Journal of Network and Computer Applications 80 (2017) 109–122

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Design and evaluation of a dual dynamic adaptive reservation approach in media production networks

Maryam Barshan*, Hendrik Moens, Bruno Volckaert, Filip De Turck

Department of Information Technology, Ghent University – iMinds Technologiepark-Zwijnaarde 15, 9052 Gent, Belgium

ARTICLE INFO

Keywords:
Advance bandwidth reservation
Media production network
Runtime adaptation
Network failure handling

ABSTRACT

Media production networks require an efficient collaboration between geographically distributed actors and offer predictable workloads, making it possible to exploit this predictability and use advance bandwidth reservation services to achieve greater bandwidth utilization and service guarantees. To offer reliable reservations, the incorporation of fault-tolerance related features in bandwidth reservation strategies is a necessity, although this imposes a waste of capacity and extra performance overhead. As a first provisional stage to offer a robust reservation system, deploying protection mechanisms ensures that the reservations remain valid when the system is in operation. To maximally utilize the network, and to ensure there is a quick response in a dynamic network environment, constant monitoring and optimization is needed. In this article, we propose an efficient dual optimization approach consisting of two consecutive processes. First, a schedule is produced by a resilient advance reservation algorithm. Then, the generated schedule is continually updated over time using a runtime adaptation approach in order to be capable of dynamically adapting the network to changing conditions and mitigating the side effects of provisioned reliability. This step uses the interconnecting network links' leftover capacity, resulting in an increased performance both in steady and unsteady network conditions. Our evaluations show that in failure-prone environments, the proposed approach leads to significant increase in the success rate of admitted requests, up to 6.77 times, compared to the resilient advance reservation algorithms.

1. Introduction

Media production processes have become more complex and more data/network-intensive as they are increasingly dealing with high bitrate videos, deadline-constrained network transfers and geographically distributed media production teams. Large quantities of data must be processed by multiple collaborating parties at different geographical locations. Media production environments are highly dynamic due to the arrival and departure of several requests of different sizes and requirements. In order to provide high-performance collaboration between different sites, next generation network reservation systems have to provide predictable performance and efficient bandwidth utilization. To ensure that bandwidth needs, delivery deadlines and requirements of different transfers are met, advance bandwidth reservation is needed. In general, advance reservation benefits the network operators as knowledge of future transmissions can be used to improve the admission control and provisioning to increase network utilization. It is also beneficial for the user as the network can provide better QoS to (future) requests with declared arrival and holding times (Charbonneau and Vokkarane, 2012), guaranteeing that the needed network capacity will be available. Advance reservation approaches can be either static or dynamic. While in a static approach all requests are known before scheduling, requests arrive one by one over time in a dynamic model.

In the media production industry, advance reservation scheduling of network transfers (Rajah et al., 2009) is very important in order to make correct decisions on rejection or acceptance of future requests. In uncertain network conditions, such as sudden changes in network configuration, network fluctuations, failures, etc., additional precautions must be taken to guarantee successful transfers. The reliability of transfers in the media production networks is of prime importance and it can be enhanced using protection mechanisms. However, there are arguments against this redundancy as a large portion of network capacity will be wasted if the capacity assigned for this redundancy cannot be reused. As such, making use of idle network capacity and updating the resilient schedule over time, based on the current state of the network and running and planned transfers is of great advantage.

This work has been performed within the context of ICON MECaNO project (ICON MECaNO project, Jan 2014–Dec 2015), which provides solutions for the transmission of large media contents over an

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

http://dx.doi.org/10.1016/j.jnca.2016.12.003
Received 9 April 2016; Received in revised form 23 November 2016; Accepted 2 December 2016
Available online 07 December 2016

IP-based infrastructure, tailored to the quality and timing requirements of current and future media production process requests. In our previous work (Barshan et al., 2015) and Barshan et al. (2016), we proposed both static and dynamic advance reservation scheduling approaches for a couple of interdependent requests of two types, video streams (VS) and file-based video transfers (FB). We have further presented the resilient version of these approaches based on a protection mechanism to improve the reliability of the advance reservation system (Sahhaf et al., 2016). The proposed scheme is capable of covering single link failures using pre-reserved disjoint backup paths. In this article, we make a tradeoff between reliability and resource usage in 3 ways: 1) the percentage of redundancy is defined for each individual request based on an input parameter provided by the customer, to influence the importance of reliability for each individual connection, 2) shared backup path protection (SBPP) (Józsa and Orincsay, 2001; Haahr et al., 2012) is used, significantly reducing the bandwidth requirements for backup purposes, and 3) redundant reservations and network leftover capacity are reused as long as those are not being used for their primary purpose.

