An Efficient Distributed Congestion Control Protocol with Stability and Fairness

Vikram A
Assistant Professor,
Department of Computer Science and Engineering,
Saranathan College of Engineering,
Panjappur, Tiruchirappalli, Tamil Nadu-620 012.
vikram.aug1984@gmail.com

Abstract:

Recent research efforts to design better Internet transport protocols combined with scalable Active Queue Management (AQM) have led to significant advances in congestion control. One of the hottest topics in this area is the design of discrete congestion control algorithms that are asymptotically stable under heterogeneous feedback delay and whose control equations do not explicitly depend on the RTTs of end-flows. In this paper, we first prove that single-link congestion control methods with a stable radial Jacobian remain stable under arbitrary feedback delay (including heterogeneous directional delays) and that the stability condition of such methods does not involve any of the delays. We then extend this result to generic networks with fixed consistent bottleneck assignments and max-min network feedback. To demonstrate the practicality of the obtained result, we change the original controller in Kelly’s work [15] to become robust under random feedback delay and fixed constants of the control equation. We call the resulting framework Max-min Kelly Control (MKC) and show that it offers smooth sending rate, exponential convergence to efficiency, and fast convergence to fairness, all of which make it appealing for future high-speed networks.

Keywords: AQM. MKC, EMKC, High Speed Networks.

I. Introduction

Over the last fifteen years, Internet congestion control has evolved from binary-feedback methods of AIMD/TCP [2], [33] to the more exciting developments based on optimization theory [22], [23], game theory [11], [19], and control theory [10], [11], [24], [26].

It is widely recognized that TCP’s congestion control in its current shape is inadequate for very high-speed networks and fluctuation-sensitive real time multimedia. Thus, a significant research effort is currently under way (e.g., [5], [6], [9], [12], [15], [16], [19], [29], [32]) to better understand the desirable properties of congestion control and develop new algorithms that can be deployed in future AQM (Active Queue Management) networks.

One of the most important factors in the design of congestion control is its asymptotic stability, which is the capability of the protocol to avoid oscillations in the steady-state and properly respond to external perturbations caused by the arrival/departure of flows, variation in feedback, and other transient effects. Stability proofs for distributed congestion control become progressively more complicated as feedback delays are taken into account, which is especially true for the case of heterogeneous delays where each user i receives its network feedback delayed by a random amount of time Di. D and do not take into account the fact that end-users in real networks are rarely (if ever) synchronized. Several recent studies [20], [24], [27] successfully deal with heterogeneous delays; however, they model Di as a deterministic metric and require that end-flows (and sometimes routers) dynamically adapt their equations based on feedback delays, which potentially leads to RTT-unfairness, increased overhead, and other side-effects (such as probabilistic stability).

In this paper, we set our goal to build a discrete congestion control system that maintains both stability and fairness under heterogeneously delayed feedback, allows users to use fixed parameters of the control equation, and admits a low-overhead implementation inside routers. We solve this problem by showing that any single-link max-min fair system with a stable radial Jacobian remains asymptotically stable under arbitrary directional delays, extend this result to multilink networks under fixed bottleneck assignments, and apply it to the original controller proposed by Kelly et al. [15].

We call the result of these efforts Max-min Kelly Control (MKC) and demonstrate that its stability and fairness do not depend on any parameters of the network (such as delay, path length, or the routing matrix of end-users). We also show that with a proper choice of AQM feedback, MKC converges to efficiency exponentially fast, exhibits stability and fairness under random delays, converges to fairness almost as quickly as AIMD, and does not require routers to estimate any parameters of individual flows.
By isolating bottlenecks along each path and responding only to the most-congested resource, the MKC framework allows for very simple stability proofs, which we hope will lead to a better understanding of Kelly’s framework in the systems community and eventually result in an actual implementation of these methods in real networks.

Our initial thrust in this direction includes ns2 simulations of MKC, which show that finite time-averaging of flow rates inside each router coupled with a naive implementation of end-user functions leads to undesirable transient oscillations, which become more pronounced when directional delays $D_i$ and $D_j$ to/from each router increase. We overcome this drawback with simple changes at each end-user and confirm that the theoretically predicted monotonic convergence of MKC is achievable in real networks, even when the routers do not know the exact combined rate of end-flows at any time instant $n$. We also show that our algorithms inside the router incur low overhead (which is less than that in XCP [12] or RED [7]) and require only one addition per arriving packet and two variables per router queue.

