Towards Resource-Optimal Routing Plans for Real-Time Traffic

Alessandro Lori, Giovanni Stea, and Gigliola Vaglini

Dipartimento di Ingegneria dell’Informazione, University of Pisa
Via Diotisalvi 2, I-56122 Pisa, Italy
{a.lori,g.stea,g.vaglini}@iet.unipi.it

Abstract. We discuss the issue of computing resource-optimal routing plans in a network domain. Given a number of known traffic demands, with associated required delays, we discuss how to route them and allocate resources for them at each node so that the demands are satisfied. While a globally optimal routing plan requires joint computation of the paths and of the associated resources (which was claimed to be NP-hard), in this paper we stick to existing approaches for path computation, and use mathematical programming to model resource allocation once the paths are computed. We show that the problem is either convex or non-convex, depending on the scheduling algorithms adopted at the nodes. Our results show that, by computing resources per-path, instead of globally, the available capacity can be exceeded even at surprisingly low utilizatons.

1 Introduction

Real-time traffic over IP networks has become a reality. Several applications, e.g. industrial control, remote sensing and surveillance systems, live IPTV and VoIP etc., all requiring real-time guarantees (i.e., a bound on the end-to-end delay) are increasingly being deployed. Internet Service Providers are already facing, or will soon face, the challenge of configuring their network domains so as to provide deterministic delay bound guarantees to their customers – whether single users or lower-tier providers themselves – by negotiating real-time oriented Service Level Agreements (SLAs). Supporting SLAs with real-time constraints requires proper Traffic Engineering (TE) and resource optimization practices. Multi-Protocol Label Switching (MPLS, [2]) with TE extensions (MPLS-TE, [17]) allows traffic trunks to be routed along arbitrary paths, and resources to be allocated on those paths at the time of flow setup.

Supporting delay constrained traffic requires in fact both computing paths and reserving resources along those paths. The usual assumption (to which we stick in the rest of the paper) is that traffic trunks are scheduled at each node so as to be reserved a minimum guaranteed rate. As far as path computation is concerned, a relevant amount of literature has been published since the late ’90s, under the name of QoS routing, a good review of which can be found in [25]. Most papers (see, e.g. [21-23]), assume that delays are static and/or additive per-link metrics. However, delay bounds do depend on the amount of reserved resources at each link, i.e. on the number and amount of flows traversing them, and the expression of the delay bound is not linear.
in the number of links. Other papers tackle the problem from a probabilistic point of view, assuming a stochastic characterization of traffic and attempting to minimize or bound the average delay, which is hardly relevant for real-time traffic (e.g. [25]). A limited number of works [18-19] propose path computation techniques constrained by deterministic (non additive) delay bound constraints, taking resource allocation into account. [18] shows that it is possible to compute a shortest path for a single flow, subject to end-to-end delay bounds, also computing the rate to be reserved on each node during path computation, at a polynomial cost. It also assumes that an equal rate has to be reserved at each node for the path. [19] proposes lower-complexity approximate solutions to the problem solved exactly in [18].

As far as resource allocation is concerned, the problem is often referred to as QoS partitioning in the literature. On that topic, several works exist that achieve optimal partitions for additive delays on a given path (see, e.g., [25]). An interesting work [20] shows that, when using end-to-end delay bounds as constraints, reserving the same rate (as done in [18-19]) may be suboptimal and lead to failing of paths which might indeed be admissible. Authors propose an algorithm that allows a delay-feasible resource allocation to be computed on a given path, if such an allocation exists.

To the best of our knowledge, the problem of making a global routing and resource allocation plan under delay bound constraints has received little attention so far. [19] claims that the problem is NP-hard. However, this does not mean that it is not solvable for practical dimensions (i.e., comparable to those of today’s and tomorrow’s network domains), nor it implies that good suboptimal solutions cannot be computed in reasonable time, even for large dimensions. Besides, global routing plans do not need to be computed in real time. Network engineering and optimization cycles – where new routing plans are made from scratch, based on the traffic forecast and negotiated SLAs – do not take place more frequently than daily or weekly, hence computation time can be traded for optimality. Second, per-path computation and resource allocation is feasible in a dynamic environment (online TE), but is clearly suboptimal when routing plans are considered (offline TE).

In this paper we mark a first step in this direction by tackling global resource allocation with delay bound constraints in a network domain. We assume that paths have been selected (and we evaluate several existing options for the path computation phase), and we exploit optimization techniques to minimize the amount of rate reserved in the network domain. We show that the problem can be solved optimally for several classes of schedulers. For some schedulers it has a convex formulation, which means that optimal solutions can be found in a reasonable time. Our first results show that, even at surprisingly low network loads, global allocation is necessary to be able to guarantee delay bounds when it is indeed feasible to do so, as per-path solutions are generally ineffective.

While in this work we assume that routing is given, and we only aim at optimizing the resource allocation, the long-term goal of this stream of research is to provide effective algorithms for joint path computation and resource reservation in a network domain, which is actively being pursued at the time of writing.

The rest of the paper is organized as follows: in section 2 we describe the system model. We formulate the resource optimization problem in Section 3. Section 4

---

1 The paper reports the statement as “proposition 1”, without a proof or a reference to one.