9189	**0.0067**
Europe/Warsaw	4.4962	0.5672	4.6569	0.4055	4.6643	0.0306
North America/Chicago	4.3405	**1.6085**	4.3536	**1.6122**	4.4755	**1.4626**
North America/Detroit	4.9377	**0.1022**	**4.9590**	**0.0017**	**4.9539**	0.0351
North America/New York	4.5427	1.0665	4.5345	1.1052	4.5246	1.1043
North America/Los Angeles	4.8740	0.2195	4.8695	0.2492	4.8460	0.3012
North America/Indiana	**3.5985**	**1.3884**	**3.4246**	**1.5583**	**3.4474**	**1.5389**
North America/Vancouver	4.7484	0.2943	**4.9594**	N/A	**4.9548**	0.0038
North America/Toronto	**4.9429**	**0.0558**	4.9510	**0.0005**	4.9507	**0.0027**
South America/Belem	4.2118	0.2607	4.3752	0.1805	4.3512	0.2461
South America/Sao Paulo	4.8244	0.3687	4.9566	0.0137	4.9296	0.0071

Table I: Mean and STD of all session QoE in different time zones for 3 Cloud CDNs

CDN varies across regions. In North America, Google cloud CDN scales poorly as they have more session crashes when more users are co-located. However, in Asia, Google cloud CDN has no session crashes even 1024 users are co-located. Similar difference can be observed on Amazon CloudFront. At the time of testing, Amazon CloudFront scales properly in Asia but has a lot of session crashes in Europe regardless of the scale of co-located users. Third, we observe that for one cloud CDN, the scalability performance may not be consistent over time. When observing the session crashes in Europe for Google Cloud CDN and Amazon CloudFront, we notice that there are fewer session crashes at higher scales. As experiments on different scales were conducted on different days, with long idle period in between, we suspect that the scalability of resource provisioning also changed due to the dynamics in the system. An example of such dynamics can be the changing workload in a region. Overall, the results show that if the number of bursty co-located requests is less than 256, there would be no session crashes on all cloud CDNs. We also show how the QoE degrades as the scale of bursty co-located users increases in Figure 7. We calculate the average session QoE on 3 cloud CDNs across regions. Crashed sessions are counted as $QoE = 0$. In North America, the user QoE on all cloud CDNs decreases slightly as the scale of bursty users increases. In Europe, Azure CDN scales properly and the user QoE does not drop significantly over the increasing scales. In Asia, the QoE drops significantly at the scale of 1024 on all cloud CDNs, especially on Azure Verizon CDN. Many factors can impact the scalability, including the amount of available resources at the edges, the dynamic resource usage of edge servers, the caching strategies, etc. Our stress testing cannot identify those factors. However, as such information are not revealed for production cloud CDNs, we believe it is necessary for video providers to evaluate the scalability of these cloud CDNs via similar stress tests. The QoE based scalability evaluation is especially important for scenarios of popular live streaming events, where a large number of co-located users stream videos at the same time.

E. Cost

Cloud vendor adopt different pricing schemes to charge CDN services. It is necessary for video providers to understand the potential costs when choosing among these cloud vendors. As of September 2017, all cloud vendors charge on outbound data transfers, namely the CDN egress traffic. The per GB price drops as the volume of data transfers increases and the price varies across regions. Both Amazon CloudFront and Google Cloud CDN pose extra charges for HTTP requests. Besides, Azure CDNs and Google Cloud CDN charge on cache fills, which is the traffic fetched from the origin server or the storage when there are cache miss