II. Background

A. Delay-Dependent Congestion Control

Recently, a large amount of theoretical and experimental work has been conducted on designing robust congestion controls. One direction is to model the network from an optimization or game-theoretic point of view [11], [17], [18], [19], [23]. The original work by Kelly et al. [14], [15] offers an economic interpretation of the resource-user model, in which the entire system achieves its optimal performance by maximizing the individual utility of each end-user. To implement this model in a decentralized network, Kelly et al. describe two algorithms (primal and dual) and prove their global stability in the absence of feedback delay.

However, if feedback delay is present in the control loop, stability analysis of Kelly controls is non-trivial and currently forms an active research area [4], [10], [20], [24], [27], [29]. Recall that in Kelly’s framework [15], [24], each user $i \in [1,N]$ is given a unique route $r_i$ that consists of one or more network resources (routers). Feedback delays in the network are heterogeneous and directional. The forward and backward delays between user $i$ and resource $j$ are denoted by $D_{ij}$ and $D_{ji}$, respectively. Thus, the round-trip delay of user $i$ is the summation of its forward and backward delays with respect to any router $j \in r_i; D_i = D_{ij} + D_{ji}$.

B. Delay-Independent Congestion Control

To the best of our knowledge, the first delay-independent stability condition is due to Vinnicombe, who proposes and examines the following continuous fluid model of a network with sources operating TCP-like algorithms [28]:

$$x_i(t) = \frac{x_i(t-D_i)}{D_i}$$(1)

where $a_i(t)=a(x_i(t-D_i))$, $b_i(t)=b(x_i(t-D_i))$, $a$, $b$, $m$, $n$ are constants, $\eta_i(t)$ is the network feedback and link price $p_i(t)=y_i(t/C_i)$ is an approximation of packet loss at link $j$ of capacity $C_j$ and buffer size $B$.

III. Classic Kelly Control

In this we discuss intuitive examples that explain the cryptic formulas in the previous section and demonstrate in simulation how delays affect stability of Kelly controls (1). We then show that the original Kelly control [15], or any mechanism that relies on the sum of feedback functions from individual routers, exhibits a tradeoff between linear convergence to efficiency and persistent stationary packet loss.

A. Delayed Stability Example

The following example illustrates stability problems of (1) when feedback delays are large. We assume a single-source, single-link configuration and utilize a congestion indication function that computes the estimated packet loss using instantaneous arrival rates:

$$p(n)=\frac{x(n)-C}{x(n)}$$ (2)

Where $C$ is the link capacity and $x(n)$ is the flow rate at discrete step $n$. We note that the price function $p(n)$ in the original Kelly control is nonnegative; this results in slow linear AIMD-like probing for link capacity until the slowest link in the path is fully utilized, which is generally considered too slow for high-speed networks. Thus, under AQM feedback assumed throughout this paper, we allow negative values in which signals the flow to increase its sending rate when $x(n) < C$. We show that the negative component of packet-loss improves convergence to efficiency from linear to exponential.

Applying $p(n)$ to Kelly control yields a linear end-flow equation:

$$x(n)=x(n-1)+kw-k(x(n-D)-C)$$ (3)

Next, assume a particular set of parameters: $K=1=2$, $\omega=10$mb/s, and $C=1,000$ mb/s. We have that the system is stable if and only if delay $D$ is less than four time units. As illustrated in Figure 1(a), delay $D=1$ keeps the system stable and monotonically convergent to its stationary point.

Figure 1: Stability of Kelly control under different feedback delays ($k=1/2,\omega=10$ mb/s, and $C=1,000$ mb/s)
Under larger delays $D = 2$ and $D = 3$ in Figures 1(b) and (c), the flow exhibits progressively increasing oscillations before entering the steady state. Eventually, as soon as $D$ becomes equal to four time units, the system diverges as shown in Figure 1(d).

