Right buffer sizing matters: some dynamical and statistical studies on Compound TCP

Debayani Ghosh, Krishna Jagannathan and Gaurav Raina

Abstract

Motivated by recent concerns that queuing delays in the Internet are on the rise, we conduct a performance evaluation of Compound TCP (C-TCP) in two topologies: a single bottleneck and a multi-bottleneck topology, under different traffic scenarios. The first topology consists of a single bottleneck router, and the second consists of two distinct sets of TCP flows, regulated by two edge routers, feeding into a common core router. We focus on some dynamical and statistical properties of the underlying system. From a dynamical perspective, we develop fluid models in a regime wherein the number of flows is large, bandwidth-delay product is high, buffers are dimensioned small (independent of the bandwidth-delay product) and routers deploy a Drop-Tail queue policy. A detailed local stability analysis for these models yields the following key insight: smaller buffers favour stability. Additionally, we highlight that larger buffers, in addition to increasing latency, are prone to inducing limit cycles in the system dynamics, via a Hopf bifurcation. These limit cycles in turn cause synchronisation among the TCP flows, and also result in a loss of link utilisation. For the topologies considered, we also empirically analyse some statistical properties of the bottleneck queues. These statistical analyses serve to validate an important modelling assumption: that in the regime considered, each bottleneck queue may be approximated as either an $M/M/1/B$ or an $M/D/1/B$ queue. This immediately makes the modelling perspective attractive and the analysis tractable. Finally, we show that smaller buffers, in addition to ensuring stability and low latency, would also yield fairly good system performance, in terms of throughput and flow completion times.

Index Terms

Compound TCP, Drop-Tail, Buffer sizing, Local stability, Hopf bifurcation

I. Introduction

There is an increasing concern regarding large queuing delays in today's Internet. This rise in queuing delays has primarily been attributed to a phenomenon called *bufferbloat*; *i.e.*, the presence of large and persistently full buffers

D. Ghosh, K. Jagannathan and G. Raina are with the Department of Electrical Engineering, IIT Madras, Chennai 600036, India. Email: {ee12s052, krishnaj, gaurav}@ee.iitm.ac.in

A part of this work appeared in [14].

in Internet routers [6], [12]. Excessive queuing delays, caused by these large buffers, would be a hindrance to the efficient functioning of various real-time delay-sensitive applications such as Voice over IP (VoIP), live streaming video and online gaming. Several factors impact the end-to-end latency, and hence the quality of service in the Internet: namely, size of buffers in routers, the choice of TCP, and the queue management scheme implemented at the routers. Currently, three buffer sizing regimes have been proposed in the literature [29]: a large, an intermediate, and a small buffer regime. In practice, today's router buffers follow the large buffer rule. In particular, this rule mandates the buffer size $B = C \times \overline{RTT}$, where C is the link capacity of the router, and $\overline{RTT}$ is the harmonic mean of the round trip times of the flows accessing the router [7], [32]. In practice, $\overline{RTT}$ is typically chosen to be 250 ms. This rule leads to larger buffers as the capacity of the router increases. The intermediate buffer rule mandates the buffer size $B = C \times \overline{RTT}/\sqrt{N}$, where N is the number of long-lived flows multiplexed at the router [35]. In the small buffer regime, the buffer size at the router is chosen independent of the bandwidth-delay product [29].

There have also been numerous proposals for TCP flavours in the literature. However, Compound TCP [31] (C-TCP) is the default protocol in the Windows operating system and Cubic TCP [16] is used in Linux. Recent studies [36] have shown that $15\% \sim 25\%$ of $30,000$ web servers implement Compound TCP. Given the large fraction of web servers that currently use Compound TCP, we primarily focus on Compound TCP for our study.

As far as queue management is concerned, solutions have been proposed in an attempt to eradicate the pervasive problem of excessively large queuing delays. The primary aim of an active queue management strategy is to maintain the bottleneck buffers small by dropping or marking packets before the buffers become full. Some common examples of active queue management strategies are RED [10], CODEL [23], and PIE [25]. However, in practice, router buffers widely deploy a simple Drop-Tail policy which drops incoming packets if the buffer is full.

In this paper, we conduct a performance evaluation of Compound TCP, in conjunction with small Drop-Tail buffers, in two topologies. We start with a single bottleneck topology and then proceed towards a more complex topology with three bottleneck routers. At a broad level, we distinguish between dynamical and statistical properties of the underlying system. We wish to emphasise that a single bottleneck topology has been widely used in the literature to understand the properties of TCP [19], [28]–[30]. However, to the best of our knowledge, this work is the first to analyse the properties of Compound TCP in a multiple bottleneck topology.

A. Dynamical properties

The first topology we consider consists of a single bottleneck router. A large number of TCP flows feed into this router, either with equal round trip times or with heterogeneous round trip times. For the traffic, we consider three scenarios. In the first scenario, the traffic consists of a large number of only long-lived flows. In the second scenario, apart from the presence of many long-lived flows, short flows arrive and depart the network. In the third scenario, we assume file sizes to be drawn from a heavy-tailed distribution [2], [20]. This is motivated by the fact that measurements on real Internet traffic show the presence of high variability at the connection level [34].

For the single bottleneck topology, we first outline the fluid models for various scenarios. We then conduct a local stability analysis, in the small buffer regime, and derive conditions that ensure local stability and non-oscillatory

convergence. A key insight obtained from our stability analysis is that *smaller buffers are favourable for local stability*. In fact, even minor variations in sizing router buffers would drive the underlying dynamical systems into a locally unstable regime.

The second topology is comprised of two edge routers fed by two sets of long-lived TCP flows, each with a different round trip time. The outputs from the edge routers feed into a core router. For this topology, deriving the necessary and sufficient condition for local stability in full generality is rather hard. To make progress, we conduct the stability analysis with two simplifying assumptions, and outline necessary and sufficient conditions for local stability. We also outline a rather simple sufficient condition for local stability, which could provide design guidelines to dimension router buffers in a decentralised manner. For the multiple bottleneck topology, our analysis highlights the importance of *smaller buffers in ensuring local stability*. Further, larger buffers would drive the system from a locally stable to an unstable regime.

Another important contribution of this work lies in determining the behaviour of the system as it transits from a locally stable to an unstable regime. In this work, the fluid models outlined for Compound TCP are parameterised, non-linear, time-delayed dynamical systems. Such time-delayed systems can readily lose local stability if either the feedback delay or other system parameters vary [17]. In our models, we show that the transition from stability to instability occurs via a Hopf bifurcation, which alerts us to the emergence of limit cycles [15], [18]. Such limit cycles were indeed observed in our packet-level simulations, conducted via the Network Simulator version 2.35 (NS2) [37]. In the context of TCP, the emergence of limit cycles manifest as: (i) synchronisation effects among TCP windows, (ii) periodic oscillations in the queue-size occupancy and (iii) loss of link utilisation.

B. Statistical properties

In this work, we also empirically investigate some statistical properties of the bottleneck queues, in the presence of a large number of TCP flows. Numerous studies have empirically shown that real Internet traffic exhibits long range dependence [26], [34], and a Poisson modelling for the packet arrivals might not be appropriate [26]. However, these studies are applicable to core links which typically use bandwidth-delay worth of buffering. In [2], [33], the authors have shown that with small Drop-Tail queues and a large number of TCP Reno flows, the packet arrival process can be approximated as Poisson. A statistical analysis with Compound TCP is in order. In particular, we pay attention to the arrival process to the queue as well as the queue size distribution.

For the single bottleneck topology, our empirical study reveals that in the absence of synchronisation, *the bottleneck queue may be well approximated by either an $M/M/1/B$ or an $M/D/1/B$ queue*. We would like to emphasise that this approximation holds reasonably well even in the presence of high variability at the TCP connection level. Thus, for the analysis, this allows us to approximate the drop probability of the bottleneck queue using the blocking probability of an $M/M/1/B$ queue. Notably, this insight carries over even to the multiple bottleneck topology. In particular, our empirical study highlights that, when a large number of flows feed into each of the edge routers, *each bottleneck queue can be well approximated by either an $M/M/1/B$ or an $M/D/1/B$ queue.* This validates our theoretical approximation of the drop probability at each queue using the loss probability

expression of an $M/M/1/B$ queue, for our local stability analysis.

Our analysis highlights the importance of smaller buffers in ensuring stability. Hence, it becomes imperative to study the impact of buffer sizing on both network and user performance. To that end, we consider two performance metrics: throughput and Average Flow Completion Time (AFCT), and show that it is indeed possible to significantly reduce buffers at routers while guaranteeing acceptable network and user performance.

In summary, the primary contributions of our work are the following:

(1) From a dynamical perspective, we highlight the interplay between buffer sizes and stability of the systems in the presence of feedback delays. In particular, we show that smaller buffers aid stability. Additionally, larger buffers would increase queuing delay, in addition to inducing limit cycles in the system dynamics. We then show that this insight holds true in each of the topologies, and traffic scenarios considered in our work. This lends credence to the fact that the loss of stability and consequently the emergence of limit cycles is primarily influenced by the buffer sizes at the routers, in the presence of large feedback delays.

(2) From a statistical perspective, we show that the bottleneck queues can be well approximated by either an $M/M/1/B$ or an $M/D/1/B$ queue, with smaller buffers. Notably, this insight remains consistent across the topologies considered. This validates our modelling assumption, and makes our system models amenable to analysis.

(3) We show that smaller buffers can indeed be realised without degrading the system performance, namely, throughput and flow completion times.

The rest of this paper is organised as follows. In Section II, we briefly outline the congestion avoidance algorithm of Compound TCP. In Section III, we analyse a single bottleneck topology with long-lived, and a combination of long- and short-lived flows, with a single feedback delay. In Section IV, we study the single bottleneck topology with heterogeneous delays. In Section V, we analyse the single bottleneck topology under a traffic scenario wherein users generate heavy-tailed files. Some of our analytical insights are corroborated by packet-level simulations, conducted using NS2. In Section VI, we study a multiple bottleneck topology using a combination of analysis and packet-level simulations. In Section VII, we investigate the impact of our buffer sizing recommendations on the system performance. Finally, in Section VIII we summarise our contributions.

II. Compound TCP

Compound TCP (C-TCP) [31] is a widely implemented Transmission Control Protocol (TCP) in the Windows operating system. Transport protocols like Reno and HighSpeed (HSTCP) use packet loss as the only indication of congestion, and protocols like Vegas [3] uses only queuing delay as the measure of network congestion. C-TCP is a synergy of both loss and delay-based feedback. The motivation behind incorporating both forms of feedback in C-TCP is to achieve high link utilisation and also to provide fairness to other competing TCP flows.

Compound TCP incorporates a scalable delay-based component into the congestion avoidance algorithm of TCP Reno. C-TCP controls its packet sending rate by maintaining two windows, a loss window $cwnd$ and a delay window $dwnd$. In a time period of one round trip time, C-TCP updates its sending window w as follows:

$$w = \min\left(cwnd + dwnd, awnd\right). \tag{1}$$

Here, $awnd$ is the advertised window at the receiver side. The decision function (1) governing the evolution of the sending window guarantees flow control between the source and the destination if the end systems operate at different speeds. In our paper, we assume that the sending window is constrained only by the congestion in the network path and not by the congestion at the receiver. Hence, for C-TCP, the source's sending window will always be $cwnd + dwnd$. The loss window $cwnd$, behaves similar to the loss window of TCP Reno and follows the Additive Increase and Multiplicative Decrease (AIMD) rule whereas the delay window $dwnd$, controls the delay-based component. If there is no loss detected, $cwnd$ increases by one packet in one round trip time and reduces by half whenever a loss is signalled. The algorithm for the delay-based component of Compound TCP is motivated from TCP Vegas. A state variable, $baseRTT$ gives the transmission delay of a packet in the network path. If the current round trip time of the TCP flow is RTT, then

$$diff = \Big(\frac{w}{baseRTT} - \frac{w}{RTT}\Big)\, baseRTT,$$

gives the amount of backlogged data in the bottleneck queue. If $diff$ is less than the congestion threshold γ, the network is considered underutilised and the TCP flow increases its packet sending rate. If $diff$ exceeds γ, congestion is detected in the network path which prompts the TCP flow to decrease its delay-based component. In C-TCP implementation, the default value of γ is fixed to be 30 packets. The overall behaviour of the window size of a C-TCP sender can then be summarised by combining the loss window and the delay window. When there is no congestion in the network path, neither in terms of increased queuing delay nor packet losses, a C-TCP sender increases its window size in its congestion avoidance phase as follows:

$$w(t+1) = w(t) + \alpha w(t)^k. \tag{2}$$

Here, α, k are the increase parameters and their default values are $\alpha = 0.125$ and $k = 0.75$ [31] respectively. If a packet loss is detected, the window size is multiplicatively reduced as follows:

$$w(t+1) = w(t)(1-\beta). \tag{3}$$

Here, β is the decrease parameter and its default value is $\beta = 0.5$ [31].

III. Single bottleneck with homogeneous delay

This topology consists of a single bottleneck link with *many* TCP flows feeding into a bottleneck router (see Fig. 1). We consider the case where the buffer at the core router is sized *small* [29] and deploys a Drop-Tail queue policy. The flows are subject to a common round trip time τ, and the bandwidth-delay product is assumed to be *large*. Let the *average* window size of the flows be $w(t)$. Then the average rate at which packets are sent is approximately $x(t) = w(t)/\tau$. Let the average congestion window increase by $i(w(t))$ for each received acknowledgement and decrease by $d(w(t))$ for each packet loss detected. The following non-linear, time-delayed differential equation describes the evolution of the *average* window size in the congestion avoidance phase [30]

$$\frac{dw(t)}{dt} = \frac{w(t-\tau)}{\tau}\left(i\left(w(t)\right)\Big(1 - p(t-\tau)\Big) - d\left(\left(w(t)\right)p(t-\tau)\right)\right), \tag{4}$$

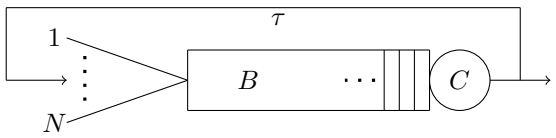


Fig. 1: Single bottleneck topology with many TCP flows feeding into a router. The flows have a common round trip time τ. The bottleneck queue has buffer size B and link capacity C.

where $p(t)$ denotes the loss probability experienced by packets sent at time t. We assume that the packet losses are independent across all TCP flows. If the server capacity is high, then it is easy to see that the packet loss probability would depend on the instantaneous arrival rate and consequently on the window size $w(t)$. We can then rewite (4) as

$$\dot{w}(t) = \frac{w(t-\tau)}{\tau}\left(i\left(w(t)\right)\left(1-p(w(t-\tau))\right) - d\left((w(t))\,p(w(t-\tau))\right)\right). \tag{5}$$

We analyse and study system (5) in three scenarios. In the first scenario, we assume that all TCP flows are long-lived, *i.e.*, each TCP source sends an infinite sized file. In the second scenario, the traffic is a mix of long and short-lived flows. In the third scenario, each TCP source sends a Poisson stream of finite sized connections, and the size of each connection is sampled from a heavy-tailed dstribution.

Our focus will be on C-TCP, however other variants of TCP, like TCP Reno and HighSpeed TCP, can also be analysed via (5). The following are the functional forms of $i\left(w(t)\right)$ and $d\left(w(t)\right)$ for Compound, Reno and HighSpeed TCP.

- Compound TCP

$$i\left(w\left(t\right)\right) = \frac{\alpha\left(w(t)\right)^{k}}{w(t)} \text{ and } d\left(w\left(t\right)\right) = \beta w(t). \tag{6}$$

- TCP Reno

$$i\left(w\left(t\right)\right) = \frac{1}{w(t)} \text{ and } d\left(w\left(t\right)\right) = \frac{w(t)}{2}. \tag{7}$$

- HighSpeed TCP

$$i\left(w\left(t\right)\right) = \frac{f_1\left(w(t)\right)}{w(t)} \text{ and } d\left(w\left(t\right)\right) = f_2\left(w(t)\right)w(t), \tag{8}$$

where $f_1(.)$ and $f_2(.)$ are continuous functions of the window size, and are given as

$$f_1\left(w(t)\right) = \frac{0.156w^2(t)f_2\left(w(t)\right)}{w^{1.2}(t)\left(2-f_2\left(w(t)\right)\right)}, \text{ and}$$
$$f_2\left(w(t)\right) = \frac{-0.4\left(\log\left(w(t)\right)-\log\left(38\right)\right)}{\left(\log(83000-\log(38)\right)} + 0.5.$$

The equilibrium of system (5) satisfies

$$i(w^*)(1-p(w^*)) = d(w^*)p(w^*). \tag{9}$$

We let $u(t) = w(t) - w^*$, and linearise (5) about its non-trivial equilibrium point w^* to get

$$\dot{u}(t) = -au(t) - bu(t-\tau), \tag{10}$$

where

$$\begin{aligned} a &= -\frac{w^*}{\tau}\left(i'(w^*)(1-p(w^*)) - d'(w^*)p(w^*)\right), \text{ and} \\ b &= \frac{w^*}{\tau}p'(w^*)\left(i(w^*)+d(w^*)\right). \end{aligned} \tag{11}$$

Looking for exponential solutions of (10), we get

$$s + a + be^{-s\tau} = 0. \tag{12}$$

We now outline the necessary and sufficient and sufficient conditions for (10) to be asymptotically stable. This would then yield the corresponding local stability conditions for the original non-linear system (5).

