

Cloud Assisted Overlay Routing

Suat Mercan

American University of the Middle East, Kuwait
Telecommunications and Networking Technology (TNT)
Kuwait City, Kuwait
suat.mercan@aum.kw.edu

Abstract—Delay is a crucial factor on the Internet especially for real-time multimedia and interactive applications like video-conferencing, gaming. The best-effort service model doesn't provide essential assistance to these applications because of the unstable queuing delays. Cloud infrastructures have being used heavily for content delivery by replicating the original content prior to demand. In this work, we investigate if cloud can also help to improve delay-sensitive real-time interactive applications. In other words, is it possible to obtain shorter and more stable delays through cloud nodes. In order to answer this question, we have done a measurement study which are performed between some vantage points using PlanetLab and public cloud platforms. We investigate the experiment outcomes from different perspectives by analyzing various statistics and characteristics. We infer from the results that utilizing cloud nodes as proxy reduces path length while navigating between two points on the Internet.

Internet measurement; cloud infrastructure; real-time applications

I. INTRODUCTION

Latency is a crucial factor that affects the quality of real-time and interactive applications. The Internet is using the best-effort service model in which no guarantee is provided for packet delivery time. In the absence of such a guarantee of QoS (quality of service) in the Internet's original design, researchers developed some methods, such as QoS, CDN (Content Delivery Network), and overlay techniques to improve the perceived quality for users.

Cloud infrastructures; Google Cloud, Microsoft Azure, Amazon Cloud (EC2); have direct links to Internet Exchange Points (IXP) in many locations. They also have inter-datacenter links which are expected to have better connectivity than regular links. So, they provide lower latency for data packets on Internet. This can help to implement better multimedia applications for third party providers without having a complete infrastructure. Cloud infrastructure is already being used extensively for delivering content to end user in order to achieve higher throughput and lower start up time by copying the content to edge servers. Can cloud infrastructure be used to enhance user experience in interactive applications by reducing end-to-end delay. We have done some measurement studies using PlanetLab [1] and Cloud platforms to test proposed idea.

The idea is to improve quality of connections between users in terms of latency by utilizing inter-data center networks as in Figure 1. Connecting user1 and user2 through cloud nodes ($p2 + p3 + p4$) might be better than using direct connection

($p1$) because of the possible high quality connections between cloud nodes.

In order to test this proposal, we developed a agent that runs on PlanetLab nodes and sent periodic Ping messages to other nodes and cloud nodes. The purpose is to measure the RTT (Round-Trip Time) among the vantage points at determined intervals during a period of time.

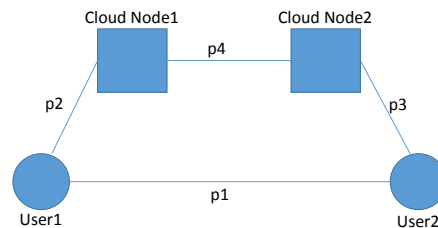


Fig. 1. One Proxy.

Results show that proposal is true for some pairs of nodes especially if the nodes are far from each other. Three hops route doesn't pay off for short distances. Even though the initial idea was utilizing the inter-datacenter networks, while investigating the results it was interesting to see that using only one proxy Figure 2 helps more than previous one. In this case, both user connects to the same cloud node and communication through one proxy ($p2 + p3$) becomes shorter than direct communication ($p1$) and also shorter than three hops most of the time ($p2 + p3 + p4$ in Fig1).

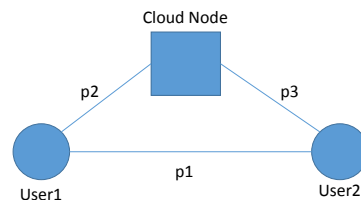


Fig. 2. Two proxy.

The remaining of the paper is organized as follows: We give brief information about related work in Section II. We explain experiment setup, present the experiment results and analyze findings in Section III. Then, in Section IV, we summarize our work and discuss possible future work.