Using the same parameter $\cdot$ and reducing $\omega$ to 20 kb/s, we examine ns2 simulations, in which a single flow passes through a link of capacity 50 mb/s. We run the flow in two network configurations with the round-trip delay equal to 90 ms and 120 ms, respectively. As seen in Figure 2, the first flow reaches its steady state after decaying oscillations, while the second flow exhibits no convergence and periodically overshoots capacity $C$ by 200%. Since Kelly controls are unstable unless condition is satisfied, a natural strategy to maintain stability is for each end-user $i$ to adaptively adjust its gain parameter $K_i \sim 1/D_i$. However, this method depends on reliable estimation of round-trip delays $D_i$ and leads to unfairness between the flows with different RTTs.

**B. Stationary Rate Allocation**

In this the price function should allow negative values, such that the convergence speed of Kelly control is improved from linear to exponential. However, we show next that this modification presents a problem in the stationary resource allocation. Consider a network of $M$ resources and $N$ homogeneous users (i.e., with the same parameters $K$ and $\omega$). Further assume that resource $j$ has capacity $C_j$, user $i$ utilizes route $r_i$ of length $M_i$ (i.e., $M_i = |r_i|$), and packet-loss $\eta(n)$ fed back to user $i$ is the aggregate feedback from all resources in path $r_i$. We further assume that there is no redundancy in the network (i.e., each user sends its packets through at least one resource and all resources are utilized by at least one user). Thus, we can define routing matrix $A_{N \times M}$ such that $A_{ij} = 1$ if user $i$ passes through resource $j$ (i.e., $j \in r_i$) and $A_{ij} = 0$ otherwise. Further denote the $j$-th column of $A$ by vector $V_j$. Clearly, $V_j$ identifies the set $s_j$ of flows passing through router $j$.

**IV. Max-min Kelly Control**

We start our discussion with the following observations. First, we notice that in the classic Kelly control (1), the enduser decides its current rate $x_i$ based on the most recent rate $x_i(n-1)$ and delayed feedback $\eta(n-D_i)$. Since the latter carries information about $x_i(n-D_i)$, which was in effect RTT time units earlier, the controller in (1) has no reason to involve $x_i(n-1)$ in its control loop. Thus, the sender quickly becomes unstable as the discrepancy between $x_i(n-1)$ and $x_i(n-D_i)$ increases. One natural remedy to this problem is to retard the reference rate to become $x_i(n-D_i)$ instead of $x_i(n-1)$ and allow the feedback to accurately reflect network conditions with respect to the first term of (1).

Second, to avoid unfairness between flows, one must fix the control parameters of all end-users and establish a uniform set of equations that govern the system. Thus, we create a new notation in which $K_{\text{det}} = \alpha$, $K_i = \beta$ and discretize Kelly control as following:

$$x_i(n) = x_i(n-D_i) + \alpha \cdot \eta_i(n-D_i)$$

where $\eta_i(n)$ is the congestion indication function of user $i$.

Next, we overcome the problems of proportional fairness described in the previous section and utilize negative network feedback, we combine with max-min fairness under which the routers only feed back the packet loss of the most-congested resource instead of the combined packet loss of all links in the path:

$$\eta_i(n) = \max_j \eta_j(n-D_i)$$

where $\eta_j(n)$ is the congestion indication function of individual routers that depends only on the aggregate arrival rate $y_j(n)$ of end-users. We call the resulting controller Max-min Kelly Control (MKC) and emphasize that flows congested by the same bottleneck receive the same feedback and behave independently of the flows congested by the other links. Therefore, in the rest of this paper, we study the single-bottleneck case since each MKC flow is always congested by only one router.

**V. Exponential MKC**

Consider a particular packet.loss function $p(n)$

$$p(n) = \sum_{i=1}^{N} x_i(n-D_i) - C$$

where we again assume a network with a single link of capacity $C$ and $N$ users. This is a rather standard packet-loss function with the exception that we allow it to become negative when the link is under-utilized and achieves exponential convergence to efficiency, which explains the Exponential MKC (EMKC).

