Quality–of–Service Routing

Sudip Misra
Cornell University, USA

INTRODUCTION

The area of quality-of-service (QoS) routing is concerned with selecting routing paths while meeting strict end-to-end service requirements involving resource constraints, while achieving optimum throughput in the network. The usefulness of QoS routing is not new. QoS routing is quite popular in the telecommunications industry because of the increased demand for satisfying multiple customer demands and obtaining increased utilization of network resources, while satisfying the varied user requirements.

BACKGROUND

Two basic considerations in QoS routing in integrated services packet-switched networks concern: (1) routing traffic with bandwidth guarantees, and (2) routing traffic with delay guarantees. Some of the algorithms proposed traditionally for solving the former class of problems are the widest-shortest path (WSP) (Guerin, Orda, & Williams, 1997), and the shortest-widest path (SWP) (Wang & Crowcroft, 1996) algorithms. More recently, Vasilakos, Saltouros, Atlassis, and Pedrycz, (2003) proposed the stochastic estimator learning automata (SELA) routing algorithm for QoS routing in asynchronous transfer mode (ATM) networks.

Online routing using traffic engineering (TE) principles has recently drawn considerable attention. We mention here a few TE algorithms proposed in the literature (e.g., Iliadis & Bauer, 2002; Kar, Kodialam, & Lakshman, 2000; Suri, Waldvogel, Bauer, & Warkhede, 2003; Szeto, Boutaba, & Iraqi, 2002; Wang, Su, & Chen, 2002). Of all the online TE algorithms, we believe that the one that has attracted the most attention is the minimum interference routing algorithm (MIRA) designed by Kar et al. (2000).

In addition to the MIRA algorithm, there are a few other TE routing algorithms that were proposed by other researchers, some of which are: the profile-based routing (PBR) (Suri et al., 2003); the dynamic online routing algorithm (DORA) (Szeto et al., 2002); Iliadis and Bauer's (2002) algorithm; Wang et al.'s (2002) algorithm; and the random races-based traffic engineering routing algorithm (RRATE) (Oommen, Misra, & Granmo, 2006). Some of these are also described in the sections to follow.

QUALITY-OF-SERVICE ROUTING

A network is said to support QoS (Guerin et al., 1997; Peterson & Davie, 2000), if it has the capability of treating different packets differently. QoS technology has enabled service providers to support different levels of service to different customers, thereby capacitating them with the option to provide better levels of paid services to some customers more than to others. For example, some groups of customers may be concerned with a service that guarantees packet delivery, even if that means paying a higher price for these services, others may just as well be satisfied with relatively less reliable data transfer by paying less for their subscribed services. Networks that transport multimedia traffic, that is, voice, data, and video need differential treatments of different packets—while voice traffic is highly sensitive to time delay and the orderly delivery of packets, data traffic is relatively less sensitive to these, and videoconferencing traffic requires a dedicated connection for a fixed amount of time for the real-time, orderly delivery of packets.

Typical QoS routing-based performance metrics are bandwidth, delay, and throughput. While some applications require bandwidth guarantees, some others mandate the satisfiability of strict end-to-end delays, and others still require a high throughput, or a combination of both of these criteria (Ma, 1998).

Routing protocols in the pre-QoS era did not consider QoS requirements of connections (e.g., delay, bandwidth, and throughput). Furthermore, optimization of resource utilization was not a primary goal. As a result, while there were flow requests that were rejected because of nonavailability of sufficient underlying resources, there were some other resources that remained available. To address such deficiencies with conventional routing protocols, QoS routing algorithms were devised that could locate network paths which satisfy QoS requirements and which made better use of the network. Routing of QoS traffic requires stringent performance guarantees of the QoS metrics (e.g., delay, bandwidth, and throughput) over the paths selected by the routing algorithms. Accordingly, whereas the traditional shortest path algorithms, for example, Dijkstra (1959)'s or Bellman (1958)'s, indeed, have the potential of selecting a feasible path for routing, QoS routing algorithms must consider multiple QoS and resource utilization constraints, typically making the problem intractable. QoS routing is, thus,

different from that of routing in traditional circuit-switched and packet-switched networks.

In ATM networks, for example, a connection is accepted or not by the Connection Admission Control (CAC), depending on whether or not sufficient resources (e.g., bandwidth) are available. The CAC operates by taking into account factors such as the incoming QoS requests and the available resources in the network. The CAC operates the QoS routing algorithms, which identify whether the different possible candidate paths satisfy the QoS requirements or not. The network architects and the engineers want to choose such a routing algorithm that will help the service providers for maximizing the network resources (e.g., the amount of bandwidth to be routed), while satisfying the requirements from the customers. Therefore, the design of any good QoS routing algorithm takes into account factors such as satisfying the QoS requirements, optimizing the consumption of the network resources (e.g., buffer space, link bandwidth), and balancing the traffic load across different paths. In addition to these, a good QoS routing algorithm should characterize itself by its ability to adapt to the periodic dynamic behavior of the network.

Conventional QoS Routing Algorithms

In this section we present an overview of some of the traditional algorithms for QoS Routing.