A. Local stability and Hopf bifurcation analysis with long-lived flows

As the number of traffic sources feeding into a small buffer router increases, in the limiting regime, the buffer overflow probability tends to the corresponding probability with Poisson arrivals [5]. This finding motivates us to approximate the packet drop probability of the core router by the corresponding probability of an $M/M/1/B$ queue, where B is the buffer size. For analytical purposes, we use the buffer exceedance probability of an $M/M/1$ queue as a surrogate for the overflow probability of an $M/M/1/B$ queue. This approach has been commonly used in the literature. This yields the packet loss probability at the core router as

$$p(w(t)) = \left(\frac{w(t)}{C'\tau}\right)^B, \tag{13}$$

where C' is the service rate per flow of the bottleneck link, and B is the buffer size. A comment is in order. Note that owing to TCP's congestion control algorithm, our scenario constitutes a *closed loop* feedback system. However, we use the loss probability of an *open loop* queueing system as a substitute for the packet loss probability at the bottleneck queue. A justification for this assumption is provided in [29], which says that for a queue with a small buffer, traffic characteristics over very short timescales matter, and if the average arrival rate to the queue does not exhibit much variation in such a short timescale, then this approximation is benign. We will empirically validate this approximation later.

We now outline some stability conditions for (10) to be asymptotically stable. From [27], if $a \geq 0$, $b > 0$, $b > a$ and $\tau > 0$, a sufficient condition for stability of (10) is

$$b\tau < \frac{\pi}{2}, \tag{14}$$

the necessary and sufficient condition for stability of (10) is

$$\tau\sqrt{b^2 - a^2} < \cos^{-1}(-a/b), \tag{15}$$

and the system undergoes the first Hopf bifurcation at

$$\tau\sqrt{b^2-a^2} = \cos^{-1}(-a/b). \tag{16}$$

We can now easily particularise the above conditions for Compound, Reno and HighSpeed TCP. However, we outline conditions for local stability only for Compound TCP, with long-lived flows. The necessary and sufficient condition for local stability with Compound TCP flows is [30]

$$\alpha\left(w^*\right)^{k-1}\sqrt{B^2-\left((k-2)(1-p(w^*))\right)^2} < \cos^{-1}\left(\frac{(k-2)(1-p(w^*))}{B}\right), \tag{17}$$

and a Hopf bifurcation would occur at

$$\alpha\left(w^*\right)^{k-1}\sqrt{B^2-\left((k-2)(1-p(w^*))\right)^2} = \cos^{-1}\left(\frac{(k-2)(1-p(w^*))}{B}\right). \tag{18}$$

A sufficient condition for local stability with Compound TCP flows is

$$\alpha B\left(w^*\right)^{k-1} < \frac{\pi}{2}.$$

Clearly, buffer thresholds and protocol parameters all greatly influence stability. In particular, if condition (17) gets violated, the system would lose local stability via a Hopf bifurcation. This would lead to the emergence of limit cycles in the system dynamics. In the context of TCP, these limit cycles manifest as oscillations in the packet arrival process and queue size dynamics [20], [29]. This would lead to synchronisation effects among TCP windows, and loss of link utilisation, and hence would be detrimental to the overall network performance.

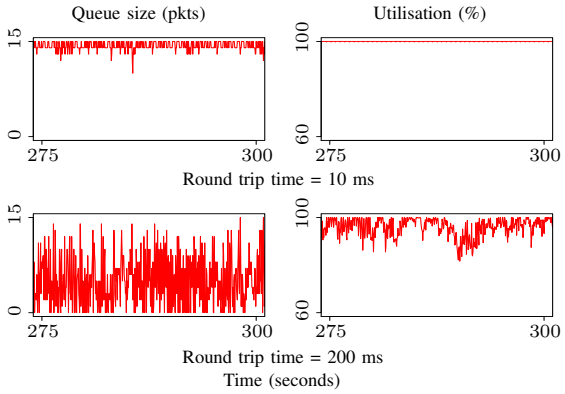
B. Robust stability analysis with long-lived flows

We now investigate the dynamical properties of system (5) under parametric uncertainties. Note that for networks with small buffer routers, the queueing delay can be assumed to be negligible as compared to the propagation delay, which is constant for fixed source destination pair. Hence, in this scenario, the feedback delay or the round trip time is also fixed. Further, the capacity of the network is known apriori, and protocol parameters are fixed. However, a protocol designer would not be aware of the exact buffer size at the core router. Hence, we allow uncertainties in the buffer size at the core router and assume that it lies in an interval, say $B \in [\underline{B}, \overline{B}]$. Here, both $\underline{B}$ and $\overline{B}$ lie in the small buffer regime. We then derive bounds on the protocol parameters and network parameters which would ensure that the system is locally asymptotically stable for all values of $B \in [\underline{B}, \overline{B}]$.

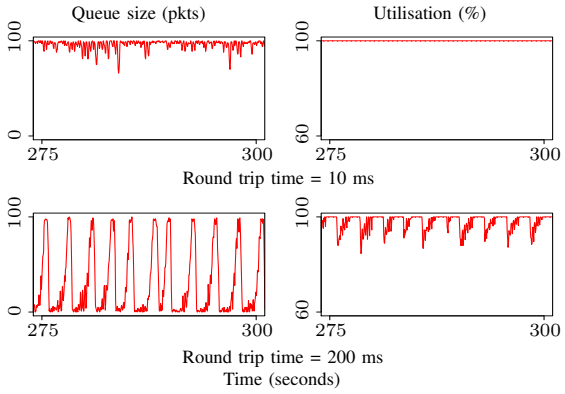
Recall that the characteristic equation of the linearised system (10) is

$$s + a + be^{-s\tau} = 0$$

Note that the above can be written in the form $P(s) + Q(s)e^{-s\tau}$, where $P(s) = s + a$ and $Q(s) = b$. Since $a > 0$, it is clear that $P(s)$ is a stable polynomial, and $deg(P) > deg(Q)$. As stated in [21], system (10) would be robust stable independent of the delay if and only if $P(j\omega) > Q(j\omega)$, $\forall\, \omega \geq 0$. This condition translates into $b < a$, $\forall\, \omega \geq 0$. However, we can easily show that this is not true for default values of protocol parameters of Compound, Reno and HighSpeed TCP, and physically relevant values of network parameters. Hence, system (5)



(a) Buffer size = 15



(b) Buffer size = 100

Fig. 2: *Long-lived flows*. 60 long-lived Compound TCP flows over a 2 Mbps link feeding into a single bottleneck queue with link capacity 100 Mbps. Note that as the buffer threshold of the bottleneck router increases, we see the emergence of limit cycles in the queue size.

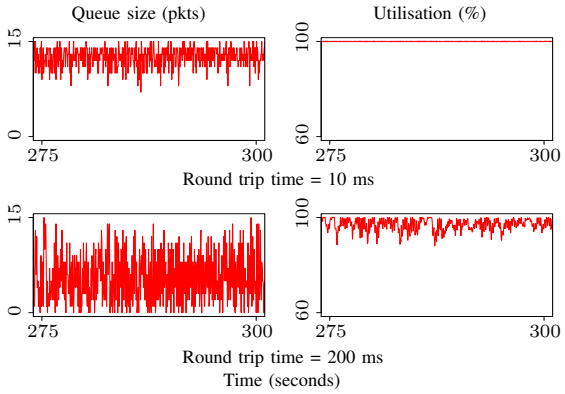
cannot be locally robust stable independent of the feedback delay. This motivates us to look for conditions for delay dependent robust stability.

As stated in [21], a sufficient condition for stability of (10) is $b\tau < 1$. It is easy to show that the equilibrium coefficient b is a monotonically increasing function of the buffer size of the core rourter B. This implies that if B lies in the interval $[\underline{B}, \overline{B}]$, then b would also lie in some interval $[\underline{b}, \overline{b}]$. Here, $\underline{b}$ and $\overline{b}$ can be evaluated by substituting $\underline{B}$ and $\overline{B}$ respectively in (11). Then, a sufficient condition for robust stability of system (5) is

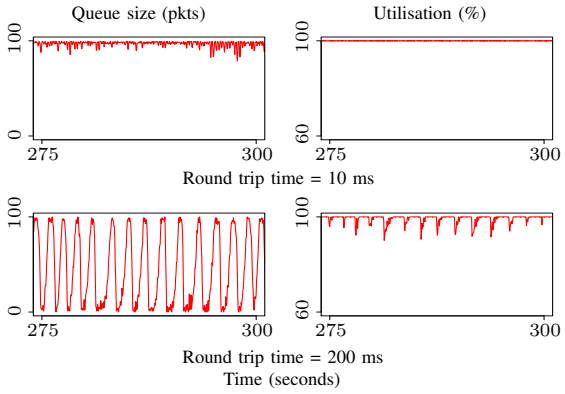
$$\overline{b}\tau < 1. \tag{19}$$

Using the functional forms (6), a sufficient condition for robust stability with Compound TCP flows can be outlined as

$$\alpha \overline{B} \left(w^*\right)^{k-1} < 1.$$



(a) Buffer size = 15



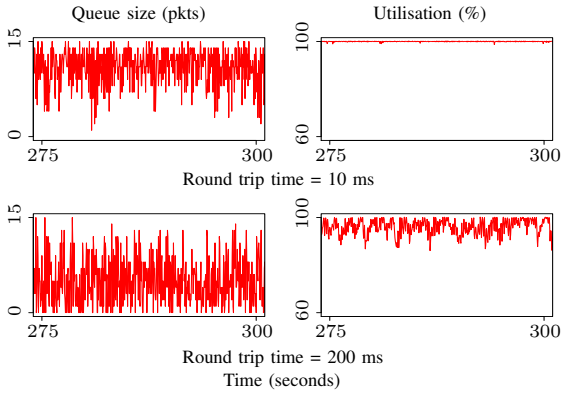
(b) Buffer size = 100

Fig. 3: *Long-lived and short-lived flows*. 55 long-lived Compound TCP flows over a 2 Mbps link, and *exponentially distributed short files*, feeding into a single bottleneck queue with link capacity 100 Mbps. Observe the emergence of limit cycles in the queue size, as the buffer threshold at the bottleneck router increases.

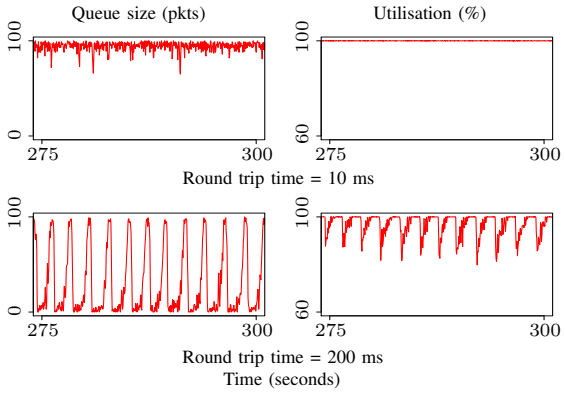
Observe that, condition (20) could provide guidlines to protocol designers to design transport protocols at end-systems, which could ensure stability of the system for a wide range of values of buffer thresholds at core routers.

C. Local stability and Hopf bifurcation analysis with long-lived and short-lived flows

We now deviate from the assumption that the system has only long-lived flows and consider the scenario where in addition to a large number of long-lived flows, short flows arrive and depart the network. On a short time scale, short TCP connections may act as an uncontrolled and random background load on the network. Suppose the workload per flow arriving at the bottleneck queue over a time period T is modelled as Gaussian with mean x^*T and variance $x^*\sigma_1^2 T$ and the background load per flow due to the short transfers over the time period T is also modelled as Gaussian with mean vT and variance $v\sigma_2^2 T$. Then the loss probability at the bottleneck queue can be



(a) Buffer size = 15



(b) Buffer size = 100

Fig. 4: *TCP and UDP flows*. 55 long-lived Compound TCP flows over a 2 Mbps link feeding into a single bottleneck queue with link capacity 100 Mbps. We consider 10 UDP flows each over a 1 Mbps link. As the buffer threshold at the bottleneck router increases, observe that the queue size dynamics exhibits limit cycles.

expressed as [22]

$$p(w^*) = \exp\left(\frac{-2B\left(C'\tau - w^* - v\tau\right)}{w^*\sigma_1^2 + v\sigma_2^2\tau}\right). \tag{20}$$

Recall that (14) gives a sufficient condition for local stability and (16) gives the condition for which the system undergoes a Hopf-type bifurcation. We now particularise the sufficient condition only for Compound TCP, however conditions for TCP Reno and HighSpeed TCP also can easily be outlined. We are also in a position to state the Hopf bifurcation conditions, which are left out due to space constraints.

For Compound TCP, a sufficient condition for local stability of the system is

$$2B\alpha\left(w^*\right)^k \tau \frac{v\sigma_2^2 + \left(C' - v\right)\sigma_1^2}{\left(w^*\sigma_1^2 + v\sigma_2^2\tau\right)^2} < \frac{\pi}{2}. \tag{21}$$

Condition (21) captures the relationship between the various protocol and network parameters. It is interesting to note that in general, larger the value of parameter B, greater the possibility of driving the system to an unstable state. In Compound TCP, there appears to be an intrinsic trade off in the choice of the parameter α and the queue

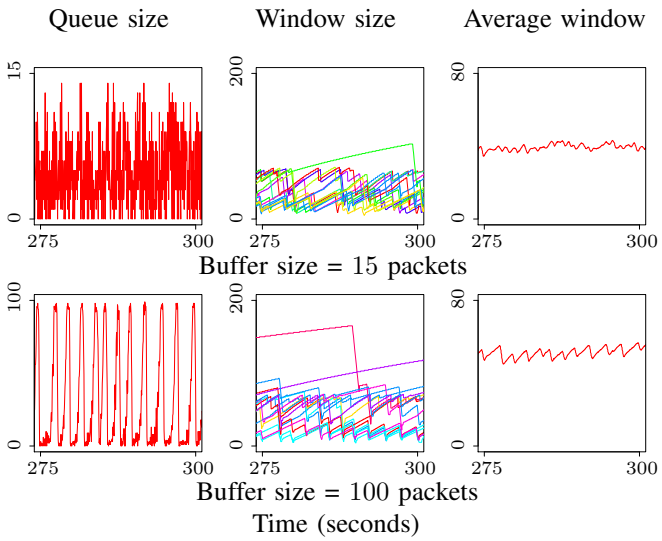


Fig. 5: *Compound TCP in single bottleneck topology.* 60 long-lived Compound TCP flows each with an access speed of 2 Mbps, over a single bottleneck link with a capacity 100 Mbps. Observe the emergence of limit cycles in the queue size, and synchronisation among Compound TCP windows, as the buffer threshold at the core router increases.

threshold parameter B. Indeed, it can be easily shown that increasing buffer sizes would prompt the system to lose local stability. The presence of short-lived flows, which are modelled here as random uncontrolled traffic, does not change the requirement of choosing smaller values of B to ensure stability. We now present some packet-level simulations, which will enable us to comment on the dynamical and statistical properties of the system.

D. Simulations

Dynamical Properties: We now conduct packet-level simulations, using NS2 [37], for the single bottleneck topology in a small buffer sizing regime. With small buffers, we employ 15 and 100 packets. We consider two scenarios (i) only long-lived flows, and (ii) a combination of long and short flows. The bottleneck link has a capacity of 100 Mbps. The packet size is fixed at 1500 bytes. In the scenario wherein only long-lived flows are present, we consider 60 and 120 long-lived flows where each flow has an access link speed of 2 Mbps and 1 Mbps respectively. With a combination of long- and short-lived flows, we consider 55 long flows each with an access speed of 2 Mbps. The file size of each of the short flows is exponentially distributed with a mean file size of 12.5 KB. The mean rate at which short flows arrive is 200 flows per second, and the total data rate contributed by these short flows is restricted to 20 Mbps.

Fig. 2 depicts the simulations where the system only has long-lived flows. With a buffer size of 15 packets, as expected, the queuing delay is negligible and the system is stable in the sense that there are no limit cycles in the queue size. With buffer size of 100 packets, with smaller round trip times, the queues are full which yields full link utilisation but at the cost of extra latency. With larger delays, limit cycles will emerge in the queue size which

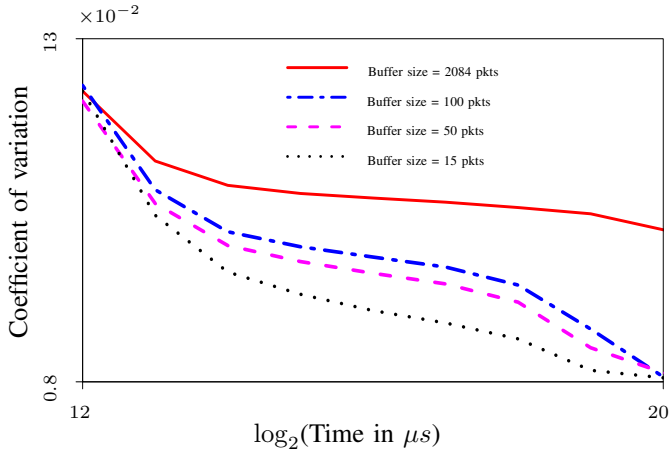


Fig. 6: *Statistics of the arrival process.* 60 long-lived Compound TCP flows each over a 2 Mbps link. The round trip time is fixed at 200 ms. The capacity of the bottleneck router is 100 Mbps. We consider three regimes: stable ($B = 15$ packets), presence of synchronisation ($B = 50$, and 100 packets), and bandwidth-delay product rule ($B = 2084$ packets). Observe that for smaller buffers (15 packets), the aggregate arrival process exhibits less burstiness or variability.

