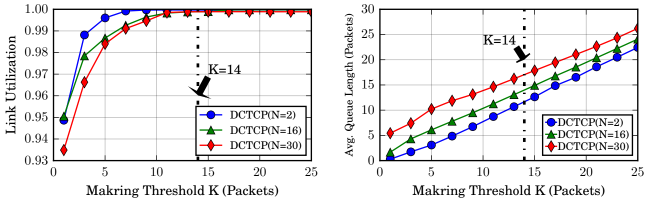
Adaptively Adjusting ECN Marking Thresholds for Datacenter Networks

Shuo Wang*‡, Jiao Zhang*†‡, Tao Huang*†‡, Tian Pan*‡, Jiang Liu*‡ and Yunjie Liu*†‡
{shuowang, jiaozhang, htao, pan, liujiang, yunjieliu}@bupt.edu.cn
*State Key Laboratory of Networking and Switching Technology, BUPT, China
†Beijing Advanced Innovation Center for Future Internet Technology, Beijing, China
‡Beijing Laboratory of Advanced Information Networks, Beijing, China

***Abstract*—ECN thresholds have limited operational range and very strict scope. Lower thresholds exacerbate the queue underflow while higher thresholds increase the queueing delays. In this paper, an Adaptive ECN (A-ECN) marking scheme is proposed to enhance the performance of ECN. A-ECN can adaptively adjust ECN marking thresholds in different scenarios to achieve good generality. Therefore, network operators can directly deploy A-ECN in various environments regardless of underlying queue types and bandwidth.**

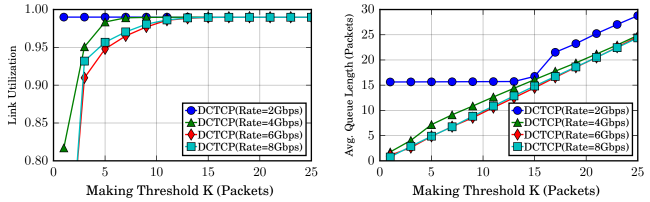
***Index Terms*—Datacenter network; DCTCP; ECN; Congestion control;**



(a) Throughput variation

(b) Average queue length variation

Fig. 1: [Single-Queue] Throughput and average queue length variations when $C = 10Gbps$, K is varied from 1 to 25 packets.



(a) Throughput variation

(b) Average queue length variation

Fig. 2: [Multi-Queue] Throughput and average queue length variations when $N = 16$, K is varied from 1 to 25 packets.

I. Introduction

ECN has been employed by Data Center TCP (DCTCP) [1] to achieve fine-grained window control. Instead of dropping packets, active queue management schemes can use ECN to inform sources about the congestion before the queue overshoot. As a result, DCTCP will react to ECN marks and reduce the window size in proportion to the fraction of ECN marked packets. Because of its simplicity and good performance, DCTCP and ECN-based window control schemes have attracted great attentions from both academia and industry.

Although DCTCP delivers both high throughput and low latency, its performance is mainly determined by the ECN marking threshold K. K has limited operation range and very strict scope. Lower or higher K can greatly decrease the performance in both single-queue and multi-queue environments. Indeed, through our observation from the simulation of DCTCP, we demonstrate that, in single-queue environments, the setting of K is related to the number of flows N. Using a large K ensures high throughput when N is also large, but will cause bufferbloat [2] and thus increase queueing delays. In contrast, using a small K ensures low queueing delays, but will throttle flows when N is large. In the multi-queue scenario, the setting of K is more complex, besides N, and it is also directly related to the bandwidth of queues. More specifically, if a port has multiple queues, the bandwidth of each queue is not static. The bandwidth of a queue may increase when some queues are idle, and thus the queue will be throttled by K. The bandwidth of a queue may decrease when other queues are active, and thus the queueing delays will increase.

Therefore, in this paper, we propose A-ECN, an adaptive ECN marking mechanism that provides high throughput and low latency simultaneously regardless of underlying queue types and bandwidth. The threshold K_a is the key to our mechanism, and it is adaptively adjusted to adapt to the traffic dynamics. It guarantees the queue has minimal queueing delays while keeping optimal throughput. More importantly, K_a makes our mechanism have good generality to adapt to different types of queues and network environments.

II. Observation

Using the sawtooth model of DCTCP, prior work [1] shows in order to fully use the bandwidth, the minimal K should be set as follows:

$$K \approx 0.17CD \tag{1}$$

where C and D is the capacity and delay of the path.

Observation 1: As shown in Figure 1, in single queue, using K given by Equation (1) makes sure bandwidth is fully utilized, but it causes large queueing delay when N is small.

978-1-5090-6501-1/17/$31.00 ©2017 IEEE

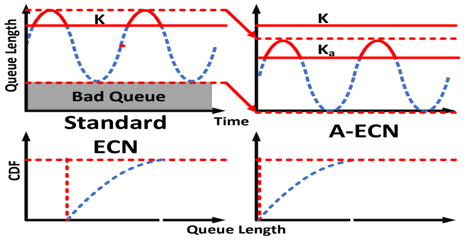


Fig. 3: The marking method of standard ECN and A-ECN

Observation 2: As shown in Figure 2, in multi-queue, if we set the threshold according to the guaranteed bandwidth of a queue, the throughput can hardly be fully utilized when the bandwidth of the queue increases.

