Computer Communications 77 (2016) 26–40

Contents lists available at ScienceDirect

Computer Communications

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

Deadline-aware advance reservation scheduling algorithms for media production networks

Maryam Barshan[a,*], Hendrik Moens[a], Jeroen Famaey[b], Filip De Turck[a]

[a] Department of Information Technology, Ghent University – iMinds, Gaston Crommenlaan 8/201, Gent B-9050, Belgium
[b] Department of Mathematics and Computer Science, University of Antwerp – iMinds, Middelheimlaan 1, Antwerpen 2020, Belgium

ARTICLE INFO

Article history:
Received 10 April 2015
Revised 27 August 2015
Accepted 31 October 2015
Available online 10 November 2015

Keywords:
Advance bandwidth reservation
Media production network
Video streaming
Deadline-aware scheduling

ABSTRACT

In the media production process a substrate network can be shared by many users simultaneously when different media actors are geographically distributed. This allows sophisticated media productions involving numerous producers to be concurrently created and transferred. Due to the predictable nature of media transfers, the collaboration among different actors could be significantly improved by deploying an efficient advance reservation system. In this paper, we propose a model for the advance bandwidth reservation problem, which takes the specific characteristics of media production networks into account. Flexible and time variable bandwidth reservations, meeting delivery deadlines, supporting splittable flows and interdependent transfers and all types of advance reservation requests imposed by the media production transfers are incorporated into this model. In addition to the optimal scheduling algorithms, which are presented based on this model, near optimal alternatives are also proposed. The experimental results show that the proposed algorithms are scalable in terms of physical topology and granularity of time intervals and obtain a satisfactory performance, executing significantly faster than an optimal algorithm and within 8.78% of the optimal results.

1. Introduction

In the media production industry, a team of artists, editors, reporters and producers works simultaneously at geographically distributed locations producing and processing content, music, commentary, special effects, etc. Various producers and actors could then access these individual elements over a shared network to integrate them and thereby produce a complete product. In the media creation process, reliability of the transport is of crucial importance.

Predictability is a key feature of traffic in media production networks. Traffic characteristics in terms of bandwidth requirements, the time when the contents are ready, and the deadline for the data to be completely transferred to the destinations, are mostly known several hours ahead of time. The predictable nature of these transfers makes it possible to use resource reservation techniques. Therefore, a management system can efficiently manage the transmission. In general, two types of resource reservation can be distinguished [1]: Immediate Reservation (IR) and Advance Reservation (AR). While just-in-time reservation is applied in IR, the principle behind AR relies on the resource reservation times before the actual time when the resource is used. Assuming prior knowledge of the network structures and different requests, advance reservation makes it possible to schedule network requests optimally.

In computer networks, bandwidth is a valuable resource. Particularly for multimedia transfers, where large amounts of content, such as video files, have to be transmitted, efficient bandwidth management is an important factor [2]. In bandwidth-limited networks, an efficient bandwidth reservation mechanism needs to be defined to meet the QoS requirements and deadlines. The next generations of media production networks are expected to efficiently support advance reservation systems for different delivery services, so the desired QoS requirement and resource utilization could be ensured.

In this paper, we propose a set of novel AR scheduling algorithms, optimized for media production networks. Such networks impose requirements not supported by existing AR scheduling techniques. First, the start time of requests is generally flexible, the deadline is fixed, and the reserved bandwidth may vary over the lifetime of the reservation. This combination of flexible start times and elastic bandwidth allocation has not received much attention in research to date [3]. Second, in media production networks, multiple requests may depend on each other. Until now to the authors' knowledge, this aspect has remained unexplored. Third, it should be possible to split requests over multiple paths, in order to further optimize bandwidth utilization.

We propose a centralized advance reservation system which based on our evaluation scales to the size of realistic media

* Corresponding author. Tel.:+32 485975066.
E-mail address: maryam.barshan@intec.ugent.be (M. Barshan).

http://dx.doi.org/10.1016/j.comcom.2015.10.016
0140-3664/

production networks and demand patterns. We present a model to solve this variant of the AR scheduling problem, and propose various advance reservation algorithms based on our designed model. This model is an instance of time variable scheduling which is known as multicommodity over time problem. In this context, commodity corresponds to a telecommunication traffic demand between two media actors. It has been proven that the complexity of multicommodity flow over time without caching is strongly NP-hard [4]. Therefore, we also came up with efficient and near-optimal heuristic solutions which are more practical. In both optimal and heuristic approaches the main goal is threefold: (1) delivery of the requests before their deadline; (2) maximizing the number of admitted requests; (3) processing requests as quickly as possible.

Both approaches can be used in static and dynamic settings. The Static Advance Reservation Algorithm (SARA) assumes all requests are known at the start of the reservation period. By contrast, the Dynamic Advance Reservation Algorithm (DARA) supports rescheduling in order to incorporate new requests at runtime. The dynamic advance reservation system tries to admit new arrival requests while rescheduling the previously admitted requests is a must. We provide a thorough analysis of the algorithms based on in-depth simulation results. They are compared and the impact of their parameters on the solution quality is evaluated.