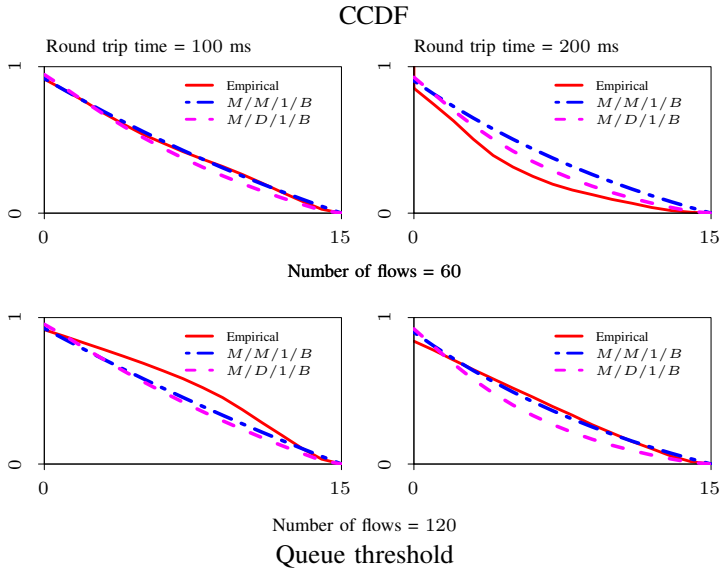


Fig. 7: *Statistics of the queue size.* Empirical queue length distribution for single bottleneck topology with 60 and 120 long-lived flows each having an access speed of 2 Mbps and 1 Mbps, with round trip times 100 ms and 200 ms respectively. We compare the empirical queue statistics with the queue length distributions of $M/M/1/B$, and $M/D/1/B$ for two different round trip times. Observe that as the number of long-lived flows increases, the approximation becomes better at a larger bandwidth-delay product.

also start to hurt link utilisation. Fig. 3 depicts the simulation results where the system has a combination of long and short-lived flows. Qualitatively, the results are very similar to those shown in Fig. 2. This is expected as the models did indeed predict that despite the presence of short flows the system could readily lose stability if key system parameters were not properly dimensioned.

Today's Internet is heterogeneous in nature. While applications like File Transfer Protocol (FTP) uses the services of TCP (closed loop), real-time applications like VoIP and online gaming use UDP (open loop). Further, the use of real-time applications is rapidly increasing. Hence, it is imperative to understand the impact of buffer sizes on the system stability, when both TCP and UDP traffic co-exist [33]. To that end, we consider a scenario with 55 long-lived Compound TCP flows each over an access link with a speed of 2 Mbps, and 10 UDP flows each over a 1 Mbps link. Fig. 4 shows the packet-level simulations for this scenario. We can immediately observe that the results obtained are qualitatively similar to the scenario with only long-lived flows. This suggests that the requirement of choosing buffer sizes carefully does not change even when open-loop UDP traffic is present.

The loss of local stability of the underlying dynamical system and hence the emergence of limit cycles indicates synchronisation among the TCP flows. For smaller buffer thresholds, all TCP flows would be totally de-synchronised and the mean window size would have small oscillations, see Fig. 5. As the buffer thresholds are increased, the mean window size would exhibit bigger oscillations owing to the synchronisation of the flows. We now empirically analyse some statistical properties of the bottleneck queue.

Statistical Properties: Statistical properties of the arrival process: We first conduct an empirical study on the statistical properties of the aggregate arrival process to the bottleneck queue. In particular, we capture the impact of the buffer size at the bottleneck queue on the burstiness or variability of the arrival process to the queue at different time scales. One way to characterise this is to measure the coefficient of variation (ratio of standard deviation to mean) of the arrival traffic at different time scales. In particular, we closely follow the method presented in [33] for our study.

For our empirical study, we consider three representative regimes: (i) $B = 15$ packets, wherein the underlying dynamical system is stable, (ii) $B = 50, 100$ packets, wherein the system dynamics exhibits limit cycles and synchronisation among TCP windows, and (iii) $B = 2084$ packets, which corresponds to the case of bandwidth-delay product worth of buffering. Note that for this buffer sizing rule, we have used a delay value of 250 ms, which is typically used in practice. This study would enable us to understand the buffer sizing regime in which the Poisson approximation for the aggregate arrival process seems justified. For our simulations, we consider the number of long-lived flows in the system to be 60, and the average round trip time to be 200 ms. Note that we are interested in measuring the burstiness of the arrival traffic at short time scales. Hence, we vary the time scale of aggregation from $2^{12}\,\mu s = 4$ ms to $2^{20}\,\mu s = 1$ second.

Fig. 6 depicts the coefficient of variation curves for the aggregate traffic arriving at the bottleneck queue, at different time scales and buffer sizes. For a buffer size of 15 packets, we can observe that the coefficient of variation curve falls quite rapidly for very short time scales, and does not change its slope significantly for larger time scales. On the contrary, as we gradually increase the buffer size, the coefficient of variation curve exhibits a

relatively slower decay over very short time scales, and flattens over larger time scales. Further, with a buffer size of 15 packets, the coefficient of variation values are lower than that of 50, 100 and 2084 packets. A larger value of coefficient of variation signifies the presence of higher variability or burstiness in the traffic arrival process to the bottleneck queue. This implies that when buffers at the bottleneck queue are large enough to cause synchonisation, the arrival traffic to the queue would be bursty. However, for a buffer size of 15 packets, there is no synchronisation among the TCP windows, and we can observe reduced burstiness in the traffic arrival process over short time scales. This suggests that only when the buffer size is sized small enough to avoid synchronisation, the aggregate arrival process behaves qualitatively similar to short-range dependent processes, a typical example of which is a Poisson process. This lends credence to the fact that packet arrivals to the bottleneck queue can be reasonably approximated as Poisson, with smaller buffers when there is no synchronisation.

Next, we empirically validate that the queue size distribution can be reasonably approximated by that of either an $M/M/1/B$ or an $M/D/1/B$ queue, with smaller buffers.

Statistical properties of the queue size: We have already established that when the buffer size at the bottleneck router is small enough to mitigate synchronisation effects, *i.e.*, 15 packets, the packet arrival process can be approximated by a Poisson process. Hence, to study the statistical properties of the queue size, we fix the buffer size at 15 packets.

In Fig. 7, we demonstrate the empirical Complementary Cumulative Distribution Function (CCDF) of the queue size. We consider 60 long-lived flows, each with an access speed of 2 Mbps. We also consider a scenario with 120 long-lived flows each with an access speed of 1 Mbps. We consider the time scale to be 50 ms, consistent with our time scale of interest. The round trip times which we choose are 100 ms and 200 ms.

For each of these cases, we perform a comparative study of the empirical queue length distribution with the theoretical queue distributions of $M/M/1/B$ and $M/D/1/B$ queues. From the simulations we can infer that, with 60 long-lived flows, the empirical queue distribution can be reasonably approximated by the corresponding queue distribution of either an $M/M/1/B$ or an $M/D/1/B$ queue when the bandwidth-delay product is large. Notably, as the number of flows is increased to 120, this approximation holds true for larger round trip times perhaps due to increased statistical multiplexing. Indeed, we verify that, as the number of long-lived flows is increased further, this approximation holds true for even larger bandwidth-delay product values. Hence, with increased statistical multiplexing, the approximation holds at a larger bandwidth-delay product.

Thus, our empirical study serves to validate a very important modelling assumption: even with TCP (closed loop) controlled traffic, the packet drop probability at the bottleneck router can be reasonably approximated using that of an $M/M/1/B$ queue, in the absence of synchronisation.

IV. Single bottleneck with heterogeneous delays

Till now, we have highlighted the interplay between buffer thresholds and stability of the underlying dynamical systems in a scenario wherein all flows are subject to a common round trip time. At this juncture, a natural question which might arise is: are the synchronisation effects of TCP windows for larger buffer thresholds an artefact of the

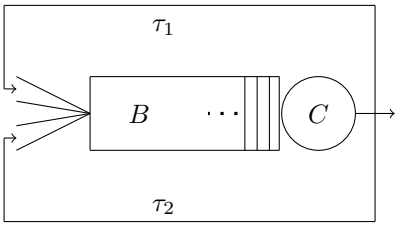


Fig. 8: *Single bottleneck topology with heterogeneous round trip times.* Two sets of TCP flows with round trip times τ_1 and τ_2, feeding into a common bottleneck router. The buffer size at the router is B and the link capacity is C.

assumption that all flows have a common round trip time? To answer this, we now investigate the impact of buffer thresholds on the dynamical properties of the system in the single bottleneck topology for the following scenario.

This model consists of a single bottleneck link with two distinct sets of *many* long-lived TCP flows feeding into a common core router, as shown in Fig. 8. The core router has a buffer size of B, with service rate per flow as C'. From a modelling perspective, both sets of TCP flows can be of different flavours and hence, can have different increase and decrease rules to govern the evolution of the corresponding window sizes. However, in this paper, we assume that the traffic in both sets is controlled by Compound TCP. Let the average window sizes of the two sets of flows be $w_1(t)$ and $w_2(t)$ respectively. For each acknowledgement received, the average window sizes increase by $i(w_1(t))$ and $i(w_2(t))$, and for each packet loss detected, the average window sizes decrease by $d(w_1(t))$ and $d(w_2(t))$ respectively. Thus, for generalised TCP flows, the non-linear, time-delayed, fluid model of the system can be outlined

$$\dot{w}_j(t) = \frac{w_j(t-\tau_j)}{\tau_j}\left(i\left(w_j(t)\right)\Big(1 - q(t,\tau_1,\tau_2)\Big) - d\left(\left(w_j(t)\right) q(t,\tau_1,\tau_2)\right)\right), \quad j = 1, 2, \tag{22}$$

where $q(t,\tau_1,\tau_2)$ represents the packet loss probability at the core router, and depends on the sending rates of both sets of TCP flows. Recall that the buffer size at the core router is dimensioned small, and the router deploys a Drop-Tail queue policy. When the bandwidth-delay product is large, the fluid model for the loss probability at the core router can be approximated as

$$q(t) = \left(\frac{w_1(t)/\tau_1 + w_2(t)/\tau_2}{\widetilde{C}}\right)^B. \tag{23}$$

Here, $\widetilde{C} = 2C'$. Using this functional form for the loss probability at the core router, we now perform a local stability analysis for the system given by (22). This would enable us to understand the impact of buffer thresholds on stability of the underlying dynamical system in presence of heterogeneous feedback delays.

A. Local stability and Hopf bifurcation analysis with long-lived flows

Suppose (w_1^*, w_2^*) is a non-trivial equilibrium of (22) and let $u_1(t) = w_1(t) - w_1^*$ and $u_2(t) = w_2(t) - w_2^*$ be small perturbations about w_1^* and w_2^* respectively. Linearising (22) about this equilibrium, we obtain

$$\begin{aligned}\dot{u}_1(t) &= -\mathcal{M}_1 u_1(t) - \mathcal{N}_1 u_1(t-\tau_1) - \mathcal{P}_1 u_2(t-\tau_2),\\ \dot{u}_2(t) &= -\mathcal{M}_2 u_2(t) - \mathcal{N}_2 u_2(t-\tau_2) - \mathcal{P}_2 u_1(t-\tau_1).\end{aligned} \tag{24}$$

Here, the increase and decrease functions for Compound TCP given by (6), and the functional form of the loss probability at the core router given by (23) yield the following equilibrium coefficients:

$$\begin{aligned}\mathcal{M}_j &= -\frac{\alpha}{\tau_j}(k-2)\ \left(w_j^*\right)^{k-1}\left(1 - \frac{1}{(2C')^B}\left(\frac{w_1^*}{\tau_1} + \frac{w_2^*}{\tau_2}\right)^B\right),\\ \mathcal{N}_j &= \frac{B\left(w_j^*\right)^2}{\tau_j^2\,(2C')^B}\left(\alpha\left(w_j^*\right)^{k-2} + \beta\right)\left(\frac{w_1^*}{\tau_1} + \frac{w_2^*}{\tau_2}\right)^{B-1},\\ \mathcal{P}_j &= \frac{B\left(w_j^*\right)^2}{\tau_1\tau_2\,(2C')^B}\left(\alpha\left(w_j^*\right)^{k-2} + \beta\right)\left(\frac{w_1^*}{\tau_1} + \frac{w_2^*}{\tau_2}\right)^{B-1},\end{aligned} \tag{25}$$

for $j = 1, 2$. Looking for exponential solutions, we obtain the characteristic equation for the linearised system (24) as

$$\begin{aligned}&\lambda^2 + \lambda\left(\mathcal{N}_1 e^{-\lambda\tau_1} + \mathcal{N}_2 e^{-\lambda\tau_2}\right) + \left(\mathcal{M}_1\mathcal{N}_2 e^{-\lambda\tau_1} + \mathcal{M}_2\mathcal{N}_1 e^{-\lambda\tau_2}\right)\\ &\quad + \lambda\left(\mathcal{M}_1 + \mathcal{M}_2\right) + \mathcal{M}_1\mathcal{M}_2 = 0.\end{aligned} \tag{26}$$

Now, for the linearised system (24) to be asymptotically stable, all roots of the characteristic equaion (26) should have negative real parts. We then aim to find conditions on different system parameters which would ensure asymptotic stability of the linearised system (24). However, obtaining the necessary and sufficient condition for the asymptotic stability of a system having a characteristic equation of the form (26) analytically seems rather hard. Hence, to investigate how different system parameters impact the asymptotic stability of (24), we outline a sufficient condition for stability for this system. To that end, we make use of a result which was derived in [21]. This would then yield a sufficient condition for stability of (22) about its equilibrium.

Note that the equilibrium (w_1^*, w_2^*) satisfy the following conditions:

$$\begin{aligned}\alpha\left(w_1^*\right)^{k-1} &= \left(\alpha\left(w_1^*\right)^{k-1} + \beta w_1^*\right)\frac{1}{(2C')^B}\left(\frac{w_1^*}{\tau_1} + \frac{w_2^*}{\tau_2}\right)^B,\\ \alpha\left(w_2^*\right)^{k-1} &= \left(\alpha\left(w_2^*\right)^{k-1} + \beta w_2^*\right)\frac{1}{(2C')^B}\left(\frac{w_1^*}{\tau_1} + \frac{w_2^*}{\tau_2}\right)^B.\end{aligned} \tag{27}$$

From (27), it can be shown that

$$\frac{\alpha(w_1^*)^{k-1}}{\alpha(w_1^*)^{k-1} + \beta w_1^*} = \frac{\alpha(w_2^*)^{k-1}}{\alpha(w_2^*)^{k-1} + \beta w_2^*} \tag{28}$$

This implies that $(w_1^*)^{2-k} = (w_2^*)^{2-k}$. It can also be shown that this equilibrium is unique. At this equilibrium, the coefficients (55) reduce to

$$\mathcal{M}_j = \frac{\mathcal{A}}{\tau_j},\ \mathcal{N}_j = \frac{\mathcal{B}}{\tau_j^2},\ \text{and}\ \mathcal{P}_j = \frac{\mathcal{B}}{\tau_1\tau_2},\ \ j = 1, 2. \tag{29}$$

Here,

$$\mathcal{A} = -\alpha\,(k-2)\ \,(w^*)^{k-1}\left(1 - \frac{1}{(2C')^B}\left(\frac{w^*}{\tau_1} + \frac{w^*}{\tau_2}\right)^B\right),\ \text{and}$$
$$\mathcal{B} = \frac{B\,(w^*)^2}{(2C')^B}\left(\alpha\,(w^*)^{k-2} + \beta\right)\left(\frac{w^*}{\tau_1} + \frac{w^*}{\tau_2}\right)^{B-1}. \tag{30}$$

Observe that, the linearised system (24) can be re-written in the following matrix form

$$\underbrace{\begin{bmatrix}\dot{u}_1(t)\\ \dot{u}_2(t)\end{bmatrix}}_{\dot{\mathbf{U}}(t)} = \underbrace{\begin{bmatrix}-\mathcal{M}_1 & 0\\ 0 & -\mathcal{M}_2\end{bmatrix}}_{A}\underbrace{\begin{bmatrix}u_1(t)\\ u_2(t)\end{bmatrix}}_{\mathbf{U}(t)} + \underbrace{\begin{bmatrix}-\mathcal{N}_1 & 0\\ -\mathcal{P}_2 & 0\end{bmatrix}}_{A_1}\underbrace{\begin{bmatrix}u_1(t-\tau_1)\\ u_2(t-\tau_1)\end{bmatrix}}_{\mathbf{U}(t-\tau_1)}$$
$$+ \underbrace{\begin{bmatrix}0 & -\mathcal{P}_1\\ 0 & -\mathcal{N}_2\end{bmatrix}}_{A_2}\underbrace{\begin{bmatrix}u_1(t-\tau_2)\\ u_2(t-\tau_2)\end{bmatrix}}_{\mathbf{U}(t-\tau_2)}.$$

Succinctly, the above can be written as

$$\dot{\mathbf{U}}(t) = A\mathbf{U}(t) + \sum_{i=1}^{2} A_i\mathbf{U}(t-\tau_i) \tag{31}$$

As stated in [21], a sufficient condition for the asymptotic stability of the linear system (24) is

$$\sum_{i=1}^{2}\|A_i\|\tau_i < 1. \tag{32}$$

To obtain a sufficient condition for the asymptotic stability of (24), we now use the Frobenius norm of a matrix in (32) which yields

$$\mathcal{B}\sqrt{\frac{1}{\tau_1^2} + \frac{1}{\tau_2^2}} < \frac{1}{2}. \tag{33}$$

Substituting $\mathcal{B}$ in (33) and simplifying yields the following condition:

$$\alpha B\,(w^*)^{k-1}\sqrt{1 - \frac{2}{\frac{\tau_1}{\tau_2} + \frac{\tau_2}{\tau_1} + 2}} < \frac{1}{2}. \tag{34}$$

The above would then yield a sufficient condition for the local stability of system (22) about its equilibrium with Compound TCP flows. The above condition evidently highlights that buffer thresholds need to be dimensioned rather carefully to ensure stability. In particular, larger buffer thresholds might prompt the system to lose local stability and transit into a locally unstable regime.

Observe that in addition to buffer threshold of the core router and protocol parameters, the condition (34) depends on the ratios of the round trip times. To facilitate a better understanding of the impact of heterogeneous feedback delays on local stability, we consider two cases: (i) both τ_1 and τ_2 are comparable to each other, and (ii) either τ_1

or τ_2 is significantly smaller than the other. In both cases, increasing buffer thresholds might destabilise the system. This suggests that, even if one round trip time is large, local stability might be lost with increasing buffer sizes. However, condition (34) highlights that in the latter case, the stability region might be smaller as compared to the former.

