Computer Communications 35 (2012) 565–578

Contents lists available at ScienceDirect

Computer Communications

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

Taming traffic dynamics: Analysis and improvements

Pedro Casas [a,b], Federico Larroca [b,c,*], Jean-Louis Rougier [c], Sandrine Vaton [a]

[a] Télécom Bretagne, Technopôle Brest-Iroise, CS 83818, 29238 Brest Cedex 3, France
[b] Facultad de Ingeniería, Universidad de la República, Julio Herrera y Reissig 565, C.P. 11.300, Montevideo, Uruguay
[c] Télécom ParisTech, 46 rue Barrault, F-75634 Paris Cedex 13, Paris, France

ARTICLE INFO

Article history:
Available online 29 July 2010

Keywords:
Traffic uncertainty
Traffic management
Robust optimization
Robust Routing
Dynamic Load Balancing

ABSTRACT

Internet traffic is highly dynamic and difficult to predict in current network scenarios, which enormously complicates network management and resources optimization. To address this uncertainty in a robust and efficient way, two almost antagonist Traffic Engineering (TE) techniques have been proposed in the last years: Robust Routing and Dynamic Load Balancing. Robust Routing (RR) copes with traffic uncertainty in an off-line preemptive fashion, computing a single static routing configuration that is optimized for traffic variations within some predefined uncertainty set. On the other hand, Dynamic Load Balancing (DLB) balances traffic among multiple paths in an on-line reactive fashion, adapting to traffic variations in order to optimize a certain congestion function. In this article we present the first comparative study between these two alternative methods. We are particularly interested in the performance loss of RR with respect to DLB, and on the response of DLB when faced with abrupt changes. This study brings insight into several RR and DLB algorithms, evaluating their virtues and shortcomings, which allows us to introduce new mechanisms that improve previous proposals.

1. Introduction

As network services and Internet applications evolve, network traffic is becoming increasingly complex and dynamic. The convergence of data, telephony and television services on an all-IP network directly translates into a much higher variability and complexity of the traffic injected into the network. To make matters worse, the presence of unexpected events such as network equipment failures, large-volume network attacks, flash crowd occurrences and even external routing modifications induces large uncertainty in traffic patterns. Moreover, current evolution and deployment-rate of broadband access technologies (e.g. Fiber To The Home) only aggravates this uncertainty.

But these are not the only problems network operators are confronted with. The ever-increasing access rates available for end-users we just mentioned is such that the assumption of infinitely provisioned core links could soon become obsolete. In fact, recent Internet traffic studies from major network technology vendors like Cisco Systems forecast the advent of the Exabyte era [1,2], a massive increase in network traffic driven by high-definition video. In this context, simply upgrading link capacities may no longer be an economically viable solution. Moreover, even if overdimensioning would be possible, its environmental impact is not negligible. For instance, the Information and Communication Technology sector alone is responsible for around 2% of the man-made CO_2, a similar figure to that of the airline industry, but with higher increasing perspectives [3]. An efficient and responsible usage of the resources is then essential.[1]

In the light of this traffic scenario, we study the problem of intradomain Traffic Engineering (TE) under traffic uncertainty. This uncertainty is assumed to be an exogenous traffic modification, meaning that traffic variations are not produced within the domain for which routing is optimized but are due to external and difficult to predict events. More in particular, we are interested in two almost antagonist approaches that have emerged in the recent years to cope with both the increasing traffic dynamism and the need for cost-effective solutions: Robust Routing (RR) [6–8] and Dynamic Load Balancing (DLB) [9–11].

In RR, traffic uncertainty is taken into account directly within the routing optimization, computing a single routing configuration for all traffic demands within some *uncertainty set* where traffic is assumed to vary. This uncertainty set can be defined in different ways, depending on the available information: largest values of links load previously seen, a set of previously observed traffic demands (previous day, same day of the previous week), etc. The

* Corresponding author. Present Address: Facultad de Ingeniería, Universidad de la República, Julio Herrera y Reissig 565, C.P. 11.300, Montevideo, Uruguay. Tel.: +598 2 711 09 74; fax: +598 2 711 74 35.

E-mail addresses: pedro.casas@telecom-bretagne.eu (P. Casas), federico.larroca@telecom-paristech.fr, flarroca@fing.edu.uy (F. Larroca), rougier@telecom-paristech.fr (J.-L. Rougier), sandrine.vaton@telecom-bretagne.eu (S. Vaton).

[1] To learn more about this emerging discipline, the interested reader should consult works related to so-called "green networking" [4].

0140-3664/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2010.07.009

criterion to search for this unique routing configuration is generally to minimize the maximum link utilization (i.e. the utilization of the most loaded link in the network) for all traffic demands of the corresponding uncertainty set. While this routing configuration is not optimal for any single traffic demand within the set, it minimizes the worst case performance over the whole set.

DLB copes with traffic uncertainty and variability by splitting traffic among multiple paths in real-time. In this dynamic scheme, each origin–destination (OD) pair of nodes within the network is connected by several a priori configured paths, and the problem is simply how to distribute traffic among these paths in order to optimize a certain function. DLB is generally defined in terms of a link-congestion function, where the portions of traffic are adjusted in order to minimize the total network congestion. Ideally, the traffic distribution is set so that at every instant the objective function is optimized.