The remainder of this paper is organized as follows. Section 2, reviews the related work. Section 3 describes the scenarios as well as architecture and components of the proposed media production network. Section 4 explains the concepts, assumptions and AR scheduling formulation for media production networks. The proposed algorithms are explained in Section 5. A comparison of optimal and near-optimal algorithms and a performance evaluation of offline and online settings are provided in Section 6. Finally, Section 7 concludes the paper.

2. Related work

There has been a large number of theoretical as well as practical experimental work [5–8] related to the advance reservation problem. Here we study some of the most relevant work. The authors in [9] and [10] focus on re-routing in advance reservation networks. Our formal model is inspired by their ILP-based solutions called GILP and DILP [9]. The GILP assumes that the entire set of requests is known beforehand and the DILP is designed to work in an online setting. However our approach is different as their models assume only streaming requests with fixed time intervals and dedicated bandwidth remains fixed and equal to the demand during the entire reservation. Dependencies among the requests are also ignored. [10] is the extension of [9] in realistic multi-domain networks which addresses the implementation challenges related to advance reservation solutions.

Charbonneau and Vokkarane [3] survey the literature on advance reservation routing and scheduling algorithms, specifically focused on WDM networks. It has defined four types of advance reservation requests based on whether their start time and their duration are specified or not. All these classifications, which will be discussed in detail later, are supported in our approach. In addition, our approach is elastic which means that the allocated bandwidth is variable over time. According to this reference only two AR scheduling algorithms have been proposed that support elastic reservations [11,12]. However, they both assume a fixed start time. Sharma et al. [13] present an algorithm called RRPC which addresses multiple flexible requests for bandwidth reservation between two end points. RRPC is deadline-aware in which any reservation that meets the deadline is acceptable. However all the requests have a same source and destination, flow splitting is not allowed and a single path is chosen for all the requests. Another work [14] focuses on dynamically transporting of large volume of data in e-science networks. The optimization consists of two steps admission control and scheduling. Periodically the central controller gathers all the new requests, runs admission control, and then schedules new and unfinished jobs.

Furthermore, the problem addressed in this work is related to the multicommodity flow problem (MCFP). The multicommodity flow problem can be described as follows: a set of individual flows have to be transferred in a dimensioned network without violating the capacity limits [15]. The resource allocation algorithm should find an optimal routing path to transfer the flows through the network. In [16], unsplittable flow and single path MCFPs are studied. Comprehensive surveys on the approaches to solve multicommodity flow problems (MCFP) and their variants are provided in [17,18]. Our approach further deals with the problem of flow variation over time and solves an MCFP as a subproblem. In network flow problems, having variable flows over time is crucial. Dynamic flows or flow variation over time are primarily introduced by Ford and Fulkerson [19,20]. They introduced variable flows over time as equal as static flow problem, building another temporal dimension over the network. This makes use of time-expanded networks. A time-expanded network is a copy of network in each discrete time step. Also, Fleischer and Skutella [21] have mentioned that in literature hardly any results on multicommodity over time are noted.

Chen et al. in [22] have stated that the single-path approach on which the Internet routing protocols is based, could not meet the delay requirements when the video streams are transferred over bandwidth-limited networks. They proposed a multipath routing of video contents over bandwidth limited network. However the main focus of their work is on delay and over the Internet, and therefore no reservation is considered.

Balman et al. [23] have focused on advance bandwidth reservation for on-demand data transfer in scientific applications. However, their work differs from our approach as they purely focus on data transfers, not video streaming sessions, and the routing mechanism is based on single-path in contrary to our multi-path approach. In addition, our approach considers dependencies among requests. To the best of our knowledge, dependencies among requests which is explicitly incorporated in our algorithms, have not received adequate attention in the literature by state of the art approaches.

This work is an extension of our previous work [24], in which only the static and dynamic ILP-based models are introduced. The main focus of this previous work was to investigate the viability of AR mechanisms in media production networks and to find the optimal solutions, determining the most appropriate objective function in our optimal models. We defined two objective functions and compared their performance. We found that the so called ASAP objective function, which in addition to maximizing the number of admitted requests also tries to schedule requests in earlier time slots leads to better results. In this article several new heuristic approaches are proposed which are near-optimal and computationally less-complex compared to ILP-based approaches. In our evaluation, their performance is compared with the highest quality optimal algorithms (i.e., algorithms based on the ASAP objective function).

3. Media production network architecture

The envisioned media production network is depicted in Fig. 1. The different actors and locations involved in the media production process, such as for example recording studios, on-site filming crews, broadcasters, and storage datacenters, are connected to a shared wide-area network, consisting of interconnected switches. The network supports the exchange of raw and encoded multimedia content between an arbitrary set of actors, both in the form of file transfers and streaming. The management layer provides a reservation interface that allows the users of the network to reserve bandwidth over certain time periods in the future. The AR scheduling algorithms are responsible for reserving the required amount of bandwidth resources for all requests. With each request, they associate one or