III. Design

Algorithm 1 Check Queue State

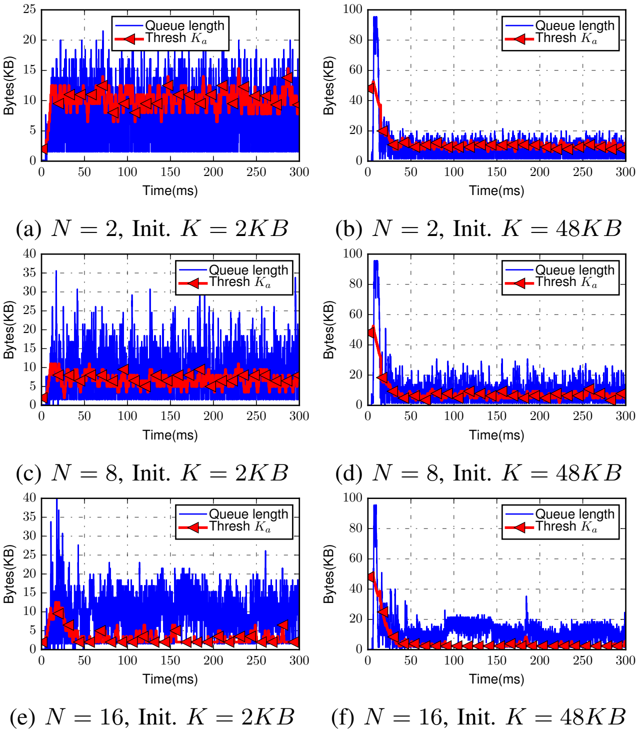
```
if $\alpha < T_{lower}$ then
    Q.State=Overshoot
end if
if $\alpha > T_{upper}$ then
    Q.State=Underflow
end if
```

A-ECN is an adaptive ECN marking mechanism for datacenter switches. One of the design goals of A-ECN is to achieve good generality for all types of queues. To achieve this goal, the mechanism should be oblivious to the queue type and only leverages the single-queue state to adjust the threshold. As shown in Figure 3, our basic idea is to adjust the threshold according to the state of a queue. We assume each queue has three states: overshoot, underflow, normal. To determine the states of a queue, we use Algorithm 1. In the algorithm, we count how many times the queue length is zero when a packet enqueues (denote as $Counter_{zero}$), and α is $\frac{Counter_{zero}}{Counter_{total}}$ when $Counter_{total}$ packets enqueue. If the queue overshoot happens, the threshold is decreased by 1 packet. If the queue underflow happens, the threshold is increased by 1 packet. Otherwise, we don't adjust the threshold.

We use a simple example shown in Figure 3 to illustrate this. In the standard ECN, to make sure the bandwidth can be fully utilized in all situations, the marking threshold K is usually set to a large value. The queue length will oscillate at a high value, which causes the bad queue [3] and increases the queueing delay. In contrast, A-ECN adjusts the marking threshold K_a to a small value that keeps queue oscillating at a relatively small scope to remove the bad queue.

IV. Testbed Evaluation

In this experiment, we inject 2, 8, 16 flows to the same receiver at the same time by varying the initial value of K_a, and then we plot the instance queue length and the evolution of K_a in Figure 4. We can find that A-ECN can quickly adaptively adjust K_a according to the queue length variation to a proper value before 20 ms in all the simulations. Specifically, if K_a is too small, A-ECN observers the queue underflows and then increases K_a, and if K_a is too large, A-ECN observers bad queue and then decreases K_a. Therefore, K_a is converged to the optimal value.



(a) $N = 2$, Init. $K = 2KB$ (b) $N = 2$, Init. $K = 48KB$

(c) $N = 8$, Init. $K = 2KB$ (d) $N = 8$, Init. $K = 48KB$

(e) $N = 16$, Init. $K = 2KB$ (f) $N = 16$, Init. $K = 48KB$

Fig. 4: [Testbed Single-Queue] Queue length variations when there are 2, 8 and 16 flows.

V. Conclusion

In this paper, we have presented A-ECN for both single-queue and multi-queue datacenter networks that can minimize queue length without a negative impact on utilization. The evaluation results show A-ECN achieves good generality for different types of queues.

Acknowledgments

This work is supported in part by National Natural Science Foundation of China (NSFC) under Grant No. 61502049, National High-Tech Research and Development Plan of China (863 Plan) under Grant No. 2015AA016101, Young talent development program of the CCF, Young Elite Scientist Sponsorship Program by CAST under Grant No.2015QNRC001, Beijing New-star Plan of Science and Technology under Grant No. Z151100000315078, and the 111 project(NO.B17007).

References

[1] M. Alizadeh, A. Greenberg, D. A. Maltz, J. Padhye, P. Patel, B. Prabhakar, S. Sengupta, and M. Sridharan, "Data Center TCP (DCTCP)," in *ACM SIGCOMM*, 2011.

[2] J. Gettys and K. Nichols, "Bufferbloat: Dark buffers in the internet," *ACM Queue*, 2011.

[3] K. Nichols and V. Jacobson, "Controlling Queue Delay," *Communications of the ACM*, 2012.