Those who promote DLB highlight among others the fact that it is the most resource-efficient possible scheme, and that given the configured paths it supports every possible traffic demand, all of this in an automated and decentralized fashion. In practice, the "always-optimized" characteristic we mentioned above is achieved by means of a distributed algorithm periodically executed by every ingress router based on feedback from the network. It is precisely this last characteristic that constitutes the most challenging aspect of DLB. In fact, the deployment of DLB has been, to say the least, limited. Two particular problems arise in DLB: convergence to the optimum is not always guaranteed, and convergence speed might be over-killing under large and abrupt changes in traffic demands. Network operators are reluctant to use dynamic mechanisms mainly because they are afraid of a possible oscillatory behavior of the algorithm used by each OD pair to adjust Load Balancing. As the early experiences in ArpaNet has proved [12], these concerns are not without reason. (In particular, before July 1987, the links' metric was defined as the packet delay averaged over a 10 s period. Although this adaptive routing scheme worked correctly under light or moderate loads, it generated oscillations under relatively heavy loads. This resulted in substituting the links' metric by a fixed value as we use it today, sacrificing optimality for stability.) Indeed, for these adaptive and distributed algorithms, a trade-off between adaptability (convergence speed) and stability must be found, which may be particularly difficult in situations where abrupt traffic changes occur.

Those who advocate the use of RR claim that there is actually no need to implement supposedly complicated and possibly oscillatory dynamic routing mechanisms, and that the incurred performance loss for using a single routing configuration is negligible when compared with the increase in complexity. RR provides a stable routing configuration for all the traffic demands within the uncertainty set, avoiding possible oscillations and convergence issues. However, RR presents some conception problems and serious shortcomings in its current state which we highlight and try to ease in this work. The first drawback of current RR is related to the objective function it intends to minimize. Optimization under uncertainty is generally more complex than classical optimization, which forces the use of simpler optimization criteria such as maximum link utilization (MLU). The MLU is not the most suitable network-wide optimization criterion; setting the focus too strictly on MLU often leads to worse distribution of traffic, adversely affecting the mean network load and thus the total network end-to-end delay, an important QoS indicator. It is easy to see that the minimization of the MLU in a network topology with heterogeneous link capacities may lead to poor results as regards global network performance. The second drawback of RR we identify is its inherent dependence on the definition of the uncertainty set of traffic demands: the uncertainty set has to be sufficiently "large" to allow traffic flexibility and to provide performance guarantees, but should not be excessively "large" to avoid wasting network resources. Thus, considering an unique RR configuration to address both traffic in normal operation and unexpected traffic variations is an inefficient strategy, as a single routing configuration cannot be suitable for both situations.

1.1. Contributions of this article

This article presents a fair and comprehensive comparative analysis between RR and DLB mechanisms. The analysis is comprehensive as it evaluates the performance of both mechanisms based on different performance indicators and considering normal operation as well as unpredicted traffic events. We believe our comparison is fair because it considers the particular characteristics of each mechanism under the same network and traffic conditions. To date and to the best of our knowledge this is the first work that conducts such a comparative evaluation, necessary indeed not only from a research point of view but also for network operators who seek cost-effective and robust solutions to face future network scenarios. Based on this comparative analysis we develop and evaluate new variants of RR and DLB mechanisms, improving some of the shortcomings found in both static and dynamic approaches.

Regarding the RR approach, we will introduce some modifications that strive to alleviate the two problems identified in current proposals. We will first study which is the best objective function to minimize, and propose the mean link utilization instead of the MLU. The mean link utilization provides a better image of network-wide performance, as it does not depend on the particular load or capacity of each single link in the network but on the average value. However, a direct minimization of the mean link utilization does not assure a bounded MLU, which is not practical from an operational point of view. Thus, we minimize the mean link utilization while bounding the MLU by a certain utilization threshold a priori defined. This adds a new, and maybe difficult to set, constraint to the problem, namely how to define this utilization threshold. We further improve our proposal by providing a multiple objective optimization criterion, where both the MLU and the mean link utilization are minimized simultaneously. We evaluate the improvements of our proposals from a QoS perspective, using the mean path end-to-end queuing delay as a measure of global performance.

The second problem we address in RR is the trade-off between routing performance and routing reliability. In [13] we have recently proposed a solution to manage this trade-off, known as Reactive Robust Routing (RRR). Basically, RRR consists of constructing a RR configuration for expected traffic in nominal operation, adapting this nominal routing configuration after the detection and localization of a large and long-lived traffic modification. RRR provides good performance for both nominal operation and unexpected traffic, but it is difficult to deploy in a real implementation, because of the routing reconfiguration step. Reconfiguring the routing of an entire Autonomous System is a nontrivial task. In this article we modify the RRR approach, using a preemptive Load Balancing algorithm to balance traffic among pre-established paths after the localization of a large volume traffic modification (preemptive in the sense of preventing a situation from occurring).

In what respects DLB, we evaluate the use of so-called *no-regret* algorithms as the distributed optimization algorithm used by ingress routers to adapt Load Balancing. The authors of a recent paper [14] proved that if all OD pairs use algorithms of this kind, convergence to the optimum is guaranteed. Special attention will be paid on the behavior of the algorithm when faced with abrupt and unexpected changes in the traffic demands. We shall introduce simple, and yet effective, modifications to the algorithm to assure a fast convergence to the new optimum in this case.