Regular Paper

Recent Empirical Evaluation of Aggregate Throughput using Parallel Transfers on the Internet

CHUNGHAN LEE,†1 HIROTAKE ABE,†2 TOSHIO HIROTSU†3
and KYOJI UMEMURA†4

It is known that aggregating parallel TCP flows have beneficial effect on sending large-size data over the Internet. Hacker et al. performed the empirical study on the effect in 2002, and Altman et al. proposed and verified its theoretical modeling in 2006. The key result of the studies was that aggregating 10 or more flows is meaningless because performance gain by adding another flow will saturate at 10 TCP flows. We re-performed the throughput measurements as Hacker et al did, but on the recent Internet environment. In this paper, we present the result of our measurements and discuss the difference we observed between Hacker's and ours.

1. Introduction

Aggregate throughput using parallel TCP flows is one of key concepts to improve data transfer. This concept is widely used by applications such as PSocket[1)], bbFTP[2)], and GridFTP[3)]. In particular, GridFTP used parallel TCP flows to provide a more reliable and high performance file transfer for Grid computing applications. But, we cannot expect the improved aggregate throughput when we use the incorrect number of flows, especially when it is excessive. Because it occurs competition among TCP flows, total throughput of aggregated flows would be decreased. Thus, it is important to understand the characteristics of aggregate throughput for the improvement on the Internet.

In order to understand the characteristics of aggregate throughput using parallel TCP flows, Hacker et al.[4)] showed the effect of parallel TCP flows on a lossy wide-area network. Moreover, Altman et al.[5)] expanded Hacker et al.'s study with a formula for the aggregate throughput. In the above works, the aggregate throughput was gradually saturated after approximately 10 TCP flows. However, the current situation in 2012 would be different from the previous works because the scale and bandwidth of networks are rapidly increasing on the recent Internet environment. Thus, it is considered important to review the effect of the aggregation in the current Internet environment.

†1 Department of Electronic and Information Engineering, Toyohashi University of Technology
†2 Cybermedia center, Osaka University
†3 Faculty of Computer and Information Sciences, Hosei University
†4 Department of Computer Science and Engineering, Toyohashi University of Technology

In this paper, we focus on the re-production of throughput measurement using parallel TCP flows as Hacker et al did, but on the recent Internet environment. We first use PlanetLab[6)7)], which is a virtualized network testbed on the Internet. Second, we select node pairs on different sites across North America and Europe. Next, we simultaneously generate parallel 10, 20, 30, 40, and 50 TCP flows on the node pairs to measure the aggregate throughput. Finally, the measurement results are analyzed with statistical analysis. In the measure results, partial results accord with the previous results by Hacker et al.[4)], thus the aggregate throughput is gradually saturated and actual throughput per TCP flow suffers bandwidth-degradation after 10 TCP flows. However, different characteristics of aggregate throughput are founded. The aggregate throughput is increased after 10 TCP flows. And also actual throughput per TCP flow is not decreased up to 30 TCP flows.

The rest of this paper is organized as follows. Related work is described in Section 2. The measurement methodology is explained in Section 3. The measurement results are shown and discussed in Sections 4. The paper concludes with a summary of the main points in Section 5.

2. Related work

There are some studies to understand the characteristics of aggregate throughput and to find the optimal number of flows for the improved aggregate throughput. Hacker et al.[4)] showed the total number of flows is like one huge flow that combines total of each flow's actual throughput and the aggregate throughput for Maximum Segment Size (MSS) 1448 is gradually saturated after 10 TCP flows. However, there is no throughput formula for parallel TCP flows. Next, Altman et al.[5] showed the aggregate throughput always had the similar characteristics depending on the capacity of the TCP flows, and it is also saturated after 10 TCP flows. Moreover, they proposed a formula for the aggregate throughput. But, it is not easy to estimate the capacity and bottleneck link on end nodes precisely. Next, Yin et al.[8] proposed a throughput prediction service for many-task computing. It provides users with the optimal number of TCP flows and an estimated time for data transfer. However, their evaluation results focused on Louisiana Optical Network Initiative (LONI) clusters, and they did not investigate the characteristics of experimental networks.