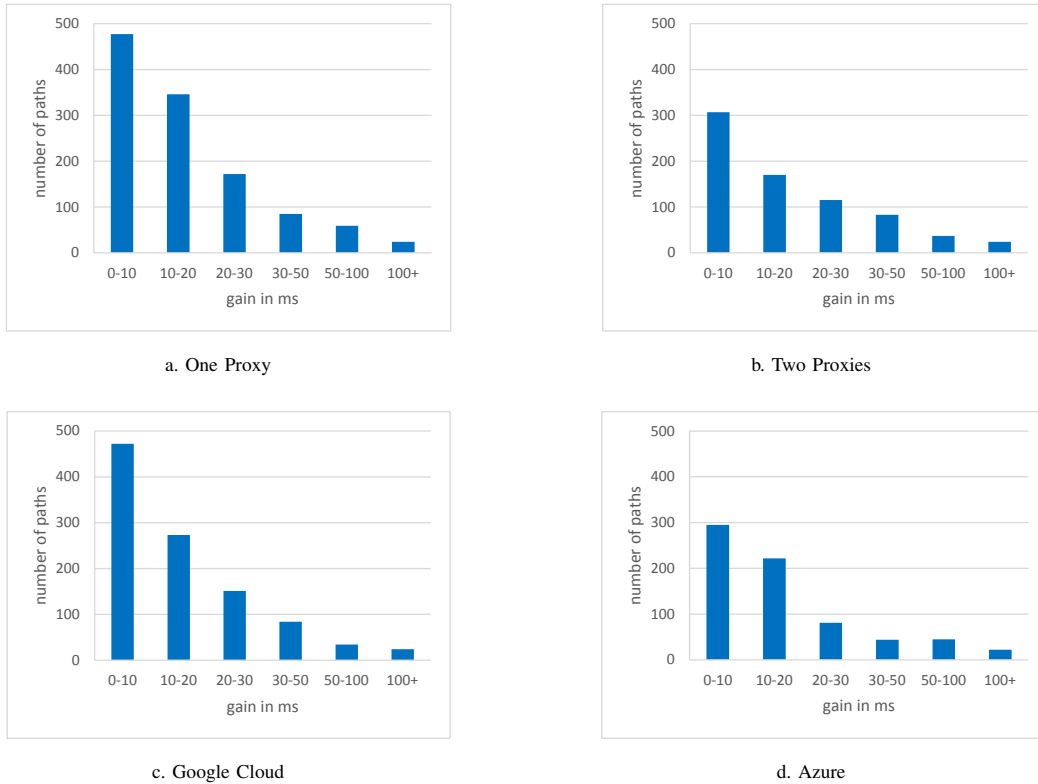


Fig. 3. Latency Gain

II. RELATED WORK

There exists some works in the literature which try to utilize cloud infrastructure to reduce the latency. A substantial amount of papers focus on cloud gaming. Since games require special and powerful hardware, researchers seek the possibility of playing games remotely by moving the software to cloud instead of installing on local machine. This requires a low latency between the user and cloud as response time to actions is important in games. Studies show that this might be achieved to some extent using the current cloud infrastructure. [2] presents a measurement study performed to investigate the latency between EC2 nodes and end-users. They selected 2500 nodes from Torrent users in the US which is believed to represent user distribution. In the time that the study was done, EC2 has only 3 datacenters in US. They claim that 70% of the users have less 80 ms latency to one of the EC2 datacenters which is accepted feasible for cloud gaming. Authors in [3] perform measurement study to understand the benefit of using three cloud infrastructure instead of using only one of them for web services. They send queries to IP prefixes from Planet-Lab nodes and cloud nodes to estimate the latency from cloud to end user in both cases. They found that multiple cloud utilization is improving the latency by 20% while management is getting more difficult. Airlift [4] aims utilizing inter-

datacenter networks to maximize the total throughput among peers. The focus is increasing bandwidth of connections rather than decreasing the delay. They implemented the video conferencing agents and run on Amazon EC2. Then, they presented the results by comparing to p2p design. [5] evaluates the performance of inter-datacenter networks for EC2 and Azure. They pick a region in each continent, then apply some traffic among these points to investigate the performance. They present throughput and delay results. [6], [7] and [8] can also be considered as related to the topic.

In our work, we concentrate on reducing the delay among users. We measure the latency between two arbitrary points using p2p and cloud overlay, then compare the results to see the contribution of cloud nodes as overlay proxy points.

III. EXPERIMENT RESULTS

In this section, we first give experiment details, then present and analyze measurement results from different perspectives.

