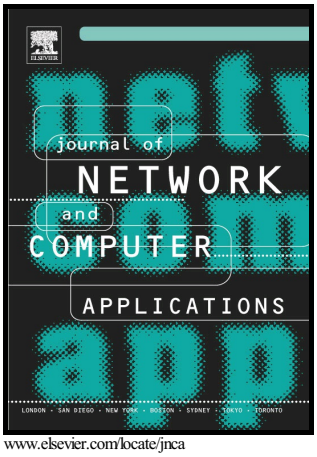
Bandwidth scheduling for big data transfer using multiple fixed node-disjoint paths

Aiqin Hou, Chase Q. Wu, Dingyi Fang, Yongqiang Wang, Meng Wang



www.elsevier.com/locate/jnca

PII: S1084-8045(16)30302-2
DOI: http://dx.doi.org/10.1016/j.jnca.2016.12.011
Reference: YJNCA1794

To appear in: *Journal of Network and Computer Applications*

Received date: 25 July 2016
Revised date: 21 November 2016
Accepted date: 2 December 2016

Cite this article as: Aiqin Hou, Chase Q. Wu, Dingyi Fang, Yongqiang Wang and Meng Wang, Bandwidth scheduling for big data transfer using multiple fixed node-disjoint paths, *Journal of Network and Computer Applications* http://dx.doi.org/10.1016/j.jnca.2016.12.011

Bandwidth Scheduling for Big Data Transfer Using Multiple Fixed Node-Disjoint Paths

Aiqin Hou[a], Chase Q. Wu[a,b,*], Dingyi Fang[a], Yongqiang Wang[a], Meng Wang[a]

[a]*School of Information Science and Technology, Northwest University, Xi'an, Shaanxi 710127, China*
[b]*Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey 07102, USA*

Abstract

Many large-scale applications require the transfer of big data over high-performance networks for remote operations. Such requirements call for a fast bandwidth scheduling solution to discover feasible and efficient reservation options in network environments with time-varying bandwidths. We formulate a generic problem of Bandwidth Scheduling with Two Node-Disjoint Paths (BS-2NDP) to support big data transfer. In BS-2NDP, we further consider two different types of paths: i) two fixed paths with fixed bandwidth (2FPFB), and ii) two fixed paths with variable bandwidth (2FPVB). We prove that both 2FPFB and 2FPVB are NP-complete, and design a heuristic approach for each of them. We implement and evaluate these scheduling algorithms in both simulated and real-life networks. Extensive results show that the proposed heuristics achieve a close-to-optimal performance in small-scale networks, and significantly outperform other heuristic approaches in large-scale networks.

Keywords: Big data, high-performance networks, bandwidth scheduling, node-disjoint paths

1. Introduction

Many large-scale applications in various domains such as e-science, e-business, and social media are now producing large amounts of data on a daily basis, which must be transferred over wide geographical areas for various remote operations [2, 3]. Typical examples include next-generation computational sciences where large simulation datasets produced on supercomputers are shared by a distributed team of scientists for collaborative visualization and analysis. Fast and reliable data transfer has become a critical task to ensure the success of these applications. Unfortunately, the sheer volume of data generated in such applications has gone far beyond the capability of traditional shared IP networks. In recent years, high-performance networks (HPNs) that provision dedicated channels through bandwidth reservation have emerged as a promising solution and their significance has been well recognized in broad science and network research communities [4, 5, 6].

As the central function unit of a generalized control plane for provisioning dedicated channels in HPNs, the bandwidth scheduler computes appropriate network paths and allocates link bandwidths to meet specific user requests based on network topology and bandwidth availability. To meet the unprecedented requirement of big data movement, it is a natural extension from single-path to multi-path transfer, which is generally more effective in terms of throughput, robustness, load balance, and congestion reduction. However, multi-path routing also introduces some extra overhead to both the control plane and the data plane of a network [7].

The complexity of multi-path routing varies depending on the type and number of constraints, and many of these routing problems are NP-complete. Several studies have shown that the Multiple Constrained Path (MCP) problems are generally NP-complete and hence are not solvable in polynomial time [8, 9]. Furthermore, finding disjoint paths with a single constraint is also an NP-hard problem [10, 11, 12]. Multiple paths usually have an additional constraint to be link-disjoint or node-disjoint. Node-disjoint paths are usually harder to find but provide more robustness in case of node failures. The two-path routing problem with reliability consideration is NP-hard in the strong sense, as opposed to the ordinary NP-completeness of the single-path problem [13].

In this paper, we formulate a generic problem of Bandwidth Scheduling with Two Node-Disjoint Paths (BS-2NDP) to support big data transfer. In BS-2NDP, we further consider two different types of paths: i) two fixed paths with fixed bandwidth (2FPFB), and ii) two fixed paths with variable bandwidth (2FPVB). We prove that both 2FPFB and 2FPVB are NP-complete, and propose a heuristic approach for each of them. We implement and evaluate these algorithms in both simulated and real-life networks. Extensive results show that the proposed heuris-

*Corresponding author, Email: chase.wu@njit.edu, Tel: 1-973-642-4579.

Some preliminary results in this manuscript have been published in QShine 2016[1].

Email addresses: houaiqin@nwu.edu.cn (Aiqin Hou), chase.wu@njit.edu (Chase Q. Wu), dyf@nwu.edu.cn (Dingyi Fang), yqwang@nwu.edu.cn (Yongqiang Wang), xidawm@stumail.nwu.edu (Meng Wang)