Unfairness among TCP flows has been researched. Nakamura et al.[9] showed a problem of actual throughput on Long Fat Pipe Network (LFN)[9]. Fast Ethernet is sometimes much faster than that of Gigibit Ethernet due to bursty transmission of Gigibit Ethernet. It implies that the aggregate throughput would be unstable when we use parallel TCP flows on LFN. Next, Kamezawa et al.[10][11] showed parallel TCP flows divide into a group of fast flows and a group of slow flows on LFN. Thus, the slow flows may be major causes of decreased aggregate throughput. These results are enough to show that there is unfairness among TCP flows when we use parallel TCP flows.

3. Measurement methodology

3.1 PlanetLab

PlanetLab[6][7] is a global testbed for network and distributed systems research. As of July 2012, it has grown to nearly 1100 machines spanning more than 500 sites and 40 countries. Linux-Vserver is used to virtualize resources on a node. Thus, node resources, such as a CPU, memory, disk space, and network connections, are shared among research activities on the node. When transmitted network traffic is more than the boundary volume, a shaper reduces the network traffic by force. In the testbed, a platform called a *sliver* is provided as a virtualized environment to users, and multiple slivers can be run simultaneously at each node. A set of these slivers participating in the same activity at different nodes is called a *slice*. Thus, PlanetLab consists of virtualized nodes on the Internet.

3.2 Node selection

To gather the aggregate throughput on the Internet, we empirically selected four pairs of nodes from PlanetLab nodes located in both North America and Europe, which we refer to as nodes (α, β), (α, γ), (δ, ϵ), and (δ, ζ). We try to choose receiver nodes that have similar Round Trip Time (RTT). We used the same sender node for the measurements to observe fluctuations in the aggregate throughput. The geographic location and mean RTT using ping for all the pairs are shown in Table 1.

Table 1 Geographic location and mean RTT at node pairs.

Geographic location	Sender node	Receiver node	Node pair	Mean RTT [s]
North America	α	β	α, β	0.049
		γ	α, γ	0.050
Europe	δ	ϵ	δ, ϵ	0.043
		ζ	δ, ζ	0.038

3.3 Throughput measurements

Here, we describe a measurement method for the aggregate throughput. Each set of parallel TCP flows consists of 10, 20, 30, 40, and 50 TCP flows. The number of TCP flows is determined based on the previous works[4][5]. To re-perform the throughput measurements as Hacker et al. did, each set of parallel TCP flows is simultaneously generated at the sender node every 5 minutes. The traffic was reduced by the shaper after approximately 20 trials at 50 TCP flows when we use 4 MB or later. Data size at each TCP flow is determined as 2 MB. This size is enough to reflect congestion window up to congestion avoidance phase. For each node pair, we conduct 100 repeated experiments. The measurement method is depicted in Fig. 1.