A. Experiment Details

In order to measure RTT values among nodes, we developed a small program and ran on different PlanetLab and cloud nodes. This program pings other nodes periodically and saves the result. The idea is simple; we measure the latency among PlanetLab nodes directly. Then we measure the distances

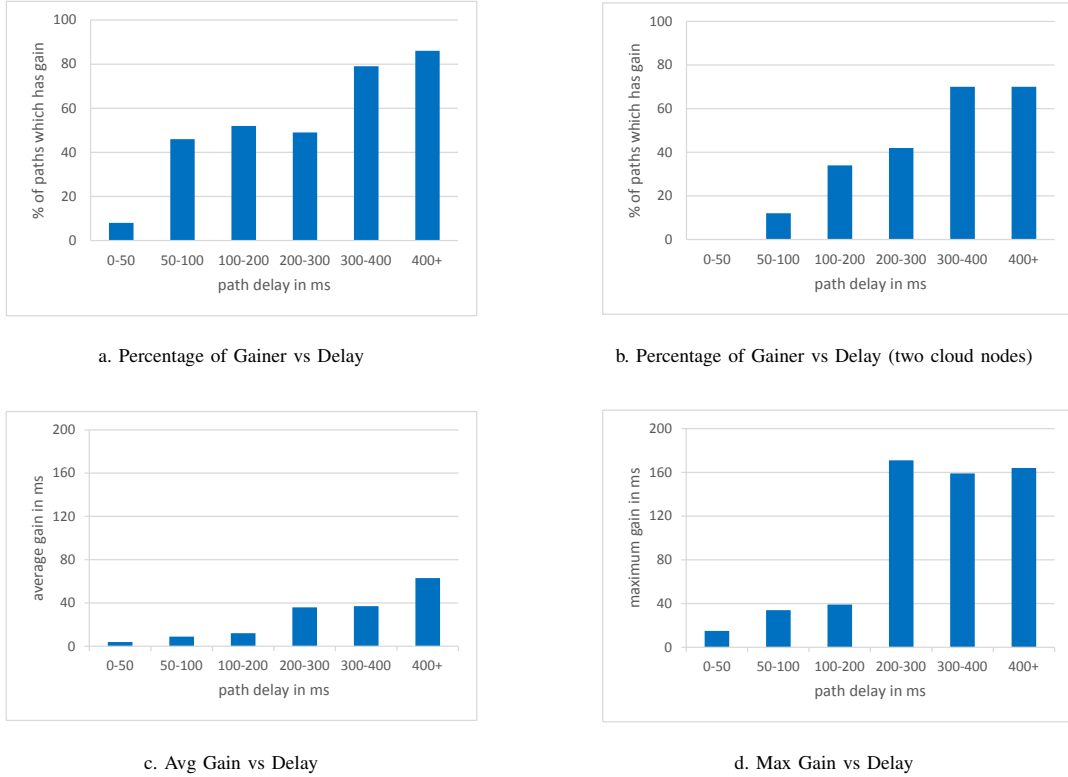


Fig. 4. More Statistics about Latency Gain

through cloud nodes to see if a shorter path can be obtained. We used 56 PlanetLab nodes, 6 Google Cloud nodes and 5 Microsoft Azure nodes. 56 PlanetLab nodes are used so that 3156 (1578 distinct paths) paths are involved in the analyses. Some paths (especially some nodes) are having very high variation. We excluded these paths to have more accurate results. So final results are based on 2672 paths. We used mode value as representative for each path. Since we used trial version of cloud accounts, we could be able to utilize limited number of sites. When all sites are benefited, better results will be certainly acquired. Histograms are utilized intensively to analyze and explain the results.

B. Gain in Delay

We try to get a general idea about how much the latency of packets on the Internet can be reduced by using cloud nodes as proxy for overlay routing. We have measured two cases; using only one proxy node (two hops in total) or using two proxy nodes (three hops). The gain of the first case in which nodes are communicating via one cloud node is shown in Figure 3.a. 45% of paths have a shorter route through one of 11 cloud nodes. Some of them have small gain which is under 10 ms while some other have more than 100 ms. The Figure 3.b shows the values of gain that are acquired when two proxies are used between two nodes. Obviously and interestingly, there

are many cases that navigating through 3 hops instead of direct access to reach destination is better. The paths in both cases (one proxy, two proxies) mostly matching which means that if there is a 3 hop path which is better than the original, then there is a best path through only one proxy. There are only rare cases that 3 hops is better than 2 hops. In next section, we also show the percentage of the gain compared to actual latency which might give additional information.