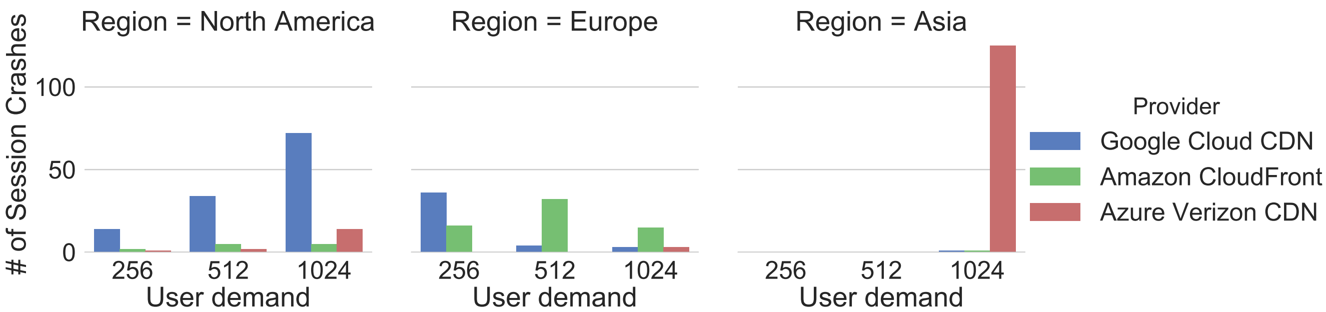


Figure 6: Number of sessions crashes vs the scale of bursty user demand

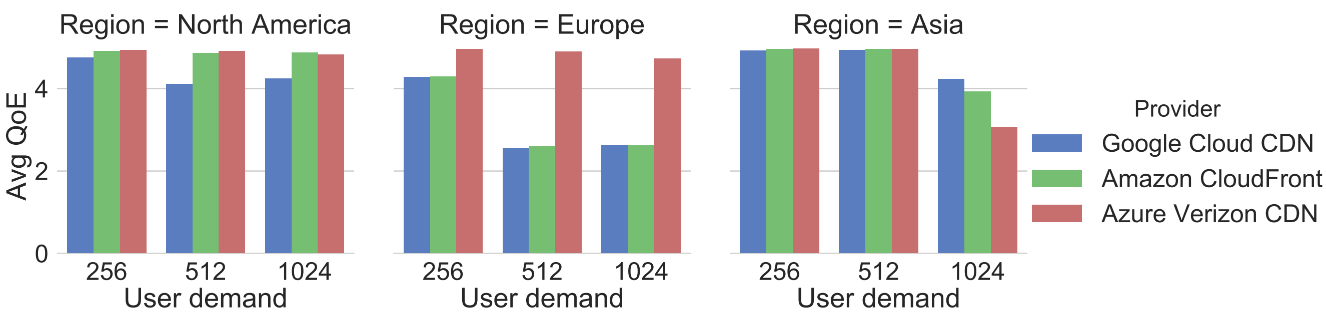


Figure 7: Average session QoE vs the scale of bursty user demand

at edges. Amazon CloudFront does not charge on cache fills as long as the origin server is deployed in AWS.

The total cost of the video streaming service depends on the usage of cloud CDNs, including the HTTP requests received, the amount of egress traffic and the amount of cache fills. Even if the total number of users and the streaming period per user are given, the CDN usage can still vary according to the user QoE and the popularity of videos. We propose a model to estimate the cloud CDN usage for a subscription based video streaming service. We assume the number of users in each region is given; how long each user watches videos and the bit-rate of videos streamed can be monitored for each user. Namely, the user u who resides in region i_u may watch h_u hours every month. Each video chunk lasts around 5 seconds. Therefore, the total number of video chunk requested is for user u per month is $J_u = h_u * 3600/5$. J_u also denotes the HTTP requests received in the CDN. The bit-rate of each video chunk is monitored as r_j. Thus, the total egress traffic user u generates can be calculated as $\sum_j^{J_u} r_j * 5$. The cache fills are determined by the popularity of videos and the caching algorithms of cloud CDNs, which are unknown to us. Thus, we use the cache hit rate $0 < \eta < 1$ to denote the percentage of traffic that are requested from edge servers.