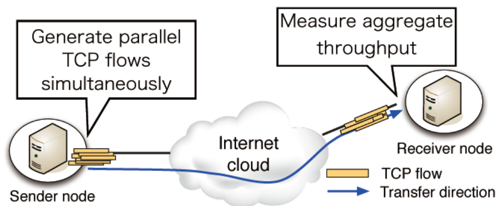
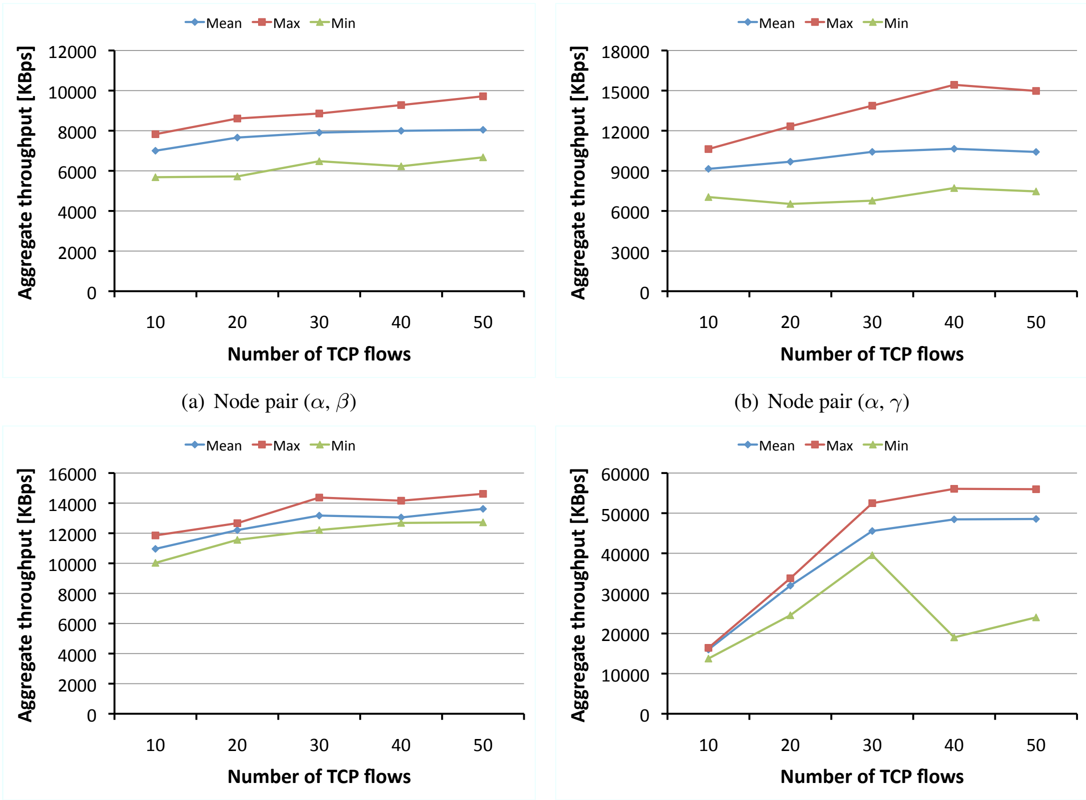


Fig. 1 Aggregate throughput measurement method.

4. Measurement results

4.1 Aggregate throughput

Although we tried 100 experiments, a few of experiments are failed. In the failed experiments, the number of TCP flows is lack or the data was not transmitted up to 2 MB. We get rid of the failed results. The statistics of aggregate throughput at node pairs using parallel TCP flows are described in Fig. 2. The mean aggregate throughput at node pair (α, β) is different from that at node pair (α, γ) though the mean RTT using ping is similar. And also the maximum value is low in comparison with that at (α, γ). We find that the aggregate throughput at these pairs is gradually saturated after 10 TCP



(a) Node pair (α, β)

(b) Node pair (α, γ)

(c) Node pair (δ, ϵ)

(d) Node pair (δ, ζ)

Fig. 2 Statistics of aggregate throughput at node pairs

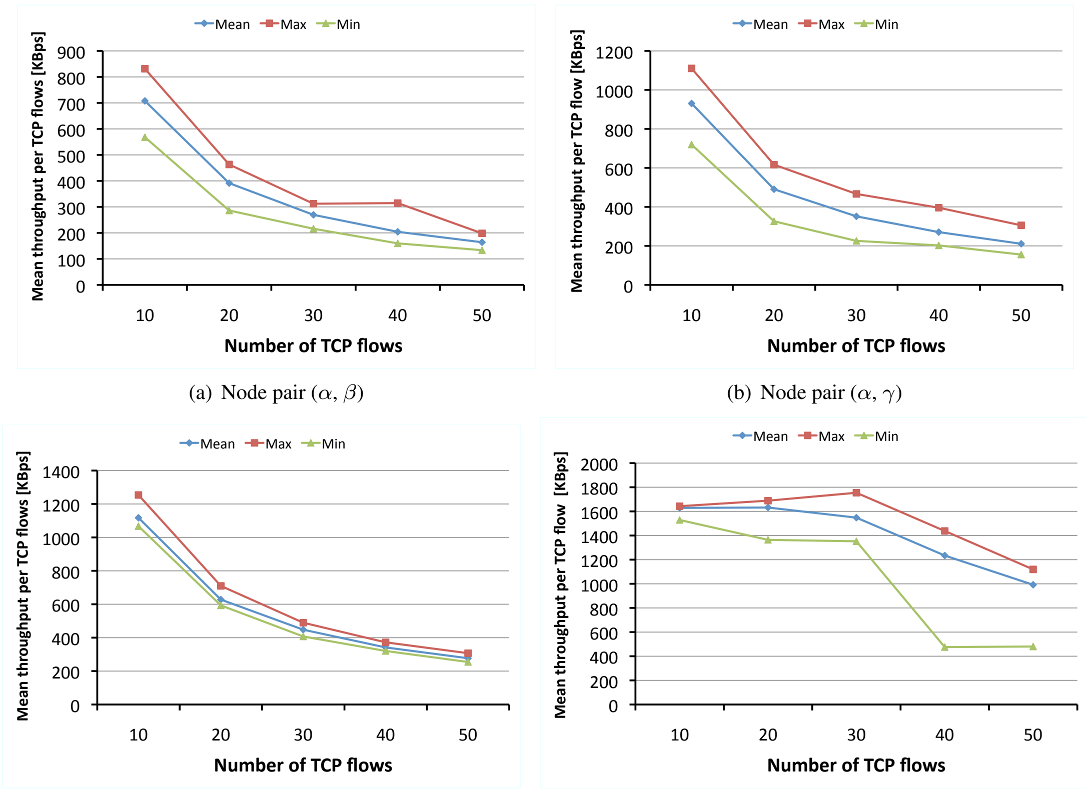
flows. The results are similar to Hacker's study. Next, we focus on the results at node pairs (δ, ϵ) and (δ, ζ). The different characteristics of statistics of aggregate throughput at node pair (δ, ζ) are observed. The aggregate throughput is increased up to 30 TCP flows, thus it is increased after 10 TCP flows and it is different from the previous results. A range of aggregate throughput is also different from the other results. Moreover, the minimum values with 40 and 50 TCP flows are 19033.7 and 24022.0 [KBps] respectively, and these are dramatically decreased in comparison with the other results.