- **SWP:** The SWP algorithm was proposed by Wang and Crowcroft (1996). They proposed two variants of the algorithm—the *distance-vector-based SWP*, and the *link-state based SWP*—depending on whether the SWP algorithm is governed by *distance-vector-based routing* or by *link-state-based routing*. Essentially, in both variants of the SWP algorithm, the algorithm finds a route with the widest path, that is, the path with the maximum bottleneck bandwidth. When there is more than one choice available, the algorithm chooses the path that has the minimum length, that is, the one that has the shortest propagation delay (Wang & Crowcroft, 1996). If still there is a tie between one or more such path(s), one of the prospective paths is randomly chosen.
- **WSP:** The WSP algorithm was proposed by Guerin et al. (1997). Unlike the SWP algorithm, WSP first attempts to compute the shortest path; if there is more than one alternative, the algorithm chooses the one with the largest residual bandwidth in the bottleneck link (i.e., the widest path). If there is still a tie with one or more such path(s), one of the prospective paths is randomly chosen.
- **SELA:** The SELA routing algorithm was proposed by Vasilakos et al. (2003), and is based on the concept

of SELA (Vasilakos & Papadimitriou, 1992). SELA is a QoS-based dynamic source routing algorithm where each source node maintains a database of the topology information between it and its k-shortest path neighbors (to each reachable destination node) (Cormen, Leiserson, & Rivest, 1990). In SELA, a learning automaton is stationed at each node in the network for determining how each call is to be routed between every node to every other reachable node in the network. For establishing a call, the source node selects one of the precomputed shortest path routes that can potentially be accepted by the algorithm based on the QoS requirements and traffic parameters. If no such path can be found, SELA rejects the request. In the heart of the SELA routing algorithm is the design of a function that is used by SELA to estimate the environmental feedback of the path selected by the algorithm. Further details of the algorithm can be found in Vasilakos et al. (2003).

Traffic Engineering Routing

TE mandates to optimize the performance of traffic handling and resource utilization on existing physical network topologies. This is, in principle, engineered by minimizing the over utilization of network capacity and distributing the traffic load on costly network resources such as, links, routers, switches, and gateways (Osborne & Simha, 2002). In the context of routing, TE is of great usefulness because traditional routing techniques are based on greedy shortest path computation techniques that lead to the over utilization of certain network resources, even when other resources remain under utilized.

Multi-protocol Label Switching (MPLS) has recently emerged for many professionals as a de facto standard in TE. MPLS is an Internet Engineering Task Force (IETF) standard which merges the layer 2 information of bandwidth, latency, and utilization of network links, with the control protocols used in layer 3 Internet protocol (IP), in order to simplify the exchange of IP packets. At the heart of the idea is the usage of a *label* (or a *tag*) to calculate shortest paths to all destinations within an autonomous system, thereby expediting the forwarding of packets. A label can be perceived as a simplified representation of an IP packet's header, with the additional advantage of enabling core backbone networks to operate at high speeds because of the exclusion of the need to re-examine each packet's IP header in detail. This, in turn, permits the differentiating between packets on an individual basis and facilitates the support of QoS. The destination of a packet is determined by observing the label and not the IP address it is destined to. MPLS helps network operators manage network route failures and makes the system able to decongest bottleneck links by providing a detour for the incoming traffic. It can also help service providers manage the

Related Content

Everyone's Watching
William L. Kuechlerand Dana Edberg (2002). *Annals of Cases on Information Technology: Volume 4 (pp. 225-240).*
www.irma-international.org/article/everyone-watching/44509

Enterprise Resource Planning (ERP) Implementation Approaches and the Performance of Procure-to-Pay Business Processes: (Field Study in Companies that Implement Oracle ERP in Jordan)
Ibrahim Kofahiand Haroun Alryalat (2017). *International Journal of Information Technology Project Management (pp. 55-71).*
www.irma-international.org/article/enterprise-resource-planning-erp-implementation-approaches-and-the-performance-of-procure-to-pay-business-processes/169830

Debunking Intermediary Censorship Framework in Social Media via a Content Retrieval and Classification Software
Baramee Navanopparatskul, Sukree Sinthupinyoand Pirongrong Ramasoota (2017). *Ontologies and Big Data Considerations for Effective Intelligence (pp. 169-195).*
www.irma-international.org/chapter/debunking-intermediary-censorship-framework-in-social-media-via-a-content-retrieval-and-classification-software/177393

An IT Project Management Framework for Assessing the Dynamism of Culture under Globalization: Evidence from Zimbabwe
Sam Takavarasha, Gilford Hapanyengwi, Donald Chimanikireand Gabriel Kabanda (2013). *International Journal of Information Technology Project Management (pp. 66-81).*
www.irma-international.org/article/an-it-project-management-framework-for-assessing-the-dynamism-of-culture-under-globalization/102481

Understanding Factors Affecting Consumer Adoption of Broadband in India: A Pilot Study
Yogesh K. Dwivedi, Michael D. Williams, Vishanth Weerakkody, Banita Laland Sneha Bhatt (2008). *Journal of Cases on Information Technology (pp. 35-47).*
www.irma-international.org/article/understanding-factors-affecting-consumer-adoption/3227