Computer Networks 66 (2014) 32–45

Contents lists available at ScienceDirect

Computer Networks

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

Optimization-based resource allocation in communication networks ☆

G. Tychogiorgos *, K.K. Leung

Electrical and Electronic Engineering Department, Imperial College, London SW7 2AZ, UK

ARTICLE INFO

Keywords:
Resource allocation
NUM
Quality of experience
Non-convex optimization

ABSTRACT

The continuously growing number of applications competing for resources in current communication networks highlights the necessity for efficient resource allocation mechanisms to maximize user satisfaction. Optimization Theory can provide the necessary tools to develop such mechanisms that will allocate network resources optimally and fairly among users. The aim of this paper is to provide a starting point for researchers interested in applying optimization techniques in the resource allocation problem for current communication networks. To achieve that we, first, describe the fundamental optimization theory tools necessary to design optimal resource allocation algorithms. Then, we describe the *Network Utility Maximization (NUM)* framework, a framework that has already found numerous applications in network optimization, along with some recent advancements of the initial NUM framework. Finally, we summarize some of our recent work in the area and discuss some of the remaining research challenges towards the development of a complete optimization-based resource allocation protocol.

1. Introduction

Since the creation of ARPANET [1], the first packet switching communication network in 1969 that connected university laboratories, industrial and government research centers in the US, there has been a tremendous change in the extent and characteristics of communication networks, and especially the *internet*, the amount of data that is shared through them and the variety of applications that generate this traffic.

Recent Cisco IP traffic studies [2]3 provide useful information and insights regarding the traffic characteristics in current communication networks. The total internet traffic currently exceeds 40 Exa-bytes per month. On the other hand, mobile traffic has seen an explosive increase in the past years. While total mobile traffic in 2008 was no more than 33 Peta-bytes per month, mobile traffic is forecasted to reach 2.1 Exa-bytes per month by the end of 2013. The reason causing this abrupt increase in the traffic in both internet and mobile networks can be justified if one looks carefully at these statistics from another perspective; that of the applications that generate the traffic. In 2008, the majority of the traffic was generated by "traditional" types of applications, such as web browsing, email and file sharing applications, that accounted for 77% of the total traffic in the internet. However, multimedia applications, such as VoIP and video streaming, dominate the traffic nowadays

☆ Research was partially sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

* Corresponding author.

E-mail addresses: g.tychogiorgos@imperial.ac.uk (G. Tychogiorgos), kin.leung@imperial.ac.uk (K.K. Leung).

http://dx.doi.org/10.1016/j.comnet.2014.03.013
1389-1286/

exceeding 57% of the total traffic in the internet. The statistics are similar in the mobile internet as well, where the video traffic alone currently accounts for two-thirds of the total mobile traffic.

This abrupt increase of the total traffic highlights the necessity for more efficient methods of sharing the available bandwidth so that users are receiving the maximum possible satisfaction and the best possible experience when using a communication network. In addition, taking into account that users are being charged by the network providers for access, the more efficient the resource allocation is, the more satisfied the users will be and consequently the more willing to continue paying the provider for the service. The heterogeneity of the provided applications also shows that all traffic does not have the same resource requirements. This strengthens the need for more sophisticated allocation methods that will be able to distinguish between different types of applications and try to allocate resources in a way that maximizes user satisfaction for each application type.

Optimization Theory can provide a powerful tool in the development of such methods for various reasons. Optimization Theory has been used successfully in many areas related to communication networks, such as optimal routing, flow control and power control. In particular, pioneering and early work that is based on optimization techniques on link capacity assignment, routing and flow control in communication networks can be found in, e.g., [4–6]. Besides these, optimization theory is also widely applied to other applications, such as chemical engineering [7], fleet management and inventory organization, since it leads to the best possible solutions for a given problem. In addition, there are techniques, such as the *Langrangian Method* that can lead to the development of distributed algorithms. Distributed calculation of the optimal solution is of significant importance in communication networks, which consist of numerous network nodes and traffic sources that behave independently and selfishly to achieve the maximum possible level of satisfaction using the resources of the network. Moreover, optimization theory can also be used to assure that the allocation of resources to each application will follow its *Quality of Service* requirements and satisfy some notion of fairness. This can be achieved by the appropriate formulation of the optimization problem and the use of specific allocation policies according to the desired type of fairness.

The purpose of this paper is twofold; first, to introduce *Lagrangian* Optimization as a technique to solve the resource allocation problem in communication networks, and then, to provide some samples of recent work in the area and outline some of the most important challenges that need to be tackled in the future. The rest of the paper is organized as follows. Section 2 provides a summary of the most important notions of Optimization Theory, necessary for someone to do research in the area. Then, Section 3 describes the Network Utility Maximization (NUM) framework and the most important enhancements of the initial NUM framework. Then, Section 4 provides an overview of our recent research in the area as a motivation to tackle some of the remaining open research challenges, which are summarized in Section 5.

2. Optimization theory overview

This section provides a brief description of the basic notions in *Optimization Theory*, based on textbooks [8,9]. The main focus of this overview is on the areas of function and problem *convexity*, optimization problem formulation, as well as on the advantages that distributed optimization techniques can offer to solve such problems. The interested reader is referred to the aforementioned textbooks for a complete presentation and analysis of *Optimization Theory*.

A set C is a *convex* set if it is a subset of $\Re^n$ and if $\alpha x + (1-\alpha)y \in C, \forall x, y \in C$ and $\forall \alpha \in [0,1]$. In accordance, a function $f : C \to \Re$, where C is a convex subset of $\Re^n$, is *convex* if

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y), \quad \forall x, y \in C, \quad \forall \alpha \in [0,1]. \tag{1}$$

An intuitive interpretation of (1) is that the line segment between $(x, f(x))$ and $(y, f(y))$, which is the chord from x to y, lies always above the graph of f. On the other hand, a function f is called *concave* if the direction of the inequality in (1) is the opposite. Relating *convex* and *concave* functions one can comment that if f is a *convex* function, then $-f$ is *concave* and vice versa. In addition, function f is called *strictly convex* if inequality (1) is strict for all $x, y \in C$ with $x \neq y$, and for all $\alpha \in (0,1)$. In general, *convex* functions are convenient for minimization problems since their local minima are also global and equivalently *concave* functions are convenient for maximization problems.

To represent an optimization problem, we use the notation

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h_i(x) \leqslant 0, \quad i = 1, \ldots, m \\ &f_i(x) = 0, \quad i = 1, \ldots, p \end{aligned} \tag{2}$$

in order to describe the problem of finding the value of the *optimization variable* x that will minimize the *objective function* $f(x)$ among all possible values of the variable so that the conditions $h_i(x) \leqslant 0, i = 1, \ldots, m$ and $f_i(x) = 0$, $i = 1, \ldots, p$ are all satisfied. A general problem formulation such as the one presented in (2) can have a number of locally and globally optimal solutions. *Global Optimization* [10]11 is the area of Optimization Theory interested in calculating the globally optimal solutions of an optimization problem that will minimize the value of the objective function $f(x)$ within the feasible region. In addition, research in Global Optimization is also interested in the feasibility characterization of optimization problems and in determining upper and lower bounds of their objective functions [12]. *Convex Optimization* is a specific area of Global Optimization where locally optimal solutions are also globally optimal. An optimization problem with the latter property is called a Convex Optimization problem. More specifically, an optimization problem such as (2) is called *convex* if the following conditions hold: (a) the objective function $f(x)$ is a convex function of the optimization variable x; (b) the inequality constraint functions $h_i(x)$ are convex; and (c) the equality constraint functions $f_i(x)$ are affine.