4.2 Mean throughput per TCP flow

In order to observe mean throughput at each TCP flow, we calculate the statistics of mean throughput per TCP flow. It is described in Fig. 3. When the number of TCP flows is increased, the mean throughput is gradually decreased. It is obvious that the saturation is occurred after 20 TCP flows. However, the mean throughput at node pair (δ, ζ) is different from the above cases. Because the mean throughput is not degraded up to 30 TCP flows, it is possible to gradually increase the aggregate throughput. With 40 and 50 TCP flows, the minimum values are 475.8 and 480.4 [KBps] respectively. These values corresponded with the minimum values of aggregate throughput. We will discuss the cause of these values later. In future, we intend to investigate what causes lead to the changes in aggregate throughput at node pair (δ, ζ).

4.3 Discussion

Here, we discuss the throughput measurement results. There are two major causes of throughput fluctuation. Network congestion is one of the causes. In congested networks, a lot of packets are dropped and retransmitted. Next, lots of TCP flows or the incorrect number of TCP flows is a major cause of poor aggregate throughput while the correct number of TCP flows can increase the aggregate throughput. Thus, TCP flows competed with each other, and unfairness would be occurred by these flows. Actual throughput per TCP flow with 10 and 50 TCP flows at all node pairs is depicted in Fig. 4. Again, we get rid of the measurement results when the experiment is failed. Actual throughput at node pairs (α, β) and (α, γ) with 10 TCP flows is widely fluctuated, thus the unfairness among the flows and the network congestion would be oc-



(a) Node pair (α, β)

(b) Node pair (α, γ)

(c) Node pair (δ, ϵ)

(d) Node pair (δ, ζ)

Fig. 3 Statistics of throughput per TCP flow at node pairs

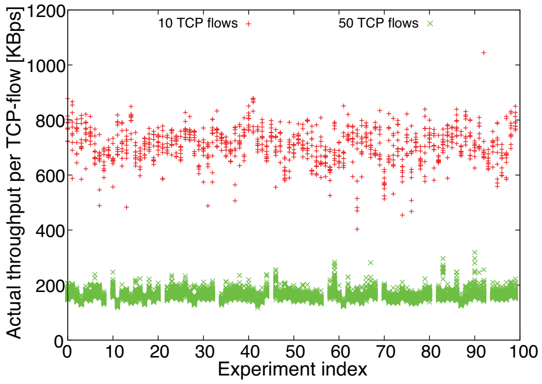
curred. Then, actual throughput at node pair (α, β) is poorer than that at node pair (α, γ) totally. However, actual throughput with 50 TCP flows is nearly concentrated at certain areas (around 200 [KBps]). It would be difficult to estimate that the unfairness is a major cause of decreased aggregate throughput. Moreover, only a few of TCP flows at node pair (α, γ) are similar to the results with 10 TCP flows. We will investigate the cause of these flows. At node pair (δ, ϵ), the similar results are also observed.