B. Numerical computations

Since (34) is a sufficient condition for local stability of (22), violating the same by varying any model parameter would not guarantee the loss of local stability. We now numerically illustrate through DDE-BIFTOOL (version 2.03) [9], that system (22) indeed loses local stability if the buffer threshold at the core router is increased beyond a critical value. Specifically, local stability is lost via a Hopf bifurcation, when exactly one pair of complex conjugate roots crosses over the imaginary axis from left half to the right half of the complex plane. At the point of criticality, the system has exactly one pair of complex conjugate roots on the imaginary axis.

For the numerical computation, we consider the following values of the protocol parameters: $\alpha = 0.125$, $k = 0.75$, and $\beta = 0.5$. Further, we fix $\widetilde{C} = 140$ packets/second. Additionally, we consider the following values for the round trip times: (i) $\tau_1 = 0.1$ seconds and $\tau_2 = 0.2$ seconds (ii) $\tau_1 = 0.01$ seconds and $\tau_2 = 0.2$ seconds. We then vary the buffer threshold B in the interval $[10, 100]$. For case (i), system (22) undergoes a Hopf bifurcation at $B = 25$ packets. For case (ii), it occurs at $B = 15$ packets.

Stability charts: We now plot some stability charts to illustrate the impact of the system parameters on the local stability of (22), see Figs. 9 and 10. For this, we again consider two cases for the round trip times: (i) $\tau_1 = 0.1$ seconds and $\tau_2 = 0.2$ seconds (ii) $\tau_1 = 0.01$ seconds and $\tau_2 = 0.2$ seconds. We vary the buffer size B in the interval $[15, 50]$ and observe the variation in the protocol parameter α at the Hopf condition, and the boundary of the sufficient condition given by (34). The remaining parameters are fixed as follows: $k = 0.75$, $\beta = 0.5$, and $\widetilde{C} = 140$ packets/second. Note that the Hopf conditions are obtained numerically through DDE-BIFTOOL. In both cases, we observe a trade off between the parameters ensure stability. In particular, even if one round trip time is large, increasing buffer sizes would destabilise the system.

C. Packet-level simulations

Now, we present some packet-level simulations, which corroborate the insights obtained from our stability analysis. In particular, we show that emergence of limit cycles in the queue size dynamics, as the buffer size at the bottleneck router is increased.

We consider two sets of 30 long-lived Compound TCP flows, each with an access speed of 2 Mbps. The flows feed into a router with a link capacity of 100 Mbps. We first consider the case when the average round trip times of both sets of flows are comparable to each other, and are fixed as 100 ms and 200 ms respectively. It can be seen from Fig. 11 that as the buffer size is increased from 15 to 100 packets, the queue size exhibits limit cycles. We then consider the case when one round trip time is much smaller than the other. In this case, we fix the average round trip times of the two sets as 10 ms and 200 ms respectively, see Fig. 12. We can observe that a similar insight

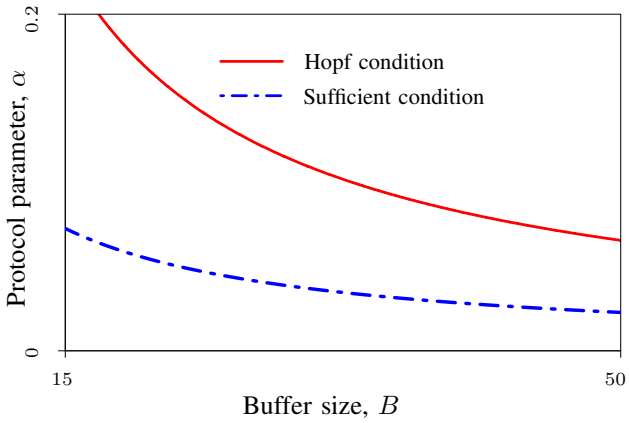


Fig. 9: *Stability Chart.* Shows the Hopf condition and sufficient condition for local stability with Compound TCP, with $\tau_1 = 0.1$ and $\tau_2 = 0.2$ seconds. The stability chart highlights the impact of the buffer size B and the protocol parameter α to ensure local stability.

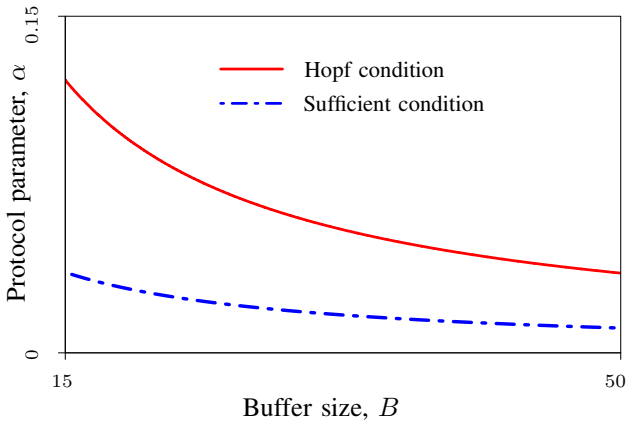


Fig. 10: *Stability Chart.* Shows the Hopf condition and sufficient condition for local stability with Compound TCP, with $\tau_1 = 0.01$ and $\tau_2 = 0.2$ seconds. The stability chart captures the trade off between the buffer size at the router B and the Compound parameter α to ensure local stability.

holds in this case also. Thus, even if one set of flows has a large round trip time, the underlying dynamical system loses stability if the buffer size at the bottleneck router is increased. This corroborates our analytical insights.

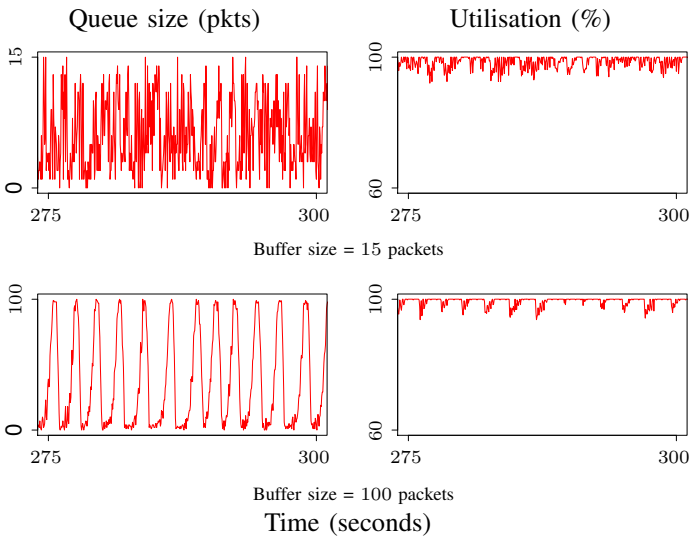


Fig. 11: *Queue size dynamics with heterogeneous round trip times.* Two sets of 30 TCP flows, each with an access speed of 2 Mbps, and feeding into a bottleneck router with a link capacity of 100 Mbps. The round trip times of the two sets of flows are fixed at 100 ms and 200 ms. It can be easily seen that that the queue size dynamics exhibits limit cycles, as the buffer size at the bottleneck router is increased.

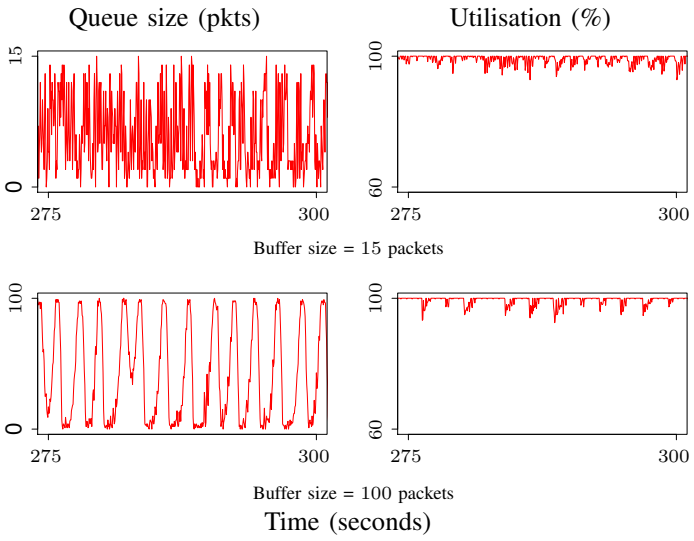


Fig. 12: *Queue size dynamics with heterogeneous round trip times.* Two sets of 30 TCP flows, each with an access speed of 2 Mbps, and feeding into a bottleneck router with a link capacity of 100 Mbps. The round trip times of the two sets of flows are fixed at 10 ms and 200 ms. As the buffer size at the bottleneck router is increased, the queue size exhibits limit cycles.

V. Single bottleneck with heavy-tailed files

Till now, we have shown the emergence of limit cycles in the queue dynamics in the presence of many TCP sources, each of which has an infinite file to transfer. At this juncture, a natural question that might arise is, are these limit cycles a consequence of the underlying assumptions on the network traffic? To answer this, we now analyse system (5) under a different workload model, which caters for more realistic traffic scenarios.

Empirical studies on real time Ethernet LAN traffic have established the presence of high variability at the TCP connection level [34]. It has been identified that heavy-tailed connections generated by individual TCP senders are a primary cause of this variability. To account for this, we consider the following workload model, which has been widely studied in the literature [2]. In this scenario, flows from each user arrive at the transport layer as a Poisson process, *i.e.,* the interarrival time between any two flows is exponentially distributed. To cater for the heavy-tailed nature of TCP connections, we assume that file sizes follow a Pareto distribution. For an overview on Pareto distribution, see [1]. The heavy-tailed nature of the file sizes ensures that each user generates a large file with a non-negligible probability. Hence, as the number of TCP senders increases, the number of long flows which are feedback controlled also increases. Apart from the presence of long flows, a salient feature of this workload model is that, a significant fraction of the total flows arriving at the transport layer is contributed by very short flows, or "mice." However, consistent with real network traffic, a major portion of the traffic is still contributed by the long flows. This motivates us to model the window evolution of the long flows in the congestion avoidance phase as (5), to capture the dynamical properties of this system.

Apart from high variability, it has been empirically shown that this source model also exhibits long-range dependence [34]. As argued in [5], even with long-range dependent sources, as the number of multiplexed sources feeding into a small buffer grows large, the aggregate packet arrival process tends towards Poisson. For analytical purposes, this gives us the confidence to use the same functional form of the packet loss probability as with long-lived flows, in this scenario. Recall that with long-lived flows, the packet loss probability at the core router is

$$p\left(w(t)\right) = \left(\frac{w(t)}{C'\tau}\right)^{B}. \tag{35}$$

Let V denote the random variable for the TCP connection sizes, and $\mathbb{E}(V)$ the expected file size in packets. Recall that the rate at which packets are transmitted is $x(t) = w(t)/\tau$. We define $\mu(t) = x(t)/\mathbb{E}(V)$.

A. Local stability and Hopf bifurcation analysis with heavy-tailed files

We now outline some stability conditions for Compound TCP, using the functional forms for the increase and decrease functions given by (6). We then obtain bounds on various model parameters, which would ensure local stability. Using(35), and $w^* = \mathbb{E}(V)\mu^*\tau$, the necessary and sufficient condition for local stability can be obtained

as

$$\alpha \left(\mathbb{E}(V)\mu^*\tau\right)^{k-1} \sqrt{B^2 - (k-2)^2 \left(1 - \left(\frac{\mathbb{E}(V)\mu^*}{C'}\right)^B\right)^2} < \cos^{-1}\left(\frac{(k-2)\left(1 - \left(\frac{\mathbb{E}(V)\mu^*}{C'}\right)^B\right)}{B}\right). \tag{36}$$

When condition (36) is met with an equality, we obtain the Hopf condition. Further, in this scenario, a simple sufficient condition with Compound TCP flows is

$$\alpha B \left(\mathbb{E}(V)\mu^*\tau\right)^{k-1} < \pi/2. \tag{37}$$

From conditions (36) and (37), we can easily observe that apart from protocol parameters and network parameters, local stability of (5) now depends on the expected file size brought by the sessions. If we assume that V follows a Pareto distribution with shape parameter $1 < \chi < 2$, then

$$\mathbb{E}(V) = \frac{\chi m}{\chi - 1}, \tag{38}$$

where m is defined as the scale parameter of the Pareto distribution. Using (38), we can then re-write the necessary and sufficient condition as

$$\alpha \left(\chi m\mu^*\tau\right)^{k-1} \sqrt{B^2 - (k-2)^2 \left(1 - \left(\frac{\chi m\mu^*}{(\chi-1)\,C'}\right)^B\right)^2} < (\chi-1)^{k-1} \cos^{-1}\left(\frac{(k-2)\left(1 - \left(\frac{\chi m\mu^*}{(\chi-1)C'}\right)^B\right)}{B}\right). \tag{39}$$

Similarly, condition (37) can be re-written as

$$\alpha B \left(\frac{\chi m\mu^*\tau}{\chi - 1}\right)^{k-1} < \pi/2. \tag{40}$$

Conditions (39) and (40) clearly highlight the interdependence of the shape parameter with network and protocol parameters to ensure stability of the system.

B. Non-oscillatory convergence with heavy-tailed files

We have already shown that both protocol parameters and network parameters such as queue thresholds or buffer sizes have to be chosen rather carefully if stability is to be ensured. Additionally, the file size distribution also impacts stability. However, even if local stability is ensured, the convergence of the system can be oscillatory or non-oscillatory. If the convergence is oscillatory, there would be some temporary degree of synchronisation among TCP flows before they can desynchronise again. There would be loss in link utilisation and bursty packet losses whenever synchronisation happens. If the system shows oscillatory convergence, and the oscillations in the queue size dynamics persist for a long time, it would affect the network performance. Hence, it becomes imperative to

design parameters such that non-oscillatory convergence of the system can be ensured. To that end, we seek bounds on various protocol and network parameters to ensure non-oscillatory convergence of (5). The following theorem outlines the necessary and sufficient condition for non-oscillatory convergence of the linearised system (10). This would then yield the necessary and sufficient condition for non-oscillatory convergence of the original system (5) in a neighbourhood of its equilibrium.

We now show that the solution of system (10) shows non-oscillatory convergence if and only if the parameters a, b and τ satisfy the condition $\ln(b\tau) + a\tau + 1 < 0$.

The boundary condition for the solution of (10) to be non-oscillatory is the point at which the curve $f(s) = s + a + be^{-s\tau}$ touches the real axis. If this point is σ, then

$$f(\sigma) = \quad \sigma + a + be^{-\sigma\tau} \quad = 0, \text{ and} \tag{41}$$

$$f^{'}(\sigma) = \quad 1 - b\tau e^{-\sigma\tau} \quad = 0. \tag{42}$$

From (42), we get

$$be^{-\sigma\tau} = \frac{1}{\tau} \text{ and } \sigma = \frac{\ln(b\tau)}{\tau}. \tag{43}$$

Substituting values of σ and $be^{-\sigma\tau}$ in (41) gives

$$\ln(b\tau) + a\tau + 1 = 0. \tag{44}$$

We now claim that the necessary and sufficient condition for non-oscillatory convergence of the equilibrium point is

$$\ln(b\tau) + a\tau + 1 < 0. \tag{45}$$

Suppose the solution of (10) exhibits non-oscillatory convergence to its equilibrium *i.e.* all roots of (12) are real and lie on the left half of the complex plane. Then, we prove that the region of non-oscillatory convergence is characterised by (45). We prove this claim by contradiction. We assume that the condition for non-oscillatory convergence is

$$\ln(b\tau) + a\tau + 1 > 0. \tag{46}$$

Let $\sigma = -\alpha$, where $\alpha > 0$ is a root of (12). Then, substituting $\sigma = -\alpha$ in (12), we obtain

$$\alpha = a + be^{\alpha\tau}. \tag{47}$$

Multiplying both sides of (47) by $\tau e^{a\tau}$ yields

$$\alpha\tau e^{a\tau} = a\tau e^{a\tau} + b\tau e^{a\tau}e^{\alpha\tau} > a\tau e^{a\tau} + \frac{e^{\alpha\tau}}{e}.$$

where, the inequality follows from (46). This leads to

$$\ln(\alpha\tau - a\tau) > -1 + (\alpha\tau - a\tau). \tag{48}$$

Now, it can be easily observed form (47) that $\alpha > a$. Hence, we obtain a contradiction (48), since $\ln x \leq x-1, \ \ \forall x > 0$. Thus, if the solution of system (10) exhibits non-oscillatory convergence, then the parameters a and b satisfy the following condition:

$$\ln(b\tau) + a\tau + 1 < 0.$$

Now, we prove the converse statement *i.e.* if the system parameters satisfy (45) and all roots of (12) have negative real parts, then all roots are real. To prove this by contradiction, we assume that all roots of (12) have non-zero imaginary part and are of the form $s = -\sigma - j\omega$, where $\sigma > 0$. Substituting s in (12) and separating real and imaginary parts, we obtain

$$\sigma = a + be^{\sigma\tau}\cos\omega\tau, \quad \text{and} \tag{49}$$

$$\omega = be^{\sigma\tau}\sin\omega\tau. \tag{50}$$

From (49) and (50), we get

$$\frac{\tan\omega\tau}{\omega\tau} = \frac{1}{(a-\sigma)\tau}. \tag{51}$$

Now, the condition given by (45) implies that $(\sigma - a)\tau \geq 1$. This in turn implies that $\omega = 0$ is the unique solution to the equation (51). Hence, the necessary and sufficient condition for non-oscillatory convergence of the solution of (10) is

$$\ln(b\tau) + a\tau + 1 < 0.$$

With Compound TCP, the necessary and sufficient condition for non-oscillatory convergence of (5) around its equilibrium is

$$\begin{aligned} &\alpha B\left(\chi m\mu^*\tau\right)^{k-1} \\ &< (\chi-1)^{k-1}\exp\left(\alpha(k-2)\left(\frac{\chi m\mu^*\tau}{\chi-1}\right)^{k-1}\left(1-\left(\frac{\chi m\mu^*}{(\chi-1)C'}\right)^{B}\right)-1\right). \end{aligned} \tag{52}$$

The above condition captures the dependence of protocol parameters α and k, network parameters like buffer size B, and shape parameter χ in ensuring the non-oscillatory convergence of the solution of (5) to its equilibrium point, for Compound TCP.