**A. Packet Loss**

EMKC converges to the combined stationary point $X^* = C/N + \alpha/N$ which is above capacity $C$. This leads to constant (albeit usually small) packet loss in the steady state. However, the advantage of this framework is that EMKC does not oscillate or react to individual packet losses, but instead adjusts its rate in response to a gradual increase in $p(n)$. Thus, a small amount of FEC can provide a smooth channel to fluctuation-sensitive applications such as video telephony and various types of real-time streaming. Besides being a stable framework, EMKC is also expected to work well in wireless networks where congestion-unrelated losses will not cause sudden reductions in the flow rates.

Also notice that EMKC’s steady-state packet loss $p^*=N\alpha/(\beta+N\alpha)$ increases linearly with the number of competing flows, which causes problems in scalability to a large number of flows. However, it still outperforms AIMD, whose increase in packet loss is quadratic as a function of $N$ [21]. Furthermore, if the network provider keeps $N = O(C)$, EMKC achieves constant packet loss in addition to exponential convergence to fairness.

Finally, observe that if the router is able to count the number of flows, zero packet loss can be obtained by adding a constant $\Delta = N\alpha/\beta C$ to the congestion indication function[3]. However, this method is impractical, since it needs non-scalable...
estimation of the number of flows $N$ inside each router. Hence, it is desirable for the router to adaptively tune $p(n)$ so that the system is free from packet loss. One such method is AVQ (Adaptive Virtual Queue) proposed in [17], [20]. We leave the analysis of this approach under heterogeneous delays and further improvements of EMKC for future work.

B. Convergence to Fairness

We next investigate the convergence rate of EMKC to fairness. To better understand how many steps EMKC requires to reach a certain level of max-min fairness, we utilize a simple metric that we call $\varepsilon$-fairness. For a given small positive constant $\varepsilon$, a rate allocation $(x_1; x_2; \ldots; x_N)$ is $\varepsilon$-fair, if:

$$f = \min_{\sum_{j=1}^N x_j} \max_{i=1}^N \frac{x_i}{x_j} \geq 1-\varepsilon \quad (7)$$

Generally, $\varepsilon$-fairness assesses max-min fairness by measuring the worst-case ratio between the rates of any pair of flows. Given the definition in (57), we have the following result.

VI. Packet Format of MKC

We next examine how to implement scalable AQM functions inside routers to provide proper feedback to MKC flows. This is a non-trivial design issue since the ideal packet loss in relies on the sum of instantaneous rates $x_i(n)$, which are never known to the router. In such cases, a common approach is to approximate model with some time-average function computed inside the router. However, as mentioned in the introduction, this does not directly lead to an oscillation-free framework since directional delays of real networks introduce various inconsistencies in the feedback loop and mislead the router to produce incorrect estimates of $X(n) = \sum_i x_i(n)$.

We provide a detailed description of various AQM implementation issues and simulate EMKC in ns2 under heterogeneous including time varying and the feedback delays.

The $\usr$ field is necessary for end-flows to determine the rate $x_i(n-D_i)$ that was in effect RTT time units earlier. The simplest way to implement this functionality is to inject the value of $x_i(n)$ into each outgoing packet and then ask the receiver to return this field in its acknowledgments.

B. The Router

Recall that MKC decouples the operations of users and routers, allowing for a scalable decentralized implementation. The major task of the router is to generate its AQM feedback and insert it in the headers of all passing packets. However, notice that the router never knows the exact combined rate of incoming flows. Thus, to approximate the ideal computation of packet loss, the router conducts its calculation of $p(n)$ on a discrete time scale of $\varepsilon$ time units. For each packet arriving within the current interval $\Delta$, the router inserts in the packet header the feedback information computed during the previous interval $\Delta$. As a consequence, the feedback is retarded by $\Delta$ time units inside the router in addition to any backward directional delays $D_i$. Since MKC is robust to feedback delay, this extra $\varepsilon$ time units does not affect stability of the system. We provide more implementation details below.