Actual throughput per TCP flow at node pair (δ, ζ) has different characteristics. With 10 TCP flows, there is a gap among TCP flows, but it is smaller than the above cases. Moreover, most results are stable and these are different from the results at the other pairs. With 50 TCP flows, actual throughput divided into two clusters: fast flows and slow flows. It would be obvious to show that a major cause of fluctuation is the unfairness. Moreover, actual throughput at index 89 is the minimum value of the aggregate throughput (24022.0 [KBps]) and the mean throughput (480.4 [KBps]). Because all of cases at index 89 have decreased throughput, two causes would be considered: one is the network congestion, and the other is a negative impact by virtualization. We find the similar results with 40 TCP flows. Thus, the minimum values with 40 and 50 TCP flows would be caused by the network congestion or the impact of virtualization. To clarify the cause of above cases, we will investigate the measurement results in packet-level. Moreover, we will investigate why aggregate throughput with 10, 20, and 30 TCP flows is increased.

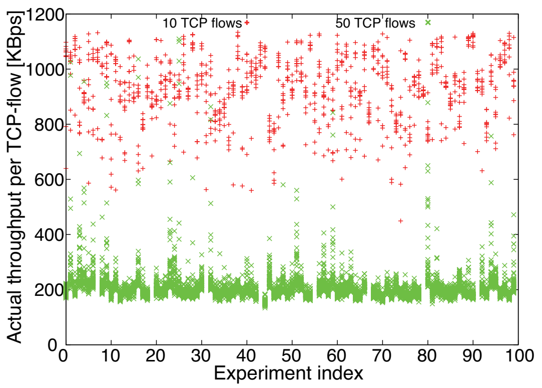
To summarize, partial measurement results accord with the previous results shown by Hacker et al. The aggregate throughput is gradually saturated, and the mean throughput per TCP flows was also saturated after 10 TCP flows. However, the results at node pair (δ, ζ) are different from the previous results. We find that the aggregate throughput is increased after 10 TCP flows. While the unfairness is occurred with 50 TCP flows, it is hard to estimate that there is the unfairness with 10 TCP flows. In future, we intend to investigate what causes lead to the above results.

5. Conclusion

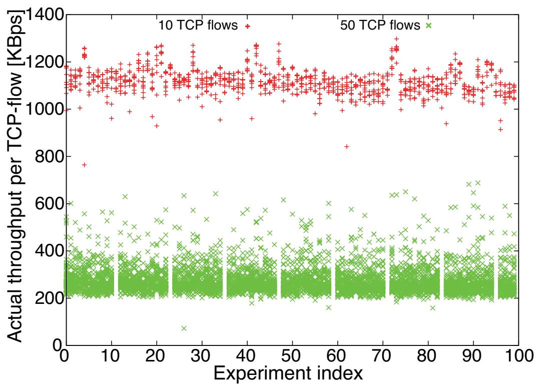
In this work, we focused on the re-production of



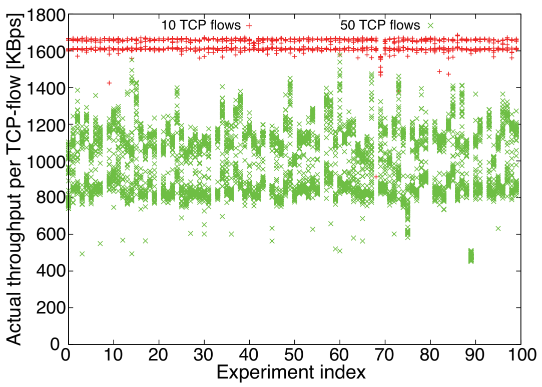
(a) Node pair (α, β)



(b) node pair (α, γ)



(c) Node pair (α, δ)



(d) node pair (δ, ζ)

Fig. 4 Actual throughput per TCP flow at node pairs.