Stability charts: Given these conditions, we now aim to gain a better understanding of how various model parameters impact local stability. To that end, we plot some stability charts, as shown in Figs. 13, 14 and 15. Further, we assume $1 - p^* \approx 1$, to simplify the equilibrium structure of (5).

Fig. 13 shows the stability chart with respect to two parameters, the buffer size at the core router B, and the increase parameter α. We first fix the values of the remaining system parameters as $C' = 140$ packets/second, $\tau = 0.2$ seconds, $\mathbb{E}(V) = 100$ packets, and $\chi = 1.5$. The protocol parameters k and β are fixed at their default values 0.75 and 0.5 respectively. We now vary the buffer size B in $[5, 50]$, and observe the variation in the protocol parameter α at the stability boundary, which is obtained when (36) is met with an equality. The stability chart also

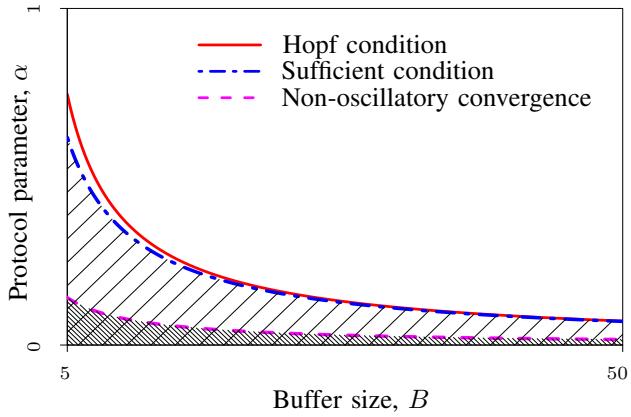


Fig. 13: *Stability chart for single bottleneck topology with heavy-tailed files.* The Hopf condition, the sufficient condition for local stability and condition for non-oscillatory convergence for compound TCP are outlined. The stability chart highlights the impact of the buffer threshold B and the protocol parameter α to ensure local stability as well as non-oscillatory convergence.

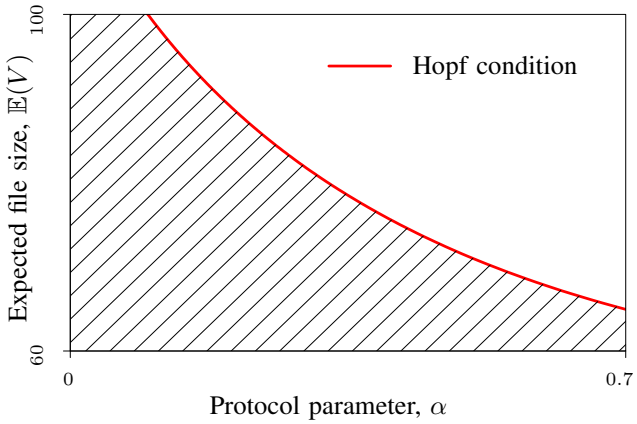


Fig. 14: *Stability chart for single bottleneck topology with heavy-tailed files.* The stability chart highlights the impact of the expected file size $\mathbb{E}(V)$ and the protocol parameter α on local stability. Observe the trade-off among the parameters to ensure stability.

illustrates the sufficient condition for local stability given by (37), and the necessary and sufficient condition for non-oscillatory convergence given by (52). It can be easily observed that if α is fixed, then increasing the buffer size B would destabilise the system. Hence, from a design point of view, buffer sizes at core routers should be

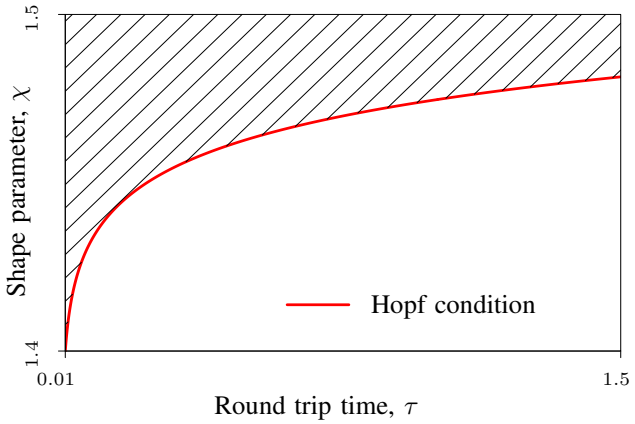


Fig. 15: *Stability chart for single bottleneck topology with heavy-tailed files.* The stability chart highlights the impact of the Pareto sized files characterised by the shape parameter χ and the round trip time τ on local stability. Note that a higher χ can accommodate a larger feedback delay.

dimensioned carefully to ensure local stability, as well as non-oscillatory convergence.

We next present a stability chart which characterises the Hopf condition for Compound TCP with respect to two parameters, the expected file size $\mathbb{E}(V)$ and the protocol parameter α, see Fig. 14. For this, we vary α in the interval $[0, 0.7]$, and find the corresponding value of B which satisfies the Hopf condition. Then, using the equilibrium condition for (5), and $w^* = \mathbb{E}(V)\mu^*\tau$, we find $\mathbb{E}(V)$ at the Hopf boundary. We fix the values of the remaining system parameters as $\beta = 0.5$, $k = 0.75$, $C' = 140$ packets/second, $\tau = 0.2$ seconds, and $\chi = 1.5$. Additionally, we choose $\mu^* = 1$. We can see that keeping α fixed, increasing $\mathbb{E}(V)$ would prompt the system to lose local stability via a Hopf bifurcation and transit into the unstable region.

At this juncture, a natural question which arises is, how does the file size distribution impact local stability? As argued before, TCP connection sizes are well modelled by the Pareto distribution, which is characterised by its shape parameter or the tail index. Hence, to better understand the dependence of local stability on the file size distribution, we plot a stability chart with respect to the shape parameter χ and the feedback delay τ, as shown in Fig. 15. For this, we vary the τ in $[0.01, 1.5]$, and find the corresponding value of α which satisfies the Hopf condition. Then, using the equilibrium condition for (5), $w^* = \mathbb{E}(V)\mu^*\tau$ and (38), we find χ at the Hopf boundary. We fix the remaining parameters as $C' = 140$ packets/second, $B = 100$ packets, $\beta = 0.5$ and $k = 0.75$. Further, we choose $\mu^* = 1$, and $m = 40$ packets. We can observe that for a fixed round trip time, increasing the shape parameter χ would have a stabilising effect on the system. Further, for larger values of χ, the system can accommodate larger feedback delays. However, if the round trip times become too large, the system could still lose local stability via a Hopf bifurcation. We now present some packet-level simulations to corroborate our analytical insights.

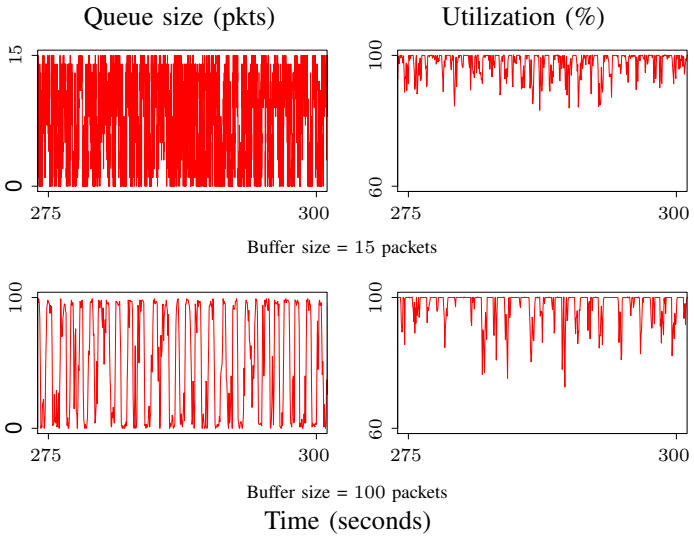


Fig. 16: *Compound TCP with heavy-tailed files.* 100 TCP sources each with an access link speed of 2 Mbps feeding into a core router of 100 Mbps. The round trip time is fixed at 200 ms. Observe the emergence of limit cycles in the queue size dynamics for a buffer size of 100 packets.

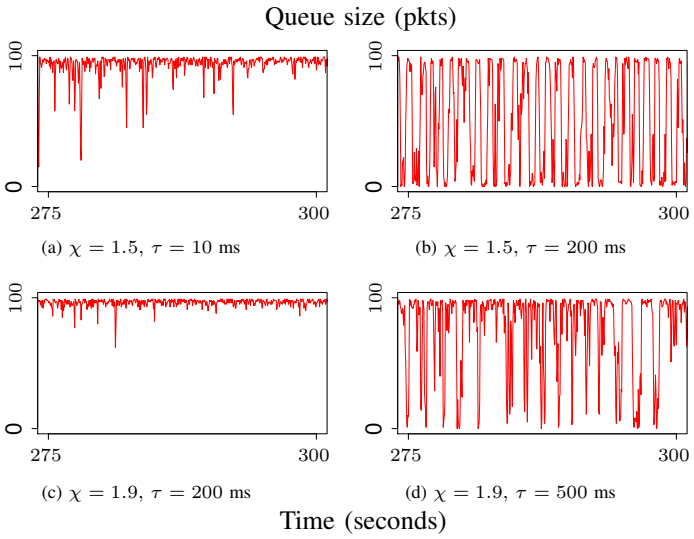


Fig. 17: *Compound TCP with heavy-tailed files.* 100 TCP sources each with an access link speed of 2 Mbps feeding into a core router of 100 Mbps. The buffer size is fixed at 100 packets. Observe the impact of the shape parameter χ and the round trip time τ on stability. As χ increases, the system stabilises. However, as the round trip time gets larger, the system again destabilises.

C. Packet-level simulations

Dynamical properties: For our packet-level simulations, we consider the following setup. Each user generates finite volume files with sizes drawn from a Pareto distribution. The files arrive at the access link as a Poisson

process with rate λ. All users are connected to the bottleneck router via the access links. If there are N users in the system, and the bottleneck link has a capacity C, then the offered load at the bottleneck link is $\rho = \mathbb{E}(V)N\lambda/C$. For our simulations, we vary the offered load to the system by varying λ.

For our simulations, we fix the number of TCP senders at 100, and the bottleneck capacity at 100 Mbps. Each access link has a speed of 2 Mbps. Further, the packet size is 1500 bytes.

Impact of the buffer size: To capture the impact of buffer sizes on stability, we observe the queue size dynamics for two values of the buffer size, 15 and 100 packets, see Fig. 16. The expected file size is chosen to be 100 kB, and the shape parameter is fixed at 1.5 [33]. Further, the average round trip time is fixed at 200 ms. With this set of parameter values, the offered load to the bottleneck link is 0.9. It can be observed that, when the buffer size at the core router is fixed at 15 packets, the queue size does not exhibit any periodic oscillations, and hence is stable. However, as the buffer size is increased to 100, the system destabilises, and we can observe the emergence of periodic oscillations in the queue size dynamics. This lends credence to the fact that these oscillations are not an artefact of the underlying assumptions for the traffic. Even if we take into account the high variability at the TCP connection level, the queue size dynamics could still exhibit periodic oscillations, if buffers are not dimensioned carefully.

Impact of the shape parameter: We now present a set of simulations which captures the impact of the file size distribution, *i.e.,* the shape parameter χ on stability. For this, we fix the buffer size at the core router at 100 packets. Fig. 17(a) shows the queue size dynamics when the shape parameter is 1.5, the expected file size is 100 kB, and the round trip time is 10 ms. For such a short round trip time, the feedback would be fast, and the queue does not have time to drain completely. Hence, in this case, the queue size is stable. As the round trip time is increased to 200 ms, we can observe the emergence of oscillations in the queue size dynamics, see Fig. 17(b). This is because, the long flows present in the network would experience synchronisation effects due to increase in the feedback delay. This would result in TCP senders backing off simulataneously and draining the queue repeatedly before it can become full again.

We now increase the shape parameter to 1.9, keeping the round trip time and the offered load fixed. We observe that increasing the shape parameter to 1.9 stabilises the system, see Fig. 17(c). A plausible explanation for this change in the qualitative behaviour of the system is as follows. As the shape parameter is increased from 1.5 to 1.9, the tail of the Pareto distribution becomes lighter. This implies that the traffic now constitutes more "mice" in the latter case. Hence, even if the long flows are synchronised and decreasing their sending rates, "mice" can arrive and depart the system without experiencing any drop. This leads to the effective utilisation of the available bandwidth and prevents the bottleneck queue from draining completely. This in turn ensures that there are no periodic oscillations in the queue size, and it is stable. However, with the shape parameter $\chi = 1.9$, if the delay gets too large, for example 500 ms, the system de-stabilises. This is because, in the congestion avoidance phase, the long flows would experience large delays before they can increase their sending rates after each packet drop. Hence, the feedback effects on the long flows would be more pronounced for larger round trip times. Additionally, a large feedback delay would affect the sending rates of the short flows. This would lead to the emergence of oscillations

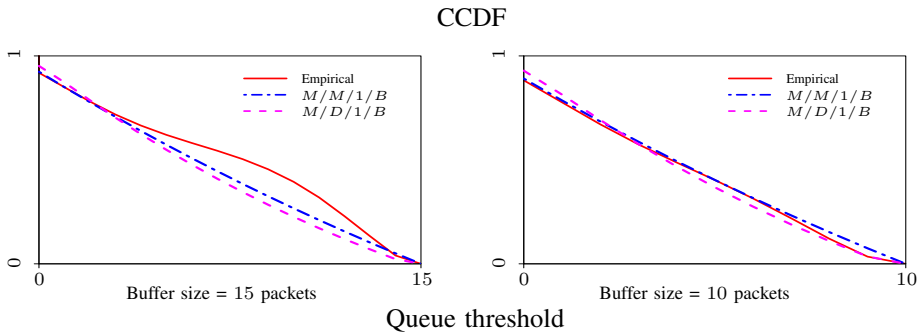


Fig. 18: *Compound TCP with heavy-tailed files.* 100 TCP sources each with an access link speed of 2 Mbps feeding into a core router of 100 Mbps. The round trip time is fixed at 200 ms. We plot the empirical queue distribution of the bottleneck queue for two buffer sizes $B = 15$ and $B = 10$ packets, for which the underlying dynamical system is stable, and compare it with that of $M/M/1/B$ and $M/D/1/B$ queues. We can observe that $M/M/1/B$ and $M/D/1/B$ approximations for the bottleneck queue seem reasonably justified.

in the queue size, as shown in 17(d). Note that the interdependence of the shape parameter and the round trip time to ensure stability of the system was indeed predicted by the necessary and sufficient condition, given by (39).

Impact of the expected file size: Figs. 17(b) and 17(c) also encapsulate the impact of the expected file size $\mathbb{E}(V)$ on stability. As we increase χ from 1.5 to 1.9, keeping the round trip time fixed at 200 ms, $\mathbb{E}(V)$ decreases from 100 kB to 70 kB, and the system stabilises. This is consistent with the insight obtained from our stability analysis, see Fig. 14.

Statistical properties: We now perform an empiricial study of the statistical properties of the bottleneck queue under this traffic scenario, as shown in Fig. 18. For this, we again consider 100 TCP senders feeding into the bottleneck router via access links with a speed of 2 Mbps. The expected file size is fixed at 100 kB, and the shape parameter is chosen to be 1.5.

We illustrate the queue distributions for two values of buffer sizes at the core router in the small buffer regime, 15 and 10 packets, for which the underlying dynamical system is stable. We can observe that for a buffer size of 15 packets, the queue distribution of the core router can be reasonably approximated by that of $M/M/1/B$ or an $M/D/1/B$ queue. An interesting observation is that this approximation holds remarkably well at a smaller buffer size of 10 packets. This strongly suggests that even with high variability at the source level, an $M/M/1/B$ approximation for the bottleneck queue is still valid in the asymptotic regime wherein, a large number of senders are present in the network, the bandwidth-delay product is high, and the buffer at the core router is dimensioned small enough to mitigate synchronisation effects. Further, from a theoretical perspective, this approximation seems benign, since packet-level simulations match our theoretical predictions well.

In the next section, we consider a multiple bottleneck topology which depicts a more realistic network scenario as opposed to the simple single bottleneck topology.

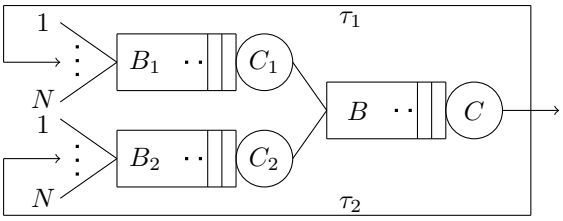


Fig. 19: Multiple bottleneck topology with two distinct sets of TCP flows, regulated by two edge routers, having round trip times τ_1 and τ_2 and feeding into a core router. The edge routers have link capacities C_1 and C_2, and buffer sizes B_1 and B_2 packets. The link capacity of the core router is C and the buffer size at the core router is B packets.