In order to numerically estimate the cloud CDN usage for video streaming services at different scales, we then make the following assumptions on the distribution of i_u, h_u, and r_j. Other distributions can be applied similarly. We assume the video streaming service is offered in I regions. Each user u is randomly located in region $i_u \in I$ following a multinomial distribution with probability p_i^u at location i. There is, $\sum_{i \in I} p_i^u = 1$. Specifically, we choose I to include three regions including North America, Europe and Asia Pacific with $p_i^u = 1/3$ for all u. We assume that the watching hours h_u for user u follows a Poisson distribution with the mean watching hours $\lambda = 28$, according to a recent study on the average monthly watching time of Netflix users [20]. As DASH streaming allows the video bit-rate changing dynamically, the higher bit-rate one user is streaming the more egress traffic the user generates. Assuming the videos are pre-encoded in K bit-rate levels. Thus the total egress traffic one user generate is determined by the bit-rate of all video chunks requested. From the overall user QoE in Figure 1, the majority of users have QoE above 4 on all 3 cloud CDNs, which indicate users often streaming in high bit-rate levels. In our estimation, we simply assume that the probability of a chunk j requested in level k for all users follow the same geometric distribution with $p = 0.7$, namely $Pr(r_j = k) = (1-p)^{K-k} p^k$ where $k = 0, 1, \cdots, K$. In real system, the chunk bit-rate distribution can vary among users according to their available bandwidth. Other distributions can be used in the estimation if needed. In our comparison, we consider $K = 9$ bit-rate levels for all videos. The bit-rate ranges from 200kbps to 10Mbps.

We then randomly generate users at scales of $N = 10^3, 10^4, \cdots, 10^7$ for cost estimation. Each user streams for

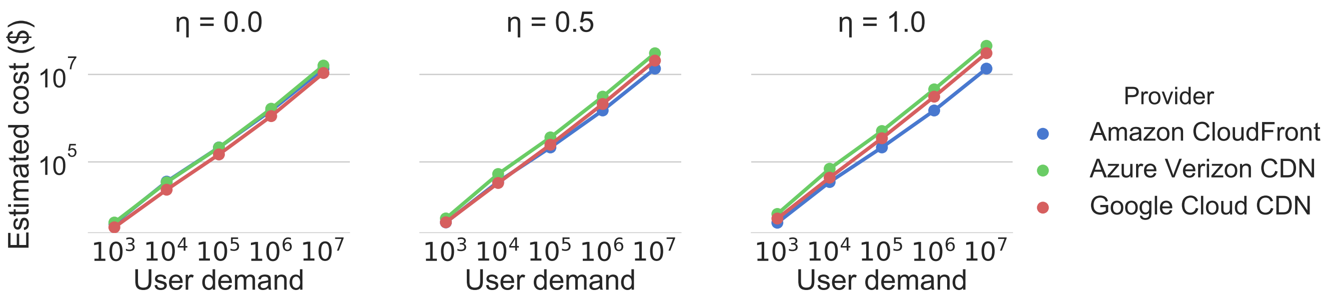


Figure 8: Numerical analysis of Video application costs at different user demand

h_u hours, requests $J_u = h_u * 3600/5$ chunks and requests r_j bit-rate for chunk j according to the assumed distributions of h_u and r_j. The total CDN usage per scale is averaged over 10 times of random experiments at each scale. We estimate the total cost of the video streaming service based on the pricing information available online [21][22][19]. In Figure 8, we compare the cost of video streaming on 3 cloud CDNs at different scales. We find that Google CDN is the most economical solution when the cache hit $\eta = 0$, denoting the scenario where all videos requested are not cached at edges and have to be fetched from the origin. When all videos are cached at edges, namely $\eta = 1$, Amazon CloudFront becomes the most economical one.

In the case that half of user requests are for cached videos and the other half are for non-cached videos, the cache hit rate would be $\eta = 0.5$. In this case, it is more cost-effective to use Google Cloud CDN at small N and more economical to use Amazon CloudFront at large N. For users with low expectation on QoE (e.g. free trial users), the video providers can predict the η and N to set the cheapest CDN as the default CDN.

V. Conclusion

We deploy testing users in PlanetLab cloud to evaluate popular cloud CDNs including Microsoft Azure CDN, Amazon CloudFront and Google Cloud CDN for video streaming services. We find that the user QoE offered by these cloud CDNs are different across regions, are changing over time and are impacted by the burstiness of user demand. The costs of using these cloud CDNs also depend on many factors including the popularity of videos requested, the total number of users, and how users experience over time. In order to guarantee the QoE for users with high expectation, we suggest that video providers monitor all users' QoE in real time on multiple cloud CDNs and enable adaptive CDN selection according to QoE measurements. Thus, one user can switch to other CDNs when his expectation in QoE is not met due to various issues including insufficient regional coverage, temporal performance drops and the poor regional scalability, etc. For users with low expectation on QoE, we suggest the video providers to predict the cache hit rate and set the most economical cloud CDN as the default to reduce the total costs.