throughput measurements as Hacker et al.[4] did, but the current Internet environment. Node pairs are selected on different PlanetLab sites across North America and Europe. The throughput measurements start with parallel TCP flows up to 50. Partial measurement results accord with previous results shown by Hacket et al. It is obvious that aggregate throughput and mean throughput per TCP flow are gradually saturated after 10 TCP flows. The aggregate throughput is fluctuated by network congestion and unfairness among TCP flows. However, most of actual throughput with 50 TCP flows is concentrated in certain areas. Moreover, a part of results is different from the previous results. The aggregate throughput is increased after 10 TCP flows. And also the mean throughput up to 30 TCP flows is close to each other. In actual throughput with 10 TCP flows, it would be hard to estimate the occurrence of the congestion or the unfairness. Contrarily, the unfairness is a major cause with 50 TCP flows.

In future work, we will firstly analyze the measurement results in packet-level, and clarify what causes lead to the above results. Finally, we intend to perform the throughput measurements with various TCP flows and different network environment, and develop a theoretical model of parallel TCP flows for recent network environment.

Acknowledgments

This work was supported in part by JSPS KAKENHI (22500072 and 24700031).

References

1) Sivakumar, H., Bailey, S. and Grossman, R. L.: PSockets: the case for application-level network striping for data intensive applications using high speed wide area networks, *Proceedings of the 2000 ACM/IEEE conference on Supercomputing (CDROM)*, Supercomputing '00, Washington, DC, USA, IEEE Computer Society (2000).
2) bbFTP: http://doc.in2p3.fr/bbftp/.
3) GridFTP: http://www.globus.org/toolkit/docs/latest-stable/gridftp/.
4) Hacker, T. J., Athey, B. D. and Noble, B.: The End-to-End Performance Effects of Parallel TCP Sockets on a Lossy Wide-Area Network, *IPDPS '02: Proceedings of the 16th International Parallel and Distributed Processing Symposium*, Washington, DC, USA, IEEE Computer Society, p. 314 (2002).
5) Altman, E., Barman, D., Tuffin, B. and Vojnovic, M.: Parallel TCP Sockets: Simple Model, Throughput and Validation, *INFOCOM 2006. 25th IEEE International Conference on Computer Communications*, Washington, DC, USA, IEEE Computer So-

ciety, pp. 1–12 (2006).

6) PlanetLab: https://www.planet-lab.org/.

7) Bavier, A., Bowman, M., Chun, B., Culler, D., Karlin, S., Muir, S., Peterson, L., Roscoe, T., Spalink, T. and Wawrzoniak, M.: Operating system support for planetary-scale network services, *NSDI'04: Proceedings of the 1st conference on Symposium on Networked Systems Design and Implementation*, Berkeley, CA, USA, USENIX Association, pp. 19–32 (2004).

8) Yin, D., Yildirim, E. and Kosar, T.: A data throughput prediction and optimization service for widely distributed many-task computing, *Proceedings of the 2nd Workshop on Many-Task Computing on Grids and Supercomputers*, MTAGS '09, New York, NY, USA, ACM, pp. 1–10 (2009).

9) Nakamura, M., Inaba, M. and Hiraki, K.: Fast Ethernet is Sometimes Faster than Gigibit Ethernet on LFN Observation of Congestion Control of TCP Streams, *PDCS '03: Proceedings of the Thirteenth IASTED International Conference, Parallel and Distributed Computing and Systems*, IASTED, pp. 392–397 (2003).

10) Kamezawa, H., Nakamura, M., Tamatsukuri, J., Aoshima, N., Inaba, M. and Hiraki, K.: Inter-Layer Coordination for Parallel TCP Streams on Long Fat Pipe Networks, *Proceedings of the 2004 ACM/IEEE conference on Supercomputing*, SC '04, Washington, DC, USA, IEEE Computer Society, pp. 24–33 (2004).

11) Kamezawa, H., Nakamura, M., Inabe, M., Hiraki, K., Jinzaki, A., Shitami, J., Kurusu, R., Nakano, S., Torii, K., Yanagisawa, T. and Ikuta, Y.: Coordination between parallel TCP streams in long distance and high bandwidth transmission, *the SACSIS (Symposium on Advanced Computing Systems and Infrastructures)*, Tokyo, Japan, IPSJ, pp. 425–432 (2004). (in Japanese).