VI. Multiple bottlenecks

The model consists of two distinct sets of *many* TCP flows having different round trip times τ_1 and τ_2 and regulated by two edge routers, as shown in Fig. 19. For this model, our focus will be on long-lived flows. The average window sizes of the two sets of flows are $w_1(t)$ and $w_2(t)$ respectively. The outgoing flows from both edge routers feed into a common core router. The buffer sizes of the edge routers are B_1 and B_2 respectively, and buffer size of the core router is B. The link capacities of the edge routers are C_1 and C_2 respectively. Let C_1' and C_2' denote the service rates per flow for the edge routers. We consider the case where both edge routers and the core router have small buffer sizes and employ a Drop-Tail queue policy. The link capacity of the core router is C. The service rate per flow for the core router is denoted by C'. Suppose $p_1(t)$ and $p_2(t)$ are the packet loss probabilities at the two edge routers for the packets sent at time instant t, for the two distinct sets of flows respectively. The packet loss probability at the core router is denoted as $q(t, \tau_1, \tau_2)$. For a generalised TCP flavour, the non-linear, time-delayed, fluid model of the system is given by the following equations:

$$\begin{aligned}\dot{w}_j(t) =& \frac{w_j(t-\tau_j)}{\tau_j}\Bigg(i\left(w_j(t)\right)\Big(1-p_j(t-\tau_j)-q(t,\tau_1,\tau_2)\Big)\\ &- d\left(w_j(t)\right)\Big(p_j(t-\tau_j)+q(t,\tau_1,\tau_2)\Big)\Bigg), j=1,2. \end{aligned} \tag{53}$$

The loss probabilities at the three routers are approximated as

$$p_1(t) = \left(\frac{w_1(t)}{C_1'\tau_1}\right)^{B_1}, \; p_2(t) = \left(\frac{w_2(t)}{C_2'\tau_2}\right)^{B_2}, \text{ and}$$

$$q(t,\tau_1,\tau_2) = \left(\frac{w_1(t-\tau_1)/\tau_1 + w_2(t-\tau_2)/\tau_2}{\widetilde{C}}\right)^{B}.$$

Here, $\widetilde{C} = 2C'$. In this section, we prove that even in the multiple bottleneck topology, the system loses stability through a *Hopf bifurcation* [15] as system parameters vary. This loss of local stability leads to the emergence of limit cycles in the queue size of the core router.

A. Necessary and sufficient condition for stability

For system (53), we will perform a local stability analysis to derive a necessary and sufficient condition for stability. Suppose the equilibrium of the system is (w_1^*, w_2^*). Let $u_1(t) = w_1(t) - w_1^*$ and $u_2(t) = w_2(t) - w_2^*$ be small perturbations about w_1^* and w_2^* respectively. Linearising system (53) about its equilibrium (w_1^*, w_2^*), we get

$$\begin{aligned}\dot{u}_1(t) &= -\mathcal{M}_1 u_1(t) - \mathcal{N}_1 u_1(t-\tau_1) - \mathcal{P}_1 u_2(t-\tau_2),\\ \dot{u}_2(t) &= -\mathcal{M}_2 u_2(t) - \mathcal{N}_2 u_2(t-\tau_2) - \mathcal{P}_2 u_1(t-\tau_1),\end{aligned} \tag{54}$$

where, for Compound TCP, the increase and decrease functions (6) yield the following coefficients

$$\begin{aligned}\mathcal{M}_j &= -\frac{\alpha}{\tau_j}(k-2)\left(w_j^*\right)^{k-1}\left(1-\left(\frac{w_j^*}{C_j'\tau_j}\right)^{B_j} - \frac{1}{(2C')^B}\left(\frac{w_1^*}{\tau_1}+\frac{w_2^*}{\tau_2}\right)^B\right),\\ \mathcal{N}_j &= \left(\alpha\left(w_j^*\right)^{k-1}+\beta w_j^*\right)\left(\frac{B_j}{\tau_j}\left(\frac{w_j^*}{C_j'\tau_j}\right)^{B_j} + \frac{B\left(w_j^*\right)^2}{(2C')^B\tau_j^2}\left(\frac{w_1^*}{\tau_1}+\frac{w_2^*}{\tau_2}\right)^{B-1}\right),\\ \mathcal{P}_j &= \left(\alpha\left(w_j^*\right)^{k-1}+\beta w_j^*\right)\frac{Bw_j^*}{\tau_1\tau_2(2C')^B}\left(\frac{w_1^*}{\tau_1}+\frac{w_2^*}{\tau_2}\right)^{B-1}, \quad j=1,2.\end{aligned} \tag{55}$$

At equilibrium, the following equations are satisfied

$$\begin{aligned}&\alpha\left(w_j^*\right)^{k-1}\left(1-\left(\frac{w_j^*}{C_j'\tau_j}\right)^{B_j} - \frac{1}{(2C')^B}\left(\frac{w_1^*}{\tau_1}+\frac{w_2^*}{\tau_2}\right)^B\right)\\ &\qquad = \beta w_j^*\left(\left(\frac{w_j^*}{C_j'\tau_j}\right)^{B_j} + \frac{1}{(2C')^B}\left(\frac{w_1^*}{\tau_1}+\frac{w_2^*}{\tau_2}\right)^B\right), \ j=1,2.\end{aligned}$$

For analytical tractability, we consider two different scenarios with simple assumptions.

Case I

In this scenario, we assume that the network parameters for all routers are the same *i.e.* $B_1 = B_2 = B, C_1' = C_2' = C'$. We further assume that the round trip times of both sets of TCP flows are identical *i.e.* $\tau_1 = \tau_2 = \tau$. Then, $w_1^* = w_2^* = w^*$ will be an equilibrium of the system, and satisfies the following equation:

$$\alpha\left(w^*\right)^{k-2} = 2\left(\alpha\left(w^*\right)^{k-2}+\beta\right)\left(\frac{w^*}{C'\tau}\right)^B.$$

Let $\mathcal{M} = \frac{\left(\alpha(w^*)^{k-2}+\beta\right)B(w^*)^{B+1}}{\tau^{B+1}C'^B}$, then the coefficients $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{N}_1$, $\mathcal{N}_2$, $\mathcal{P}_1$, $\mathcal{P}_2$ reduce to

$$\begin{aligned}\mathcal{M}_1 = \mathcal{M}_2 &= \frac{2\mathcal{M}\beta w^*}{\left(\alpha\left(w_j^*\right)^{k-1}+\beta w_j^*\right)B}(2-k) = a,\\ \mathcal{N}_1 = \mathcal{N}_2 &= \frac{3}{2}\mathcal{M} = b,\\ \mathcal{P}_1 = \mathcal{P}_2 &= \frac{1}{2}\mathcal{M} = c.\end{aligned} \tag{56}$$

Note that a, b, $c > 0$. With these assumptions the linearised system (54) becomes

$$\dot{u}_1(t) = -au_1(t) - bu_1(t-\tau) - cu_2(t-\tau),$$
$$\dot{u}_2(t) = -au_2(t) - bu_2(t-\tau) - cu_1(t-\tau). \tag{57}$$

We now show that system (57) is stable if and only if the parameters a, b, c and τ satisfy the condition $\tau < \frac{1}{\omega_1}\cos^{-1}\left(\frac{-a}{b+c}\right)$ with crossover frequency $\omega_1 = \sqrt{(b+c)^2 - a^2}$.

Looking for exponential solutions, we get the characteristic equation for the linearised system (57) as

$$\left(s + a + be^{-s\tau}\right)^2 - c^2e^{-2s\tau} = 0, \tag{58}$$

which can be written as

$$g_1(s)\, g_2(s) = 0,$$

where,

$$g_1(s) = s + a + (b+c)\, e^{-s\tau}, \text{ and}$$
$$g_2(s) = s + a + (b-c)\, e^{-s\tau}. \tag{59}$$

For stability, all roots of (58) should have negative real parts. For negligible values of delay τ, system (57) is stable, *i.e.* all roots of the characteristic equation lie of the left half of the complex plane. As the delay is increased, the system becomes unstable if one pair of complex conjugate roots of either $g_1(s)$ or $g_2(s)$ or both crosses over the imaginary axis. We aim to determine the values of delay τ at which one pair of complex conjugate roots of $g_1(s)$ and $g_2(s)$ cross over the imaginary axis. Let $\tau_{1,c}$ and $\tau_{2,c}$ denote the values of τ at which $g_1(s)$ and $g_2(s)$ have exactly one pair of purely imaginary roots. Then, the critical value of τ, denoted by τ_c, at which (58) has one pair of purely imaginary roots is $\tau_c = \min(\tau_{1,c}, \tau_{2,c})$ [4]. Substituting $s = j\omega_1$ in $g_1(s)$ and separating real and imaginary parts we get

$$(b+c)\sin\omega_1\tau = \omega_1, \text{ and} \tag{60}$$
$$(b+c)\cos\omega_1\tau = -a. \tag{61}$$

Solving (60) and (61) for ω_1 we get

$$\omega_1 = \sqrt{(b+c)^2 - a^2}, \tag{62}$$

and under the condition $b + c > a$, ω_1^2 is strictly positive. This implies that there exists a cross over frequency ω_1 at which one pair of complex conjugate roots of $g_1(s)$ crosses over to the right half of the complex plane. Solving (60) and (61) for τ, we get the critical value of delay at which the system transits from stability to instability as

$$\tau_{1,c} = \frac{1}{\omega_1}\cos^{-1}\left(\frac{-a}{b+c}\right). \tag{63}$$

Similarly, substituting $s = j\omega_2$ in $g_2(s)$ we get the cross over frequency as

$$\omega_2 = \sqrt{(b-c)^2 - a^2}. \tag{64}$$

and under the condition $b-c>a$, ω_2^2 is strictly positive. Hence, there exists a cross over frequency ω_2 at which one pair of complex conjugate roots of $g_2(s)$ crosses over to the right half of the complex plane. Solving for τ, we obtain the critical value of the delay at which the system having the characteristic equation $g_2(s)$ has exactly one pair of purely imaginary roots as

$$\tau_{2,c} = \frac{1}{\omega_2}\cos^{-1}\left(\frac{-a}{b-c}\right). \tag{65}$$

Observe that $\omega_1 > \omega_2$. Since $\cos^{-1}(x)$ is monotonically decreasing for $x \in [-1,1]$, it can be shown that $\tau_1 < \tau_2$. This implies that $\tau_c = \tau_{1,c}$. Consequently, all roots of the characteristic equation (58) lie on the left half of the complex plane for all $\tau < \tau_c$. Hence, system (57) is asymptotically stable for $\tau < \tau_c$, and unstable for $\tau > \tau_c$. Further, we will analytically show that this loss of stability occurs via a Hopf bifurcation when one pair of complex conjugate roots of (58) crosses over the imaginary axis with non-zero velocity at $\tau = \tau_c$. Therefore, the necessary and sufficient condition for local stability of (57) is

$$\tau < \frac{1}{\omega_1}\cos^{-1}\left(\frac{-a}{b+c}\right). \tag{66}$$

Substituting values of ω_1, a, b and c in (66), we get the necessary and sufficient condition for local stability of (53), with Compound TCP as

$$\alpha\left(w^*\right)^{k-1}\sqrt{B^2-(k-2)^2\left(1-2p(w^*)\right)^2} < \cos^{-1}\left(\frac{(k-2)\left(1-2p(w^*)\right)}{B}\right), \tag{67}$$

where $p(w^*) = \left(\frac{w^*}{C'\tau}\right)^B$. This condition captures the relationship between the equilibrium window size, protocol parameters k and α, and buffer size B of the core router to ensure stability of the system.

We now derive a sufficient condition for local stability of system (53). To that end, we show that system (57) is stable if the parameters a, b, c and the feedback delay τ satisfy $\tau < \frac{\pi}{2(b+c)}$.

A sufficient condition for stability for a system with the characteristic equation $g_1(s) = 0$ is $(b+c)\,\tau < \frac{\pi}{2}$. Similarly, a sufficient condition for stability for a system with the characteristic equation $g_2(s) = 0$ is $(b-c)\,\tau < \frac{\pi}{2}$. Hence, a sufficient condition for system (57) to be asymptotically stable is

$$(b+c)\,\tau < \frac{\pi}{2}. \tag{68}$$

Substituting b and c in (68), a sufficient condition for stability of system (53) with Compound TCP flows is

$$\alpha B\left(w^*\right)^{k-1} < \frac{\pi}{2}. \tag{69}$$

We now show that a simple sufficient condition for local stability of system (53) is

$$\alpha B < \frac{\pi}{2}. \tag{70}$$

We note that at equilibrium, trivially $w^* \geq 1$. Since $k = 0.75$, it is easy to see that $h(w^*) = \left(w^*\right)^{k-1}$ is a decreasing function of w^*. This implies that $\alpha B\left(w^*\right)^{k-1} \leq \alpha B$, $\forall\, w^*$. Hence, if we ensure $\alpha B < \pi/2$, then local stability of (53) would be ensured. This yields a rather simple sufficient condition (70) for local stability of (53).

Discussion: Observe that condition (70) is independent of the equilibrium window size, and hence provides a *decentralised* design guideline for a network designer to dimension router buffers. Interestingly, this condition ensures that a network designer need not have the exact knowledge of the network parameters, such as the feedback delay and the capacity of the network, to dimension router buffers.

Case II

In this scenario, we assume that the network parameters for all routers are distinct, and the average round trip time of the first set of TCP flows is much larger than the other. Further, we consider that the average round trip time of the second set of TCP flows is negligible. This implies that $\tau_1 >> \tau_2$ and $\tau_2 \approx 0$. As a consequence of this assumption, the dynamics of the second set of TCP flows will appear to be almost instantaneous. This leads to the following non-linear, time-delayed fluid model of the system:

$$\begin{aligned}\dot{w}_1(t) =& \frac{w_1(t-\tau_1)}{\tau_1}\Bigg(i\left(w_1(t)\right)\Big(1-p_1(t-\tau_1)-q(t,\tau_1,\tau_2)\Big)\\ &- d\left(\left(w_1(t)\right)\Big(p_1(t-\tau_1)+q(t,\tau_1,\tau_2)\Big)\Bigg),\\ \dot{w}_2(t) =& \frac{w_2(t)}{\tau_2}\Bigg(i\left(w_2(t)\right)\Big(1-p_2(t)-q(t,\tau_1,\tau_2)\Big)\\ &- d\left(\left(w_2(t)\right)\Big(p_2(t)+q(t,\tau_1,\tau_2)\Big)\Bigg).\end{aligned} \tag{71}$$

The loss probabilities at the three routers are approximated as

$$p_1(t) = \left(\frac{w_1(t)}{\widetilde{C}_1\tau_1}\right)^{B_1}, \; p_2(t) = \left(\frac{w_2(t)}{\widetilde{C}_2\tau_2}\right)^{B_2}, \text{ and}$$
$$q(t,\tau_1,\tau_2) = \left(\frac{w_1(t-\tau_1)/\tau_1 + w_2(t)/\tau_2}{C'}\right)^{B}.$$

Here, $C' = 2\widetilde{C}$. We now outline local stability conditions for system (71). This will enable us to characterise the stability of the system in the presence of heterogeneity in network parameters. Suppose (w_1^*, w_2^*) be a non-trivial equilibrium of system (71). Let $u_1(t) = w_1(t) - w_1^*$ and $u_2(t) = w_2(t) - w_2^*$ represent small perturbations about w_1^* and w_2^* respectively. Linearising system (71) about its equilibrium, we get the following:

$$\begin{aligned}\dot{u}_1(t) &= -\mathcal{M}_1 u_1(t) - \mathcal{N}_1 u_1(t-\tau_1) - \mathcal{P}_1 u_2(t),\\ \dot{u}_2(t) &= -\big(\mathcal{M}_2 + \mathcal{N}_2\big)u_2(t) - \mathcal{P}_2 u_1(t-\tau_1),\end{aligned} \tag{72}$$

where, the coefficients are given by (55). Looking for exponential solutions, we get the characteristic equation for the linearised system (72) as

$$s^2 + as + bse^{-s\tau_1} + ce^{-s\tau_1} + d = 0, \tag{73}$$

where,

$$\begin{aligned}a &= \mathcal{M}_1 + \mathcal{M}_2 + \mathcal{N}_2\ , \qquad b = \mathcal{N}_1,\\ c &= \mathcal{N}_1\left(\mathcal{M}_2 + \mathcal{N}_2\right) - \mathcal{P}_1\mathcal{P}_2, \quad d = \mathcal{M}_1\left(\mathcal{M}_2 + \mathcal{N}_2\right).\end{aligned} \tag{74}$$

Result 1: The stability boundary of system (72), having the characteristic equation (73) is characterised by [13] $\tau_1 = \frac{1}{\omega}\cos^{-1}\left(\frac{\omega^2(d-ab)-cd}{b^2\omega^2+d}\right)$ with the cross over frequency as

$$\omega = \sqrt{\frac{(2c-a^2+b^2)}{2} + \frac{\sqrt{(2c-a^2+b^2)^2-4(c^2-d^2)}}{2}}.$$

We will show later that if this boundary condition just gets violated, the underlying dynamical system loses local stability via a Hopf type bifurcation. Hence, system (71) is locally stable if and only if $\tau_1 < \frac{1}{\omega}\cos^{-1}\left(\frac{\omega^2(d-ab)-cd}{b^2\omega^2+d}\right)$. Substituting a, b, c and d in the above would yield a condition which captures the interdependence among different network parameters and Compound TCP parameters to ensure stability of the system.

B. Hopf Condition

We have seen that protocol parameters, buffer thresholds and feedback delay all play an important role to ensure local stability. If the local stability conditions get violated, the system could transit from a locally stable to an unstable regime Varying any of the system parameters beyond the critical value can also drive the system to instability. Thus, instead of treating delay or any of the system parameters as the bifurcation parameter, we introduce an exogenous non-dimensional parameter κ which can act as the bifurcation parameter. If κ is varied keeping the values of the system parameters constant at their critical values, the system loses stability at $\kappa_c = 1$. To show that this loss of stability occurs via a Hopf bifurcation, we proceed to verify the transversality condition of the Hopf spectrum [17, Chapter 11, Theorem 1.1] for both scenarios. To verify the transversality condition, we need to show that $\mathrm{Re}(\mathrm{d}s/\mathrm{d}\kappa) \neq 0$ at $\kappa = \kappa_c$.

