Automatica 50 (2014) 3112–3122

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

A family of multi-path congestion control algorithms with global stability and delay robustness☆

Wei-Jie Feng [a], Lei Wang [a], Qing-Guo Wang [b]

[a] Laboratory of Mathematics, Information and Behavior of the Ministry of Education, School of Mathematics and Systems Science, Beihang University, Beijing, 100191, P. R. China

[b] Department of Electrical and Computer Engineering, National University of Singapore, Singapore, 119260, Singapore

ARTICLE INFO

Article history:
Received 20 May 2012
Received in revised form
18 September 2013
Accepted 30 July 2014
Available online 4 November 2014

Keywords:
Congestion control
Dynamic multi-path
Global stability
Lyapunov–Krasovskii functional
Delay robustness

ABSTRACT

The goal of traffic management is to efficiently allocate network resources via adjustment of source transmission rates and routes selection. Mathematically, it aims to solve a traditional utility maximization problem in a fair and distributed manner. In this paper, we first develop a generalized multi-path utility maximization problem which features a weighted average of the classical Kelly's formulation and the Voice's model. Next, we design from this broader framework a family of multi-path dual congestion control algorithms whose equilibrium point can both achieve a desired bandwidth utilization and preserve a notion of fairness among competing users. Global stability can be guaranteed for the proposed schemes in the absence of delays by use of a totally novel Lyapunov function. Moreover, when heterogeneous propagation delays are taken into account, we establish decentralized and scalable sufficient conditions for robust global stability by constructing a reasonable Lyapunov–Krasovskii functional candidate. These conditions give estimates for the maximum admissible delays that the controller can tolerate without losing stability. Finally, we verify the results through simulation.

1. Introduction

Traffic management is concerned with the regulation of source transmission rates and routes adaptation to improve network performance for end users, while network resources are fully utilized but not exceeded. Ever since the substantial work of Kelly, Maulloo, and Tan (1998), nonlinear differential systems, generally called congestion control algorithms, can be viewed as decentralized schemes to solve a convex but centralized optimization problem, i.e., the maximization of an aggregate utility across all users and routes, subject to link capacity constraints. Each "utility" function represents the performance, or level of "satisfaction" of a user when a certain amount of resource is allocated to it. Following Kelly et al. (1998), it has become customary to denote by "primal" algorithms those that contain dynamics at sources, but static functions at links, by "dual" algorithms those where the opposite holds, and by "primal–dual" algorithms those where both sources and links are dynamic. For these large-scale coupled systems, stability, efficiency and steady state fairness are all particularly challenging problems partially because of unknown network size and topology. Therefore, there has long been a desire to design congestion control algorithms in a distributed manner, which has to ensure (at equilibrium) optimality, fairness and stability, but more importantly robustness to uncertainties of network topology and time delays.

Currently, the Transmission Control Protocol and the Internet Protocol, known as TCP/IP, are widespread for guiding traffic flows in the Internet. According to IP routing, a single path, typically the shortest path, is chosen between its source and destination. Then the source sending rate is varied based on congestion level along that path. Although IP routing is highly scalable, the static single-path routing scheme fails to react to instantaneous network congestion, and may limit the achievable throughput. Thus, the Internet would be more efficient and robust if traffic flows could be flexibly divided over multiple paths. Multi-path routing would offer many benefits summarized as follows: improving end-to-end reliability, avoiding congested paths and accessing more bandwidth. Recently, motivated by the applications in ad hoc networks and overlay TCP, there have been more interests in

☆ This work is supported by the National Natural Science Foundation of China under Grant Nos. 61473016, 61004106 and 10426003. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor David Angeli under the direction of Editor Andrew R. Teel.

E-mail addresses: fengwj@buaa.edu.cn (W.-J. Feng), lwang@buaa.edu.cn (L. Wang), elewqg@nus.edu.sg (Q.-G. Wang).

http://dx.doi.org/10.1016/j.automatica.2014.10.061
0005-1098/

multi-path routing schemes (Han, Shakkottai, Hollot, Srikant, & Towsley, 2006; Kelly & Voice, 2005; Voice, 2007). Unfortunately, once multiple alternative routes are available, a major difficulty is that the considered utility maximization problem is no longer strictly concave, resulting in the possible existence of multiple optimal solutions. Thus researchers resort to the duality of the primal problem. However, the dual may not be differentiable at every point when extended to the multi-path case, and hence the dual congestion control algorithm in Kelly et al. (1998) shares some oscillation behaviors, which cause barriers to the deployment of the dynamic multi-path routing scheme.

