Network Optimization and Control

Srinivas Shakkottai
Texas A&M University
USA

R. Srikant
University of Illinois
Urbana-Champaign
USA
rsrikant@uiuc.edu

Full text available at: http://dx.doi.org/10.1561/1300000007
Editorial Scope

**Foundations and Trends® in Networking** will publish survey and tutorial articles in the following topics:

- Ad Hoc Wireless Networks
- Sensor Networks
- Optical Networks
- Local Area Networks
- Satellite and Hybrid Networks
- Cellular Networks
- Internet and Web Services
- Protocols and Cross-Layer Design
- Network Coding
- Energy-Efficiency
- Incentives/Pricing/Utility-based
- Games (co-operative or not)
- Security
- Scalability
- Topology
- Control/Graph-theoretic models
- Dynamics and Asymptotic Behavior of Networks

**Information for Librarians**

Foundations and Trends® in Networking, 2007, Volume 2, 4 issues. ISSN paper version 1554-057X. ISSN online version 1554-0588. Also available as a combined paper and online subscription.
Network Optimization and Control

Srinivas Shakkottai\textsuperscript{1} and R. Srikant\textsuperscript{2}

\textsuperscript{1} Texas A\&M University, USA
\textsuperscript{2} University of Illinois, Urbana-Champaign, USA, rsrikant@uiuc.edu

Abstract

We study how protocol design for various functionalities within a communication network architecture can be viewed as a distributed resource allocation problem. This involves understanding what resources are, how to allocate them fairly, and perhaps most importantly, how to achieve this goal in a distributed and stable fashion. We start with ideas of a centralized optimization framework and show how congestion control, routing and scheduling in wired and wireless networks can be thought of as fair resource allocation. We then move to the study of controllers that allow a decentralized solution of this problem. These controllers are the analytical equivalent of protocols in use on the Internet today, and we describe existing protocols as realizations of such controllers. The Internet is a dynamic system with feedback delays and flows that arrive and depart, which means that stability of the system cannot be taken for granted. We show how to incorporate stability into protocols, and thus, prevent undesirable network behavior. Finally, we consider a futuristic scenario where users are aware of the effects of their actions and try to game the system. We will see that the optimization framework is remarkably robust even to such gaming.
Contents

1 Introduction 1

2 Network Utility Maximization 5
   2.1 Utility Maximization in Networks 6
   2.2 Resource Allocation in Wireless Networks 12
   2.3 Node-Based Flow Conservation Constraints 15
   2.4 Fairness 16

3 Utility Maximization Algorithms 21
   3.1 Primal Formulation 22
   3.2 Dual Formulation 34
   3.3 Extension to Multi-Path Routing 38
   3.4 Primal–Dual Algorithm for Wireless Networks 41

4 Congestion Control Protocols 53
   4.1 TCP-Reno 55
   4.2 Relationship to Primal Algorithm 58
   4.3 A Generalization of TCP-Reno 59
   4.4 TCP-Vegas: A Delay Based Algorithm 60
   4.5 TCP-Vegas as a Resource Allocation Algorithm 62
   4.6 Relation to Dual Algorithms and Extensions 64
5  Network Stability  69
5.1  Stability of the Primal Algorithm with Delays  71
5.2  Stability Under Flow Arrivals and Departures  80

6  Game Theory and Resource Allocation  87
6.1  VCG Mechanism  88
6.2  Kelly Mechanism  91
6.3  Strategic or Price-Anticipating Users  93

Conclusions  101

Acknowledgments  105

References  107
The Internet has been one of the most conspicuous successes of the communications industry, and has permeated every aspect of our lives. Indeed, it is a testimony to its relevance in modern lifestyle that Internet connectivity is now considered an essential service like electricity and water supply. The idea that a best-effort data service could attain such significance was perhaps difficult to imagine a couple of decades ago. In fact, the Internet was initially envisioned as a decentralized data transmission network for military use. The argument was that if there were no centralized control, such as in a telephone network, and if much of the intelligence was at the edges of the system, it would make the network that much harder to destroy. Concurrent with the indestructibility requirements of the military was the need of scientific laboratories which required a network to exchange large data files of experimental results with each other. They envisioned high-speed links for transferring data between geographically distant data sources. The two requirements, coupled with statistical multiplexing ideas that illustrated the efficiency of using packetized data transmission gave rise to the Internet.
2 Introduction

As the network grew, it was clear that unrestricted data transfer by many users over a shared resource, i.e., the Internet, could be bad for the end users: excess load on the links leads to packet loss and decreases the effective throughput. This kind of loss was experienced at a significant level in the 1980s and was termed compression collapse. Thus, there was a need for a protocol to control the congestion in the network, i.e., control the overloading of the network resources. It led to the development of a congestion control algorithm for the Internet by Van Jacobson [34]. This congestion control algorithm was implemented within the protocol used by the end hosts for data transfer called the Transmission Control Protocol (TCP). Even though TCP is a lot more than just a congestion control algorithm, for the purposes of this review, we will use the terms “TCP” and “TCP congestion control algorithm” interchangeably.

Congestion control can also be viewed as a means of allocating the available network resources in some fair manner among the competing users. This idea was first noted by Chiu and Jain [13] who studied the relationship between control mechanisms at the end-hosts which use one-bit feedback from a link and the allocation of the available bandwidth at the link among the end-hosts. In fact, some of the details of Jacobson’s congestion control algorithm for the Internet were partly influenced by the analytical work in [13]. The resource allocation viewpoint was significantly generalized by Kelly et al. [41] who presented an optimization approach to understanding congestion control in networks with arbitrary topology, not just a single link. The purpose of this review is to present a state-of-the-art view of this optimization approach to network control. The material presented here is complementary to the book [89]. While the starting point is the same, i.e., the work in [41], this review focuses primarily on the developments since the writing of [89]. The purpose of this review is to provide a starting point for a mature reader with little background on the subject of congestion control to understand the basic concepts underlying network resource allocation. While it would be useful to the reader to have an understanding of optimization and control theory, we have tried to make the review as self-contained as possible to make it accessible to a large audience. We have made it a point to provide extensive
references, and the interested reader could consult these to obtain a deeper understanding of the topics covered. We hope that by providing a foundation for understanding the analytical approach to congestion control, the review will encourage both analysts and systems designers to work in synergy while developing protocols for the future Internet.

The review is organized as follows. We state the resource allocation objective in Section 2 and present the optimization formulation of the resource allocation problem. In Section 3, we will study decentralized, dynamic algorithms that solve the optimization problem. The section explains the framework for the design of such algorithms and proves the convergence of these algorithms. We then study current and recently proposed congestion control protocols, and their relationship to the optimization framework in Section 4. We then proceed to study the question of network stability in Section 5. We study two concepts of stability — that of convergence of algorithms to the fair allocation in the presence of feedback delays, and the question of whether the number of flows in the system would remain finite when flows arrive and depart. Finally, in Section 6, we study resource allocation from a game-theoretic viewpoint. In all the previous sections, it was assumed that the users cooperate to maximize network utility. In Section 6, we study selfish users whose goal is to maximize their own utility and their impact on the total network utility.


References


References


References


References

References