During interval $\Delta$, the router keeps a local variable $S$ which tracks the total amount of data that has arrived to the queue since the beginning of the interval. Specifically, for each incoming packet $k$ from flow $i$, the router increments $S$ by the size of the packet: $S = S + x_i(k)$. In addition, the router examines whether its locally recorded estimate $\hat{p}$ of packet loss (which was calculated in the previous interval $\Delta$) is larger than the one carried in the packet. If so, the router overrides the corresponding entries in the packet and places its own router ID, packet loss, and sequence number into the header. In this manner, after traversing the whole path, each packet records information from the most congested link.3 At the end of interval $\Delta$, the router approximates the combined arriving rate $X(n) = \sum_{i=1}^N x_i(n-D_i)$ by averaging $S$ over time $\Delta$:

$$\hat{X} = \frac{S}{\Delta} \quad (8)$$

Based on this information, the router computes an estimate of packet loss $p(n)$ using

$$\hat{p} = (X-C)/X \quad (9)$$

where $C$ is the capacity of the outgoing link known to the router (these functions are performed on a per-queue basis). After computing $\hat{p}$, the router increments its packet-loss sequence number (i.e., $seq = seq + 1$) and resets variable $S$ to zero. Newly computed values $\hat{seq}$ and $\hat{p}$ are then inserted into qualified packets arriving during the next interval $\Delta$ and are subsequently fed back by the receiver to the sender. The latter adjusts its sending rate as we discuss in the next section.

C. The User

MKC employs the primal algorithm at the end users who adjust their sending rates based on the packet loss generated by the most congested resources of their paths. However, to properly implement MKC, we need to address the following issues. First, most existing congestion control algorithms are window-based, while MKC is a rate-based method. This means that, instead of sending out a window of packets at once, each MKC user $i$ needs to properly pace its outgoing packets and maintain its sending rate at a target value $x_i(n)$. We implement this mechanism by explicitly calculating the interpacket interval $\delta(k)$ of each packet $k$.
\[ \delta(k) = \frac{S(k)}{X_i(n)} \] \hspace{1cm} (10)

where \( S(k) \) is the size of packet \( k \) of user \( i \). Second, notice that ACKs carrying feedback information continuously arrive at the end-user and for the most part contain duplicate feedback (assuming \( \Delta \) is sufficiently large). To prevent the user from responding to redundant or sometimes obsolete feedback caused by reordering, each packet carries a sequence number \( seq \), which is modified by the bottleneck router and is echoed by the receiver to the sender. At the same time, each end-user \( i \) maintains a local variable \( seq_i \), which records the largest value of \( seq \) observed by the user so far. Thus, for each incoming ACK with sequence \( seq \), the user responds to it only when \( seq > seq_i \). This allows MKC senders to pace their control actions such that their rate adjustments and the router’s feedback occur on the same timescale. Third, MKC requires both the delayed feedback \( \eta_i(n) \) and the delayed reference rate \( x_i(n-D) \) when deciding the next sending rate.

VII. Conclusion

This paper investigated the properties of Internet congestion controls under non-negligible directional feedback delays. We focused on the class of control methods with radial Jacobians and showed that all such systems are stable under heterogeneous delays. To construct a practical congestion control system with a radial Jacobian, we made two changes to the classic discrete Kelly control and created a max-min version we call MKC. Combining the latter with a negative packet-loss feedback, we developed a new controller EMKC and showed in simulations that it offers smooth sending rate and fast convergence to efficiency. Furthermore, we demonstrated that EMKC’s convergence rate to fairness is exponential when the network provider scales the number of flows \( N \) as \( \Theta(C) \) and linear otherwise. From the implementation standpoint, EMKC places very little burden on routers, requires only two local variables per queue and one addition per arriving packet, and allows for an easy implementation both in end-to-end environments and under AQM support. Our future work involves improvement of the convergence speed to fairness and design of pricing schemes for EMKC to achieve loss-free performance regardless of the number of flows \( N \). It is generally accepted that future communication and computer networks will be characterised by high-speed and long-distance connectivity, and by the requirement to carry a wide variety of network services and traffic types. These demands create new challenges for network designers and researchers. It is widely recognised that transport layer enhancements are essential if high performance next generation networks are to be realized.

Our objective here is to develop a systematic framework for modifying the basic TCP algorithm that renders it suitable in a variety of network types. We describe a new TCP-variant that is suitable for deployment in high-speed and long distance networks, as well as conventional networks. The new TCP variant, H-TCP, is shown to be fair when deployed in homogeneous networks, to be friendly when competing with conventional TCP sources, to rapidly respond to changes in available bandwidth, and to utilise link bandwidth efficiently.

References