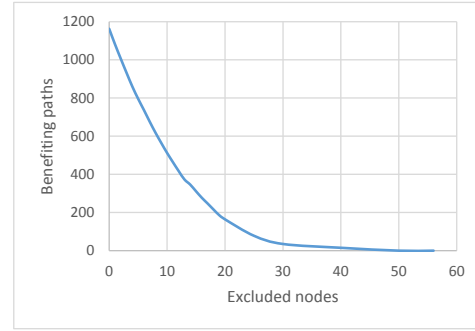
As stated before, we use some nodes from Google Cloud and some from Azure, 6 and 5 respectively. In order to see the benefit in case only one group is used, we reevaluate the paths and present the results in Figure 3.c and d in which less gain is achieved than combined one as expected for both cases. Also, it can be said that Google Cloud is yielding better results even though the purpose of this study is not comparing these two platforms.

C. More on Delay Gain

To understand the gain more, we investigate additional features in the collected data from different perspectives. Each figure take picture from another side. One insight we expect to exist is that longer paths should benefit more from rerouting and that is something desired. In Figure 4.a, we try to figure out the trend in gain with increasing delay between nodes. Since a new hop is added in case of rerouting, it is a low

Distance	Together	Google	Azure
0-10	5	3	5
10-20	16	18	8
20-30	17	16	13
30-50	7	8	18
50-100	4	5	6
100+	7	7	7
average	36	38	45

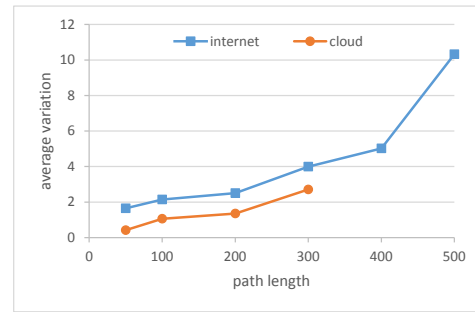
a. Distance to closest cloud node



b. Number of Benefiting Paths

Node	Dist. to Closest Node	Avg Dist. to Other	One Proxy	Two Proxy
12	215	409	12	7
9	10	173	33	33
33	1	187	34	34
23	52	188	40	30
25	116	198	39	27
19	25	146	37	25
27	48	124	0	0

c. Sample Node Records



d. Latency Variation of Paths

Fig. 5. Various Findings

probability that the delay is improved in lower delays but still there are some cases that falls in this category. When the figure is investigated, we see that for 80% of the paths whose latency is over 300, an improvement is achieved. When we look at shorter paths, number of benefiting paths (called gainer in the figure) is getting less. If the path length is less than 50, then rerouting doesn't contribute at all, or very little for only a small portion of the paths. The figure confirms the expectation (longer paths benefit more) we had at the beginning.

In case that two cloud nodes are used as relay points, Figure 4.b, improvement is reduced especially for low latencies. Higher latency paths have still improvement. 70% of paths whose latency is over 300 have a better (much or little) alternative path which consists of three parts. This is making the idea of using cloud nodes as proxy or relay nodes for overlay routing more promising.

In Figure 4.a and b, we looked at percentage of benefiting paths vs path length without considering the amount of benefit. Next two figures will evaluate the benefit by taking the amount into consideration which help to grasp the number we are talking about. Figure 4.c shows the average value of benefit vs path length. It is obvious that benefit is increasing with path length. Some nodes may not be enough close to a cloud relay node, that is why their benefit is very limited which drags the average gain down. So, looking at max values, Figure 4.d,

will reveal some information about the possible benefit. There are cases that we have 150 ms shorter paths which will really affect the quality of application.

As stated before, we have used only limited number of planet-lab nodes (so, limited vantage points) and limited cloud nodes in terms of number and geographic distribution (so, limited relay points). For this reason, the figures and statistics discussed until now reflect only a small portion of the big picture. But, the existence of nodes taking advantage of cloud overlay shows that the idea is applicable in a wider scope. When the whole internet and a broader network of cloud nodes are considered, the results will be different (probably better). Since we don't have the big picture, by looking at exemplary cases we can understand the conditions that boost the gain. Below is a list of items which states some findings from the dataset.

- Approaching to cloud nodes will help the content delivery services as well as the cloud overlay concept. The table in Figure 5.a shows the count of planetlab nodes in each interval based on the distance to the closest cloud node. There are some nodes which are just 10 ms away from the cloud. 38 nodes out of 56 (72%) has a cloud node in a distance shorter than 30ms, and 90% of them is closer than 100 ms which shows cloud nodes have good coverage. These values might be sufficient for cloud

based services like gaming, but for cloud assisted overlay routing, it might depend on the path length. If only one cloud platform was to be used, distance to the closest node would be more. The last row in the table shows the average values of all nodes.