References

[1] RightScale, "State of the cloud report," 2017. [Online]. Available: http://www.rightscale.com

[2] B. Chun *et al.*, "Planetlab: an overlay testbed for broad-coverage services," *ACM SIGCOMM*, 2003.

[3] T. Stockhammer, "Dynamic adaptive streaming over http–: standards and design principles," in *ACM MMSys*, 2011.

[4] A. Li *et al.*, "Comparing public-cloud providers," *IEEE Internet Computing*, 2011.

[5] W. Cai *et al.*, "A survey on cloud gaming: Future of computer games," *IEEE Access*, 2016.

[6] C. Huang, A. Wang, J. Li, and K. W. Ross, "Measuring and evaluation large-scale cdns," in *ACM SIGCOMM*, 2008.

[7] A. Rafetseder, F. Metzger, D. Stezenbach, and K. Tutschku, "Exploring youtube's content distribution network through distributed application-layer measurements: a first view," in *ACM International Workshop on Modeling, Analysis, and Control of CNET*, 2011.

[8] P. Casas, A. D'Alconzo, P. Fiadino, A. Bär, A. Finamore, and T. Zseby, "When youtube does not workanalysis of qoe-relevant degradation in google cdn traffic," *IEEE TNSM*, 2014.

[9] J. Jiang, V. Sekar, I. Stoica, and H. Zhang, "Shedding light on the structure of internet video quality problems in the wild," in *ACM CoNEXT*, 2013.

[10] A. Ahmed, Z. Shafiq, and A. Khakpour, "QoE analysis of a large-scale live video streaming event," in *ACM SIGMETRICS*, 2016.

[11] C. Wang, H. Kim, and R. Morla, "QWatch: Detecting and locating QoE anomaly for VoD in the cloud," in *IEEE CloudCom*, 2016.

[12] ——, "Identifying persistent and recurrent QoE anomalies for DASH streaming in the cloud," in *IEEE CloudCom*, 2017.

[13] V. K. Adhikari, Y. Guo, F. Hao, M. Varvello, V. Hilt, M. Steiner, and Z.-L. Zhang, "Unreeling netflix: Understanding and improving multi-cdn movie delivery," in *IEEE INFOCOM*, 2012.

[14] V. K. Adhikari, Y. Guo, F. Hao, V. Hilt, and Z.-L. Zhang, "A tale of three cdns: An active measurement study of hulu and its cdns," in *IEEE INFOCOM WKSHPS*, 2012.

[15] J. Jiang, S. Sun, V. Sekar, and H. Zhang, "Pytheas: Enabling data-driven quality of experience optimization using group-based exploration-exploitation." in *ACM NSDI*, 2017.

[16] C. Wang, H. Kim, and R. Morla, "Users know better: A QoE based adaptive control system for VoD in the cloud," in *IEEE GLOBECOM*, 2015.

[17] C. Wang, "Python emulation of DASH streaming," 2018. [Online]. Available: https://goo.gl/Lzzvoe

[18] ITU, "Recommendation p.10/g.100 (2006) amendment 1 (01/07)," 2007. [Online]. Available: https://goo.gl/DCKU3e

[19] Microsoft, "Azure Content Delivery Network," 2017. [Online]. Available: https://azure.microsoft.com/en-us/services/cdn/

[20] TechCrunch, "Netflix reaches 75% of US streaming service viewers, but YouTube is catching up," 2017. [Online]. Available: https://goo.gl/dq72Aa

[21] Google, "Google Cloud CDN Pricing," 2017. [Online]. Available: https://cloud.google.com/cdn/pricing

[22] A. CloudFront, "Amazon CloudFront Pricing," 2017. [Online]. Available: https://aws.amazon.com/cloudfront/pricing/