Case I

In this scenario, the linearised system, with the non-dimensional parameter κ, becomes

$$\begin{aligned}\dot{u}_1(t) &= \kappa\Big(-au_1(t) - bu_1(t-\tau) - cu_2(t-\tau)\Big),\\ \dot{u}_2(t) &= \kappa\Big(-au_2(t) - bu_2(t-\tau) - cu_1(t-\tau)\Big).\end{aligned} \tag{75}$$

Looking for exponential solutions of (75) we get

$$\left(s+\kappa a+\kappa\left(b+c\right)e^{-s\tau}\right)\left(s+\kappa a+\kappa\left(b-c\right)e^{-s\tau}\right) = 0. \tag{76}$$

Differentiating (76) with respect to κ, we get

$$\frac{\mathrm{d}s}{\mathrm{d}\kappa} = \frac{-\kappa a^2 - sa - sbe^{-s\tau} - 2\kappa abe^{-s\tau} - \kappa\left(b^2-c^2\right)e^{-2s\tau}}{s+\kappa a+\kappa be^{-s\tau} - s\kappa b\tau e^{-s\tau} - \kappa^2 ab\tau e^{-s\tau} - \kappa^2\tau\left(b^2-c^2\right)e^{-2s\tau}}. \tag{77}$$

From (76) we get,

$$e^{-s\tau} = -\frac{s+\kappa a}{\kappa\left(b+c\right)}. \tag{78}$$

Next, substituting (78) in (77), we get

$$\frac{\mathrm{d}s}{\mathrm{d}\kappa} = \frac{s}{\kappa\left(1 + s\tau + \kappa a\tau\right)}. \tag{79}$$

At $\tau = \tau_0$, $\kappa = \kappa_c$. Substituting $s = j\omega_1$ in (79) we get

$$\mathrm{Re}\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)_{s=j\omega_1} = \frac{\omega_1^2\tau_0}{\kappa_c\left(\left(1+\kappa_c a\tau_0\right)^2 + \left(\omega_1\tau_0\right)^2\right)} > 0.$$

In particular, we have proved that, $\mathrm{Re}(\mathrm{d}s/\mathrm{d}\kappa) > 0$, which implies that the roots cross over the imaginary axis with positive velocity at $\kappa = \kappa_c$.

Case II

For the second scenario, we observe that the linearised system, with the non-dimensional exogenous parameter κ is given as

$$\begin{aligned}
\dot{u}_1(t) &= \kappa\Big(\mathcal{M}_1 u_1(t) - \mathcal{N}_1 u_1(t-\tau_1) - \mathcal{P}_1 u_2(t-\tau_2)\Big),\\
\dot{u}_2(t) &= \kappa\Big(-\big(\mathcal{M}_2 + \mathcal{N}_2\big)u_2(t) - \mathcal{P}_2 u_1(t-\tau_1)\Big).
\end{aligned} \tag{80}$$

To show that system (80) loses stability via a Hopf bifurcation as the non-dimensional parameter κ is increased, we need to verify the transversality of the Hopf spectrum. Note that, for any complex number z, $\mathrm{Re}(z) \neq 0$ if and only if $\mathrm{Re}(z^{-1}) \neq 0$. Hence, for ease of analysis, we proceed to verify that $\mathrm{Re}(\mathrm{d}s/\mathrm{d}\kappa) \neq 0$ at $\kappa = \kappa_c$. Looking for exponential solutions of (80) leads us to the following characteristic equation:

$$s^2 + \kappa a s + \kappa b s e^{-s\tau_1} + \kappa^2 c e^{-s\tau_1} + \kappa^2 d = 0. \tag{81}$$

Differentiating (81) with respect to κ, we get

$$\frac{\mathrm{d}s}{\mathrm{d}\kappa} = \frac{-as - bse^{-s\tau_1} - 2\kappa c e^{-s\tau_1} - 2kd}{2s + \kappa a + \kappa b e^{-s\tau_1} - \kappa b s\tau_1 e^{-s\tau_1} - \kappa^2 c\tau_1 e^{-s\tau_1}} \tag{82}$$

From the characteristic equation (81), we get

$$e^{-s\tau_1} = -\frac{s^2 + \kappa a s + \kappa^2 d}{\kappa b s + \kappa^2 c}. \tag{83}$$

Now, substituting the value of $e^{-s\tau_1}$ in (82) and performing some algebraic manipulations, we obtain

$$\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)^{-1} = \left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)_1^{-1} + \left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)_2^{-1} + \left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)_3^{-1},$$

where,

$$\begin{aligned}
\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)_1^{-1} &= \frac{\kappa}{s}, \qquad \left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)_2^{-1} = \kappa\tau_1,\\
\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)_3^{-1} &= \frac{\kappa^2\big(s^2 ab\tau_1 - s^2 c\tau_1 + 2\kappa s b d\tau_1 + \kappa^2 cd\tau_1\big)}{s\Big(s^2 b + \kappa^2 ac + 2\kappa s c - \kappa^2 bd\Big)}.
\end{aligned} \tag{84}$$

Recall that, at the crossover point, the system has one pair of complex conjugate roots on the imaginary axis. Hence, substituting $s = j\omega$ in (84), we obtain $\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)^{-1}_{1,s=j\omega} = \frac{\kappa}{j\omega}$, which is purely imaginary. Similarly, we see that

$\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)^{-1}_{2,s=j\omega} = \kappa\tau_1$ which is strictly positive. Thus, to verify that $\mathrm{Re}\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)^{-1}_{s=j\omega} > 0$, verifying $\mathrm{Re}\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)^{-1}_{3,s=j\omega} > 0$ suffices. Now,

$$\mathrm{Re}\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)^{-1}_{3,s=j\omega} = \frac{2\omega^2\kappa^3\tau_1\big(abc - c^2 - b^2d\big)\big(\omega^2 + \kappa^2 d\big)}{4\omega^4\kappa^2c^2 + \big(\kappa^2\omega ac - \omega^3 b - \kappa^2\omega bd\big)^2}. \tag{85}$$

Recall that d is positive. Hence, the expression $\omega^2+\kappa^2 d$ is positive. Thus, it suffices to verify that $(abc-c^2-b^2d) > 0$. Substituting the values of a, d, c and d from (74), we get

$$abc - c^2 - b^2d = \mathcal{P}_1\mathcal{P}_2\left(\mathcal{N}_1\mathcal{N}_2 - \mathcal{P}_1\mathcal{P}_2\right) + \mathcal{N}_1\mathcal{P}_1\mathcal{P}_2\left(2\mathcal{M}_2 - \mathcal{M}_1\right).$$

Note that $\mathcal{M}_j$, $\mathcal{N}_j$ and $\mathcal{P}_j$ are strictly positive for $j = 1, 2$. Now, it can be easily concluded that $\mathcal{N}_1\mathcal{N}_2 > \mathcal{P}_1\mathcal{P}_2$. Hence, the first term in the above expression is positive. Recall that we consider a regime wherein the router buffers are small. Consequently, the average window sizes w_1^* and w_2^* would also be small. Additionally, we consider that the bandwidth-delay product is high. Hence, we assume that the per flow capacities of the edge routers C_1' and C_2' are large enough such that $1 - p_j^* - q^* \approx 1 \; \forall\, j = 1, 2$. With this approximation, in the regime wherein $\tau_1 \gg \tau_2$ and $\tau_2 \approx 0$, we can conclude that $\mathcal{M}_2 > \mathcal{M}_1$. This ensures that $abc - c^2 - b^2d > 0$. Hence, we can conclude that $\mathrm{Re}\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)^{-1}_{3,s=j\omega} > 0$, which in turn ensures that

$$\mathrm{Re}\left(\frac{\mathrm{d}s}{\mathrm{d}\kappa}\right)^{-1}_{\kappa=\kappa_c} > 0.$$

Thus, we observe that, the system undergoes a *Hopf Bifurcation* at $\kappa = \kappa_c$ for both scenarios. This implies that the system loses stability, as the system parameters vary, leading to the emergence of limit cycles. These limit cycles could in turn induce synchronisation among the Compound TCP flows which leads to periodic packet losses and loss in link utilisation. In turn, we expect the downstream traffic to be bursty.

C. Simulations

To validate our analytical insights, we simulate two scenarios in the multiple bottleneck topology: only long-lived flows, and with heavy-tailed files.

1) Dynamical Properties: The system consists of two distinct sets of 60 Compound TCP flows each with an access speed of 2 Mbps, regulated by two edge routers and feeding into a common core router. Each edge router has a link capacity of 100 Mbps and the core router has a link capacity of 197 Mbps.

To illustrate the impact of increasing buffer sizes on the queue size dynamics, we consider two cases: all routers have a buffer size of (i) 15 packets, and (ii) 100 packets. For the round trip times, we consider the following two cases: (i) both sets of flows have same average round trip times, 200 ms, and (ii) the average round trip time of one set is much smaller compared to the other. In this case, we choose the average round trip times as 10 ms and 200 ms respectively.

Figs. 20, and 21 show the queue size dynamics, with long-lived flows. It is evident that as the buffer thresholds are increased from 15 to 100 packets, limit cycles emerge in the queue size. In particular, even if one round trip time is large, the underlying dynamical systems lose stability if buffer sizes increase. This corroborates our analysis.

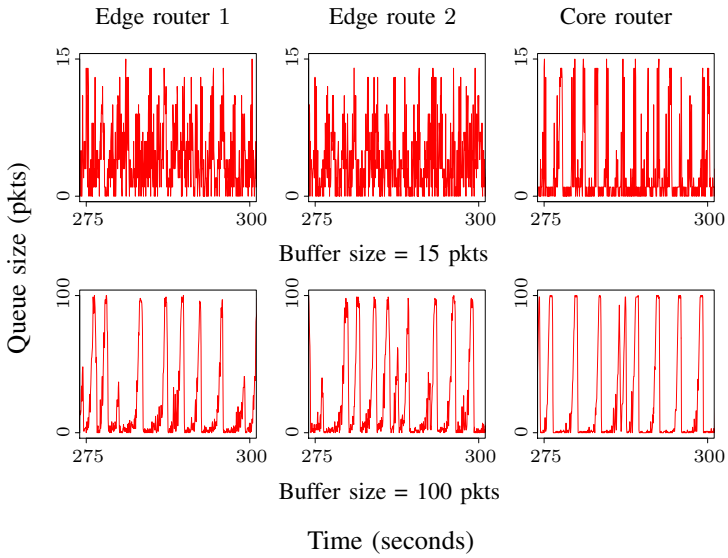


Fig. 20: *Queue size dynamics with long-lived flows:* Two sets of 60 Compound TCP flows each with an access speed of 2 Mbps, regulated by two edge routers each with a link capacity of 100 Mbps, and feeding into a core router with a link capacity of 197 Mbps. The flows in each set have an average round trip time of 200 ms. We can easily see that as buffer thresholds at routers are increased, the queue size exhibits limit cycles.

Fig. 22 shows the impact of increasing buffer sizes on the queue size dynamics, with heavy-tailed TCP connections. We can see that even with high variability at the connection level, increasing buffer thresholds would induce limit cycles in the queue size dynamics, an insight consistent with that obtained for a single bottleneck topology.

2) Statistical Properties: To establish the validity of our theoretical approximation that the packet loss probability at each bottleneck queue can be approximated by the corresponding blocking probability of an $M/M/1/B$ queue, we now empirically study the statistical properties of the arrival process, and the queue length distribution at each queue, for both topologies.

Note that in packet-level simulations, we can easily observe the loss of stability and hence a qualitative change in the dynamical properties of the system, with a reasonable number of TCP flows (60). However, for our statistical analyses, we need a larger number of long-lived flows for the statistical properties to hold. Hence, we consider two sets of 480 flows, each over an access link with a speed of 0.25 Mbps. Each set of flows is regulated by an edge router with a link capacity of 100 Mbps. The outputs of the edge routers feed into a core router with a link capacity of 194 Mbps. We choose the average round trip time of each set of flows as 300 ms.

Statistics of the arrival process: We first conduct an empirical study on the statistical properties of the traffic arrival at each bottleneck queue in a similar spirit as done for the single bottleneck topology. Specifically, we measure the burstiness of the arrival process at each queue in terms of their coefficient of variation at different time scales.

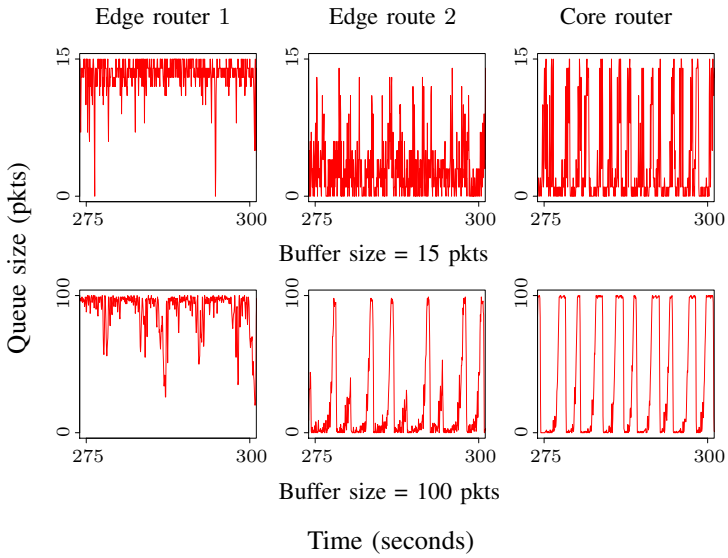


Fig. 21: *Queue size dynamics with long-lived flows:* Two sets of 60 Compound TCP flows each with an access speed of 2 Mbps, regulated by two edge routers each with a link capacity of 100 Mbps, and feeding into a core router with a link capacity of 197 Mbps. The flows in the two sets have average round trip times 10 ms and 200 ms respectively. It can be easily noted that as buffer sizes are increased from 15 to 100 packets, the queue size dynamics exhibits limit cycles.

For the empirical study, we consider three representative scenarios. In the first scenario, the buffer sizes at all routers are fixed at 15 packets, which ensures that the underlying dynamical system is stable. In the second scenario, all buffers are dimensioned at 50 and 100 packets respectively. In both these cases, the system dynamics exhibits limit cycles, and synchronisation among TCP windows. In the third scenario, the buffer sizes at all routers are chosen according to the bandwidth-delay product rule, which leads to 2084 packets at the edge routers, and 4040 packets at the core router. Since we are interested in measuring the burstiness of the arrival process at short time scales, we aggregate the arrival traffic over time scales ranging from $2^{12}\,\mu s = 4$ ms to $2^{20}\,\mu s = 1$ second.

Fig. 23 depicts the coefficient of variation curves at each router, at various time scales, for this topology. We can easily observe that similar qualitative insights obtained in the single bottleneck topology carry forward to the multiple bottleneck topology also. In particular, when all buffers are dimensioned at 15 packets, the coefficient of variation curves exhibits a relatively faster decay as the aggregation increases, as opposed to larger buffer thresholds. Further, we can observe that for larger buffer thresholds, the coefficient of variation curves for the traffic arrival at each queue flattens significantly at larger time scales. This indicates that larger buffers maintain higher variability or burstiness in the traffic arrival, in the presence of synchronisation. On the contrary, in the absence of synchronisation, we can observe reduced variability or burstiness in the traffic arrival at each queue. This suggests that the aggregate traffic arrival behaves qualitatively similar to short range dependent processes, when buffers are sized small enough

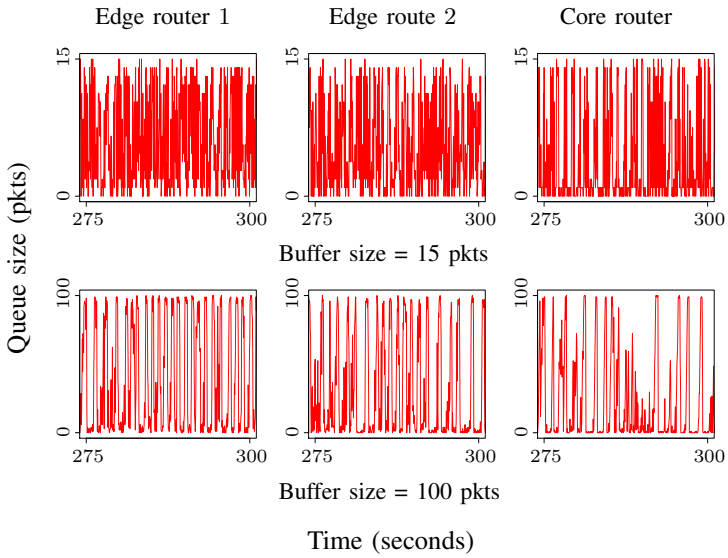


Fig. 22: *Queue size dynamics with heavy-tailed files:* Two sets of 100 Compound TCP sources, regulated by two edge routers each with a link capacity of 100 Mbps, and feeding into a core router with a link capacity of 197 Mbps. The flows in both sets have an average round trip time of 200 ms. The file sizes are drawn from a Pareto distribution with shape parameter 1.5. The average file size is fixed at 100 KB. Observe that the queue size exhibits limit cycles, as buffer sizes are increased from 15 to 100 packets.

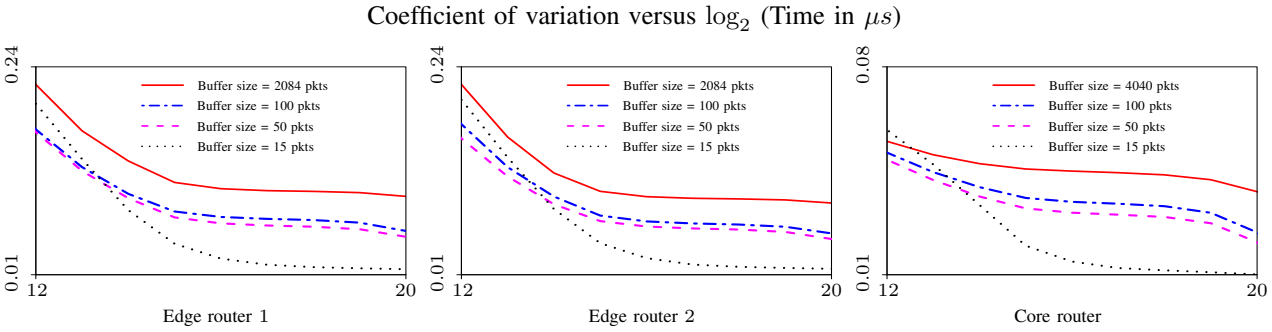


Fig. 23: *Statistics of the arrival process.* Two sets of 480 long-lived Compound TCP flows each with an access link speed of 0.25 Mbps, regulated by two edge routers and feeding into a core router of 194 Mbps. The average round trip of each set is chosen as 300 ms. We consider three representative regimes: (i) stable (each router has a buffer size of 15 packets, (ii) presence of synchronisation (each router has a buffer size of 50 and 100 packets respectively) and (iii) each router follows the bandwidth-delay product rule, used in practice. Observe that with smaller buffers (15 packets), the aggregate arrival process at each queue exhibits reduced burstiness.

to mitigate synchronisation effects. Hence, the approximation that the aggregate traffic arrival at each bottleneck queue is Poisson in the presence of a large number of TCP flows seems reasonable, in the regime considered.