- The paths which has gain is not equally distributed among nodes which means most of the gains are associated with some specific nodes. When we take these nodes out from the node set, number of benefiting paths from overlay routing is dramatically decreasing as seen in Figure 5.b.
- When the nodes are investigated in more detail, we can highlight some characteristics of various nodes. We should say that we could not derive any certain rule that brings higher number of benefiting paths or vice versa. In Figure 5.c, there are some sample node records. Column1 is node number, column2 is the distance to closest node, column3 is average length of benefiting paths from this node to other nodes, column4 and column5 are the number benefiting paths with one and two proxies. Node12 is very far from closest node, but there are still shorter paths through proxy since it is also too far to other nodes. Node9 and Node33 are very close to cloud node as the distance to others is average, they have shorter paths with many others either using one proxy or two proxy. Node23 and Node25 are can be far to cloud node, but there is a high number of benefiting nodes with one proxy which drops in case of two proxies. Node27 has no alternative shorter path. In general, we can state two things; (1) if the path between two nodes is long enough, then there is probably a better way through a cloud node even it is not very close, (2) if one node is very close to a cloud node, then there might be better paths. But, it might depend on individual circumstances.

D. Variation in the Paths

We have looked at gain up to this point in terms of latency which is an important metric for real time applications, there is another criteria, variation or jitter, which is also important. We try to analyze that if cloud overlay is also helping to reduce the jitter. We preferred using *mad*(Mean Absolute Deviation) to express dispersion of measured values. The definition of *mad* is given.

$$mad = \frac{\sum_{i=1}^n (x_i - \mu)}{n}$$

Basically it measures average distance from all points to mean. *mad* is calculated for each path among planetlab nodes and cloud nodes. Figure 5.d shows the latency variation of paths vs path length. As the path length increases, the variation gets higher. When the cloud and Internet variation is compared, cloud gives lower variation.

IV. CONCLUSIONS

In this paper, we performed some measurements using PlanetLab platform and public cloud platforms in order to understand the if the cloud nodes can be utilized to reduce the

delay among users which would help to improve the quality of real time applications. Results showed that cloud assisted overlay routing idea is shortening end-to-end delay to some extent. We are planning to take this study to further level in which IP prefixes will be used from all ASes and more cloud nodes will be used to acquire more accurate results.

REFERENCES

- [1] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, M. Bowman. "Planetlab: an overlay testbed for broad-coverage services." ACM SIGCOMM Computer Communication Review, 33(3), 2003, 3-12
- [2] C. Sharon, B. Wong, G. Simon, and C. Rosenberg. "The brewing storm in cloud gaming: A measurement study on cloud to end-user latency." In Proceedings of the 11th annual workshop on network and systems support for games, 2012
- [3] W. Zhe and Harsha V. Madhyastha. "Understanding the latency benefits of multi-cloud webservice deployments." ACM SIGCOMM Computer Communication Review 43, no. 2 (2013): 13-20
- [4] F. Yuan, Baochun Li, and Bo Li. "Airlift: Video conferencing as a cloud service using inter-datacenter networks." In Network Protocols (ICNP), 2012 20th IEEE International Conference on, pp. 1-11. IEEE, 2012.
- [5] C. Yingying, S. Jain, V. K. Adhikari, Z.-L. Zhang, and K. Xu. "A first look at inter-data center traffic characteristics via yahoo! datasets." In INFOCOM, 2011 Proceedings IEEE, pp. 1620-1628. IEEE, 2011.
- [6] Yang, Song, Philipp Wieder, Ramin Yahyapour, Stojan Trajanovski, and Xiaoming Fu. "Reliable Virtual Machine Placement and Routing in Clouds." IEEE Transactions on Parallel and Distributed Systems 28, no. 10 (2017): 2965-2978.
- [7] Makkes, Marc X., Ana-Maria Oprea, Rudolf Strijkers, Cees de Laat, and Robert Meijer. "MeTRO: low latency network paths with routers-on-demand." In European Conference on Parallel Processing, pp. 333-342. Springer, Berlin, Heidelberg, 2013.
- [8] Cai, Chris X., Franck Le, Xin Sun, Geoffrey G. Xie, Hani Jamjoom, and Roy H. Campbell. "CRONets: Cloud-Routed Overlay Networks." In Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on, pp. 67-77. IEEE, 2016.