The main contribution of this paper is to make use of backups and idle bandwidth capacities to push more data into the network as long as advance reservations are partially unused as well as rapid reaction to sudden changes in uncertain network conditions using an event-based approach. Based on the resilient advance reservation approach, backups are ready for use, but are only activated when failures occur, leaving capacity unused. In addition, we have found that reservations made for video streams, are not completely utilized throughout the requested time. Video streams can be resumed and played-back multiple times during the reserved period, which causes idle reservations between resumes and playbacks. In our proposed approach, these unused capacities can be exploited to transfer additional data. This means that we use these reserved capacities as double-purpose, prioritizing their original purpose. In doing so, as long as these reserved capacities are idle, additional data can be transferred and as soon as for example a video stream becomes active, an event will be raised to prioritize the advance reservation made for this streaming request over the extra data transfers.

The proposed approach consists of two sequential processes. First, the network and transfers status are being continually monitored and the advance reservations are periodically updated. Second, the backup and unused network capacities, e.g. unused video stream reservations, are re-utilized to transfer more data than the schedule made by the resilient advance reservation algorithms. In unreliable networks, as soon as any failure is detected, an event will trigger our proposed algorithm to adapt the ongoing network transfers according to the current state of the network. This leads to a better utilization of substrate network resources, higher success rate and rapid reaction to sudden changes when the network is in operation.

The rest of this article is organized as follows. Section 2, describes background and related work. The envisioned media production network and the runtime adaptation approach are explained in Section 3. The proposed algorithms are described in Section 4. Section 5 provides simulation results and Section 6 concludes the article.

2. Related work

2.1. Advance resource reservation

Advance network resource reservation has applications for both wide-area and grid networks and has been studied frequently in recent years (Depoorter et al., 2014; Bochenina et al., 2016; Bai et al., 2016; Guok, 2005; Gibbard et al., 2006; Chunlin et al., 2009; Gu et al., 2011). Current research mostly focuses on optical networks in combination with wavelength division multiplexing (Charbonneau and Vokkarane, 2012). Advance reservation requests can be classified in 4 individual categories (Charbonneau and Vokkarane, 2012). This classification is also valid for different types of requests in media production networks and all classes are supported in our work. In optical networks, the static advance reservation problem is first introduced by Kuri et al. (2002), Kuri et al. (2003), who focus on requests with specified start time and duration and proposed heuristics and meta-heuristics to solve the static problem. The authors in Zheng and Mouftah (2001), Zheng and Mouftah (2002) were the first to propose dynamic advance reservation in fixed time-slotted networks. By introducing the percentage of known requests in Barshan et al. (2015), both static and dynamic traffic models are considered in our approach. Xie et al. in Xie et al. (2012) proposed ILP-based models and heuristic approaches on re-routing in advance reservation networks in order to maximize admittance of new requests. The authors in Balman et al. (2010), Balman (2013) focused on advance bandwidth reservation for on-demand data transfer in scientific applications. These approaches, however, purely focus on data transfers, not video streaming requests, dependencies among different transfers are ignored and no fault tolerance techniques are considered for possible failures.

2.2. Resilient reservation

Adding resilience into a reservation system can be achieved through restoration or protection failure recovery mechanisms (Watanabe et al., 2015). In Tanwir et al. (2008) a resilient advance reservation mechanism is proposed in optical grids. Due to the lower cost of restoration mechanisms, they use the latter. Burchard et al. in Burchard et al. (2008) consider a recovery mechanism for advance reservations in grid environments. The idea is to re-schedule failed but unstarted requests whenever failure occurs, but the main focus is on estimating the downtime. The authors in Latchoumy et al. (2013) have also focused on a proactive approach by taking resource statistical failure information into account. Their method relies on failure prediction and avoiding vulnerable resources. The authors in Nazir et al. (2012) present a fault-tolerant job scheduling approach for grid environments using adaptive task replication, which is a recovery approach. Providing resiliency in optical WDM networks through shared path protection has been proposed in Chołda and Jaglarz (2016), Cavdar et al. (2009), Wang and Li (2010), Jaekel et al. (2010). Since meeting strict deadlines and QoS requirements is of great importance in our approach, using protection mechanisms tends to be more reliable.

2.3. Media production networks

The work presented in this article consists of two complementary approaches for media production networks. The combination of a customized resilient AR approach with a highly dynamic event-driven runtime adaptation approach consists of several functions which are of essential importance in the considered media production networks and have not been previously studied in the context of advance reservations.

This work proposes a dual approach which partially makes use of our previous works (Barshan et al., 2016) and Sahhaf et al. (2016). In Barshan et al. (2016), we devised an Integer Linear Programming (ILP) based model and proposed heuristic approaches for an exact solution. We showed that the heuristics yielded favorable results in much less time complexity than the linear programs. In Sahhaf et al. (2016), we enhanced the media production reservation system and made it more reliable in case of failures by following a protection mechanism and provisioning backup reservations for each request. As redundancy imposes cost and resource waste, the main motivation of our approach is to mitigate the side-effect of using redundant reservations by employing underutilized network capacities for transferring extra data as long as those are not needed for redundancy purposes. This work in the extension of Barshan et al. (2016) in which design of the proposed approach is explained in depth and Barshan et al. (2016) where the the initial evaluation of our proposed approach without considering the