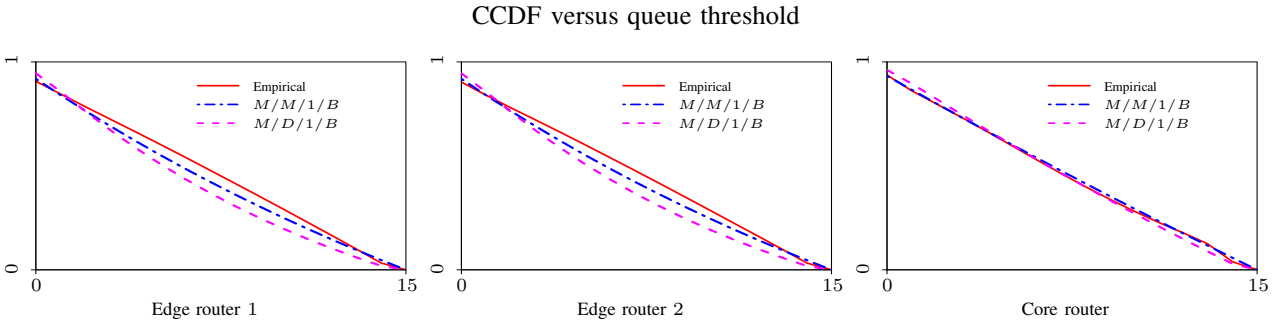


Fig. 24: *Statistics of the queue size.* Two sets of 480 long-lived Compound TCP flows each with an access link speed of 0.25 Mbps, regulated by two edge routers and feeding into a core router of 194 Mbps. The buffer size at each router is fixed at 15 packets, and the average round trip of each set is chosen as 300 ms. We compare the queue distribution at each queue with that of $M/M/1/B$ and $M/D/1/B$ queues. Observe that either an $M/M/1/B$ or $M/D/1/B$ approximation seems reasonable with a large number of long-lived flows and high bandwidth-delay product.

Statistics of the queue size: We now perform a comparative study on the queue size distribution of each bottleneck queue in each topology, with that of an $M/M/1/B$ and an $M/D/1/B$ queue. For our study, we consider the regime when all buffers are dimensioned at 15 packets. This is because, we have already established that only in this regime, the arrival process at each queue can be reasonably approximated by a Poisson process.

As shown in Fig. 24,for a buffer size of 15 packets, the queue distribution at each queue can be well approximated by that of an $M/M/1/B$ or an $M/D/1/B$ queue, in this topology. This strongly suggests that even with TCP controlled flows in a multiple bottleneck topology, each bottleneck queue can be approximated as either an $M/M/1/B$ or an $M/D/1/B$ queue, thus validating our modelling assumptions, in the asymptotic regime considered in this work.

VII. Impact of buffer sizing on system performance

Through a combination of stability analysis and extensive packet-level simulations, we have highlighted smaller buffer thresholds play an important role to ensure stability. Hence, it is imperative to study the impact of such small buffers on the system performance. To that end, we consider the multiple bottleneck topology, and choose two performance metrics, throughput and flow completion time. We conduct packet-level simulations for the same. Note that for a single bottleneck topology also, we observe similar qualitative behaviour.

A. Throughput

For this, we consider two sets of 60 long-lived Compound TCP flows, each over a 2 Mbps access link. Both sets of flows are regulated by two edge router with a link capacity of 100 Mbps. The outputs of the edge routers feed into a core router with a link capacity of 197 Mbps. To study the impact of buffer sizing on throughput, we fix the

buffer sizes at the edge routers at 15 packets, and vary the buffer size at the core router from 5 to 300 packets. We fix the average round trip time of each set of flows as 200 ms. Fig. 25 shows the variation in throughput as the buffer size at the core router varies. It is evident that even with smaller buffers, we achieve fairly good throughput.

B. Average Flow Completion Time (AFCT)

While throughput is undoubtedly an important performance metric from a network operator point of view, flow completion time is more important from a user perspective [8]. In particular, users would want their flows to complete in the shortest time possible. Hence, if buffer sizes are made much smaller that what they are today (bandwidth-delay product rule), they should not degrade flow completion times significantly.

For this, we consider two sets of 100 Compound TCP sources, regulated by two edge routers, each with a link capacity of 100 Mbps. The outputs of the edge routers feed into a core router with a link capacity of 197 Mbps. Each TCP source is connected to an edge router via an access link with a speed of 2 Mbps. Further, each TCP source performs successive transfer of files according to a Poisson process, and file sizes are drawn from a Pareto distribution. The expected file size is 100 KB and the shape parameter is 1.5. We consider the average duration between each transfer to be 0.1 seconds. To study the impact of buffer sizing on flow completion times, we consider two cases: (i) the buffer size at each router is fixed at 15 packets, and (ii) the buffer size at each router follows the bandwidth-delay product rule. With this, the buffer size is 2084 packets at each router, and 4100 packets at the core router. Fig. 26 shows the AFCT for these two cases. We can observe that smaller buffers yield comparable AFCTs as that with bandwidth-delay worth of buffering. Hence, it is indeed possible to significantly reduce buffer sizes without affecting flow completion times.

In summary, smaller buffers ensure stability without degrading the system performance.

VIII. Conclusions

In this paper, we conducted a performance evaluation of Compound TCP in two different topologies, with Drop-Tail queues, in a small buffer regime. For the traffic, we considered three scenarios. In the first scenario, we assume that only long-lived flows are present in the system. The second scenario constitutes a combination of long and short flows. The third scenario aims to capture the high variability present in real Internet traffic and considers heavy-tailed connections at the source level. Using a combination of analysis and packet-level simulations, we explored numerous dynamical and some statistical properties to obtain a few key insights.

From a *dynamical* perspective, we emphasised the interplay between buffer sizes and stability. In particular, we showed that smaller buffers tend to favour stability. On the other hand, larger buffer thresholds may help link utilisation, but they would increase queuing delay and are also prone to inducing limit cycles, via a Hopf bifurcation, in the queue size dynamics. However, such limit cycles can in turn lead to a drop in link utilisation, induce synchronisation among TCP flows, and make the downstream traffic bursty. We also noted that when a network has high variability in terms of the connections generated at the source level, such limit cycles continue to exist in the queue size despite the heterogenity in the incoming traffic. Some design considerations for protocol

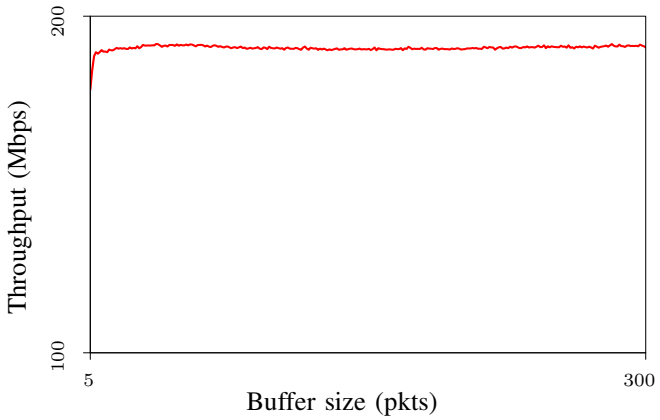


Fig. 25: *Impact of buffer sizing on throughput*. Two sets of 60 long-lived Compound TCP flows, regulated by two edge routers each with a link capacity of 100 Mbps. The outputs of the edge routers feed into a core router with a link capacity of 197 Mbps. The round trip time of each set is 200 ms. We vary the core router buffer size. Observe that smaller buffers do not degrade throughput significantly.

and network parameters, to ensure stability and low-latency queues, are also outlined. We repeatedly witnessed the existence of limit cycles in the queue size. To that end, it would be of both theoretical and practical interest, to establish the *asymptotic orbital stability* of the bifurcating limit cycles and to the determine the *type* of the Hopf bifurcation. One way to approach this analytically would be via the theory of normal forms and the centre manifold analysis [15], and such an analysis is conducted in [13].

In terms of the *statistical* properties, we examined the arrival process and the empirical queue distribution at the bottleneck queues, both in single and multiple bottleneck topologies. We observed that in the regime considered, each bottleneck queue can be well approximated by an $M/M/1/B$ or an $M/D/1/B$ queue, in the absence of synchronisation. Further, this approximation holds reasonably well even in the presence of high variability at the TCP connection level. This makes our system models amenable to analysis, and thus gives confidence in using the underlying models to better understand network performance and quality of service.

In summary, our work recommends that buffer sizes at routers should be significantly reduced to ensure stability as well as low latency. We showed that design of such small buffers is indeed possible without compromising the system performance, namely, throughput and flow completion times.

The insights obtained in this thesis could have important consequences for the modelling and the performance evaluation of communication networks. From a theoretical perspective, this opens many challenging questions centered around the development of accurate fluid models for different versions of TCP and queue management policies, and their interaction with different buffer sizing strategies. From a practical perspective, of immediate

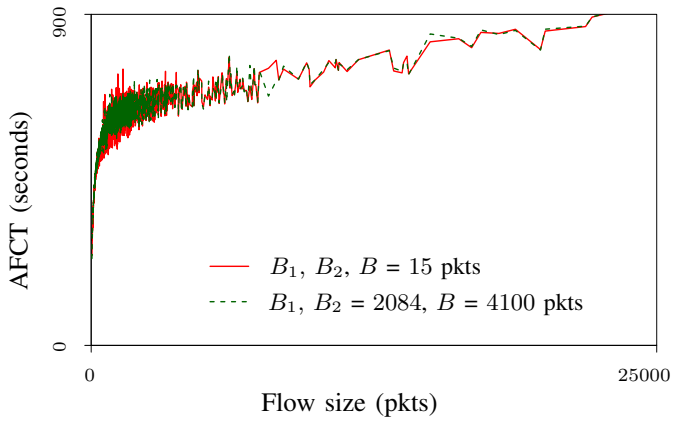


Fig. 26: *Impact of buffer sizing on AFCT*. Two sets of 100 Compound TCP sources, regulated by two edge routers each with a link capacity of 100 Mbps. The outputs of the edge routers feed into a core router with a link capacity of 197 Mbps. The round trip time of each set is 200 ms. The file sizes are drawn from a Pareto distribution with mean 100 KB and shape 1.5. We consider two cases: (i) buffer sizes at all routers are fixed at 15 packets, (ii) buffer sizes at all routers follow the bandwidth-delay product rule, used in practice. Observe that smaller buffers yield comparable AFCT as bandwidth-delay product worth of buffering.

interest would be to understand buffer sizing requirements with CUBIC TCP, which is the default protocol in the Linux OS.

REFERENCES

REFERENCES

[1] B.C. Arnold, "Pareto Distributions", *Chapman and Hall/CRC, Second Edition*, 2015.

[2] U. Ayesta, K.E. Avrachenkov, E. Altaman, C. Barakat and P. Dube, "Multilevel approach for modeling short TCP sessions", *Teletraffic Science and Engineering, Elsevier*, vol. 5, pp. 661–670, 2003.

[3] L.S. Brakmo and L.L. Peterson, "TCP Vegas: end to end congestion avoidance on a global Internet", *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1465–1480, 1995.

[4] S.A. Campbell, S. Ruan and J. Wei, "Qualitative analysis of a neural network model with multiple time delays", *International Journal of Bifurcation and Chaos*, vol. 9, pp. 1585–1595, 1999.

[5] J. Cao and K. Ramanan, "A Poisson limit for buffer overflow probabilities", in *Proceedings of IEEE INFOCOM*, 2002.

[6] V.G. Cerf, "Bufferbloat and other Internet challenges", *IEEE Internet Computing*, vol. 5, pp. 79–80, 2014.

[7] A. Dhamdhere, H. Jiang and C. Dovrolis, "Buffer sizing for congested internet links", in *Proceedings of IEEE INFOCOM*, 2005.

[8] N. Dukkipati and N. McKeown, "Why flow-completion time is the right metric for congestion control", *0ACM SIGCOMM Computer Communication Review*, vol. 36, pp.59–62, 2006.

[9] K. Engelborghs, T. Luzyanina and D. Roose, "Numerical bifurcation analysis of delay differential equations using DDE-BIFTOOL", *ACM Transactions on Mathematical Software (TOMS)*, vol. 28, pp. 1–21, 2002.

[10] S. Floyd and V. Jacobson, "Random Early Detection gateways for congestion avoidance", *IEEE/ACM Transactions on Networking*, vol. 1, pp. 397–413, 1993.

[11] S. Floyd, "HighSpeed TCP for large congestion windows", RFC 3649, December 2003.

[12] J. Gettys and K. Nichols, "Bufferbloat: dark buffers in the Internet", *Communications of the ACM*, vol. 55, pp. 57–65, 2012.

[13] D. Ghosh, K. Jagannathan and G. Raina, "Local stability and Hopf bifurcation analysis for Compound TCP", *IEEE Transactions of Control of Network Systems*, vol. 5, pp. 1668–1681, 2018.

[14] D. Ghosh, K. Jagannathan and G. Raina, "Right buffer sizing matters: stability, queuing delay and traffic burstiness in compound TCP", in *Proceedings of 52nd Annual Allerton Conference on Communication, Control, and Computing*, 2014.

[15] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, 1996.

[16] S. Ha, I. Rhee and L. Xu, "CUBIC: a new TCP-friendly high-speed TCP variant", *ACM SIGOPS Operating Systems Review*, vol. 42, pp. 64–74, 2008.

[17] J.K. Hale and S.M.V. Lunel, *Introduction to functional differential equations*, Springer Science & Business Media, 2013.

[18] B.D. Hassard, N.D. Kazarinoff and Y-H. Wan, *Theory and Applications of Hopf Bifurcation*, Cambridge University Press, 1981.

[19] C.V. Hollot, V. Misra, D. Towsley and W.B. Gong, "A control theoretic analysis of RED", in *Proceedings of IEEE INFOCOM*, 2001.

[20] Y. Joo, V. Ribeiro, A. Feldmann, A.C. Gilbert and W. Willinger, "TCP/IP traffic dynamics and network performance: A lesson in workload modeling, flow control, and trace-driven simulations", *ACM SIGCOMM Computer Communication Review*, vol. 31, pp. 25–37, 2001.

[21] V. Kharitonov and D. Melchor–Aguilar,"On delay–dependent stability conditions", *Systems & Controls Letters*, vol. 15, pp. 71–76, 2000.

[22] F.P. Kelly, "Models for a self-managed Internet", *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 358, pp. 2335–2348, 2000.

[23] K. Nichols and V. Jacobson, "Controlling queue delay", *Communications of the ACM*, vol. 55, pp. 42–50, 2012.

[24] J. Padhye, V. Firoiu, D. Towsley and J.F. Kurose, "Modeling TCP Reno performance: a simple model and its empirical validation", *IEEE/ACM Transactions on Networking*, vol. 8, pp. 133–145, 2000.

[25] R. Pan, P. Natarajan, C. Piglione, M.S. Prabhu, V. Subramanian, F. Baker and B. VerSteeg, "PIE: a lightweight control scheme to address the bufferbloat problem", in *Proceedings of 14th International Conference on High Performance Switching and Routing*, 2013.

[26] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling", *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226–244, 1995.

[27] G. Raina, "Local bifurcation analysis of some dual congestion control algorithms", *IEEE Transactions on Automatic Control*, vol. 50, pp. 1135–1146, 2005.

[28] G. Raina, S. Manjunath, S. Prasad and K. Giridhar, "Stability and performance analysis of Compound TCP With REM and Drop-Tail queue management", *IEEE/ACM Transactions on Networking*, vol. 24, pp. 1961–1974, 2016.

[29] G. Raina and D. Wischik, "Buffer sizes for large multiplexers: TCP queueing theory and instability analysis", in *Proceedings of Next Generation Internet Networks*, 2005.

[30] P. Raja and G. Raina, "Delay and loss-based transport protocols: buffer-sizing and stability", in *Proceedings of International Conference on Communication Systems and Networks*, 2012.

[31] K. Tan, J. Song, Q. Zhang and M. Sridharan, "A Compound TCP approach for high-speed and long distance networks", in *Proceedings of IEEE INFOCOM*, 2006.

[32] C. Villamizar and C. Song, "High performance TCP in ANSNET", *ACM SIGCOMM Computer Communication Review*, vol. 24, pp. 45-60, 1994.

[33] A. Vishwanath, V. Sivaraman and G.N. Rouskas, "Anomalous loss performance for mixed real-time and TCP traffic in routers with very small buffers", *IEEE/ACM Transactions on Networking*, vol. 19, pp. 933–946, 2011.

[34] W. Willinger, M.S. Taqqu, R. Sherman and D.V Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level", *IEEE/ACM Transactions on Networking*, vol. 5, pp. 71–86, 1997.

[35] D. Wischik and N. McKeown, "Part I: Buffer sizes for core routers", *ACM Computer Communication Review*, vol. 35, pp. 75–78, 2005.

[36] P. Yang, J. Shao, W. Luo, L. Xu, J. Deogun and Y. Lu, "TCP congestion avoidance algorithm identification", *IEEE/ACM Transactions on Networking*, vol. 22, pp. 1311–1324, 2014.

[37] The Network Simulator (NS2). [Online]. Available: http://nsnam.isi.edu/nsnam/index.php/User Information.