Time-delay, even though often neglected to facilitate analysis, is ubiquitous in networks. If one neglects network delays, global stability for various congestion control algorithms can be established by means of the conventional Lyapunov function, first for the single-path case (Kelly, 2003; Kelly et al., 1998; Paganini, 2002), then for the multi-path case (Kelly & Voice, 2005; Lin & Shroff, 2006; Voice, 2007). However, doing so runs the risk of performance degradation or even instability. Therefore, time-delay should be taken into account. In this situation, Lyapunov function method can deal well with some simple network instances, e.g., single bottleneck or fixed delays, as in Deb and Srikant (2003), Ranjan, La, and Abed (2006) and Wang and Paganini (2003, 2006). To tackle the case of an arbitrary network with heterogeneous delays, the first consideration is the generalized Nyquist criterion, by which local stability can be obtained through linearized analysis of such systems around the equilibrium point. See the references on this technique and its applications to congestion control (Desoer & Yang, 1980; Han et al., 2006; Kelly, 2003; Kelly & Voice, 2005; Paganini, Doyle, & Low, 2001; Paganini, Wang, Doyle, & Low, 2005; Srikant, 2004; Tian, 2005; Tian & Yang, 2004; Vinnicombe, 2002; Voice, 2007). A significant open challenge is to verify whether the designing criteria derived from such linearized analysis can guarantee the global stability. The Lyapunov–Razumikhin theorem is an effective tool used to analyze the global stability for delay-based nonlinear systems, particularly treating delay-independent sufficient conditions. Yet it only applies to a specially structured single-path primal algorithm (Ying, Dullerud, & Srikant, 2006). Rather than specific controllers, when considering the traditional Kelly's primal and dual algorithms, global stability is established from the ISS small-gain theorem with respect to delays (Fan & Arcak, 2006; Fan, Arcak, & Wen, 2004). However, the analysis is restricted to the single-path network. Very recently, delay robustness is addressed again for the Kelly's control laws in Papachristodoulou and Jadbabaie (2010), where the analysis is pursued by use of an appropriately structured Lyapunov–Krasovskii functional (LKF). Furthermore, decentralized conditions are provided for global stability, by which the controller gains can be scaled down in a similar manner to Fan et al. (2004). The main difference, however, is that Papachristodoulou and Jadbabaie (2010) presents a relatively more scalable gain choice. But the analysis is still applicable to the single-path case.

Naturally, it would be desirable to study global stability for network controllers with heterogeneous delays and multi-path routing. A stumbling block in extending the results of Papachristodoulou and Jadbabaie (2010) to the multi-path case lies in two aspects: possible existence of multiple equilibria and difficult construction of a reasonable LKF candidate. It is worth mentioning that our latest work (Feng, 2012), with the help of Voice (2007), has made some progress on these directions, including the design of a globally stable controller and the establishment of LKF. However, the fairness among different users is not considered in Feng (2012) and Voice (2007). To the best of our knowledge, the fairness for multi-path networks has not been addressed in the literature.

In this paper, following the work of Feng (2012), we develop a generalized multi-path utility maximization problem which takes a weighted average of two existing models, i.e., the classical Kelly's formulation (Kelly et al., 1998) and the Voice's model (Voice, 2007). This problem gives a broader design framework, applicable to diverse requirements of the network. Namely, it can not only guarantee the unique equilibrium point at which high utilization of resources are achieved, but also maintain the fairness among the involved users. The definition of fairness presented here is a generalization of proportional fairness for the case of single-path to multi-path. Based on dual decomposition and gradient projection methods, we propose a large family of multi-path dual congestion control algorithms with varying parameters, where our earlier work (Feng, 2012) can be interpreted as a special case by setting a proper parameter. To demonstrate global stability for such an extensive class of systems, the key is to find a suitable LKF candidate, which can be constructed by borrowing ideas of Feng (2012). Particularly, if delays are ignored, it reduces to the corresponding Lyapunov function. Furthermore, if only one route is allowed, it is the exact one frequently used in most prior studies, e.g., Srikant (2004) and Wen and Arcak (2004). In the presence of propagation delays, we show global stability for sufficiently small time-delays. And for large delays, we give the formulas by which the controller gains can be scaled to preserve global stability.

The rest of this paper is organized as follows. In Section 2, we introduce a weight parameter and describe a generalized multi-path utility maximization problem formulation. Then we derive its dual that motivates our approach, along with the discussion of fairness. In Section 3, we first propose a family of multi-path dual congestion control algorithms, then we exhibit the main results on global stability in the absence of delays and in the presence of delays, respectively. The proofs of theorems are relayed in Section 4. We present numerical simulations to illustrate the performance of the proposed schemes in Section 5 and draw conclusion in Section 6. Finally, proofs of lemmas are given in the Appendix.

2. Problem formulation

This section consists of the following five parts: network model, a generalized multi-path utility maximization problem, approximation error, deduction of the dual problem and the notion of fairness.

2.1. Network model

Consider a network with a set $\mathcal{L}$ of links (resources) with capacities c_l, $l \in \mathcal{L}$. The network is shared by a set $\mathcal{S}$ of users (sources). Let a route r be a non-empty subset of $\mathcal{L}$, and write $\mathcal{R}$ for the set of possible routes. There could be multiple routes that connect a user s to its destination, and let $\mathcal{R}_s$ denote the routes set belonging to a user s. To succinctly express the relationships between routes and links, we define a 0–1 matrix $\mathbf{A} = (A_{lr}, l \in \mathcal{L}, r \in \mathcal{R})$, where $A_{lr} = 1$ if a link l lies on a route r, and $A_{lr} = 0$ otherwise. Similarly, let $H_{sr} = 1$ if a route r serves a user s, and set $H_{sr} = 0$ otherwise. This defines a 0–1 matrix $\mathbf{H} = (H_{sr}, s \in \mathcal{S}, r \in \mathcal{R})$. To model the fluid-flow traffic, a route r is associated with it a flow rate $x_r \geq 0$, which represents the rate at which a user s is sending packets along it. Let $\mathbf{c}$ and $\mathbf{x}$ denote the vectors with components c_l and x_r, respectively, over the sets $\mathcal{L}$ and $\mathcal{R}$.

The presence of propagation delays between sources and links can be described as follows. For a link l and a route r passing through it, denote by T_{rl} (T_{lr}) the forward (backward) propagation delay from s (l) to l (s), i.e., the length of time it takes for a packet to travel from a source s (a link l) to a link l (a source s) along a route r. In the protocols, it is assumed that a packet must reach its destination before an acknowledgment containing congestion feedback is returned to its source. Denote by $T_r = T_{rl} + T_{lr}$ the round trip time for a route r.