

An SLA-Driven Framework for Dynamic Multimedia Content Delivery Federations

Jeroen Famaey, Steven Latré, Tim Wauters, and Filip De Turck

Department of Information Technology
Ghent University – IBBT

Gaston Crommenlaan 8/201, B-9050 Gent, Belgium
Email: jeroen.famaey@intec.ugent.be

Abstract—Recently, the Internet has become a popular platform for the delivery of multimedia content. However, its best effort delivery approach is ill-suited to guarantee the stringent Quality of Service (QoS) requirements of many existing multimedia services, which results in a significant reduction of the Quality of Experience. This paper presents a solution to these problems, in the form of a framework for dynamically setting up federations between the stakeholders involved in the content delivery chain. More specifically, the framework provides an automated mechanism to set up end-to-end delivery paths from the content provider to the access Internet Service Providers (ISPs), which act as its direct customers and represent a group of end-users. Driven by Service Level Agreements (SLAs), QoS contracts are automatically negotiated between the content provider, the access ISPs, and the intermediary network domains along the delivery paths. These contracts capture the delivered QoS and resource reservation costs, which are subsequently used in the price negotiations between content provider and access ISPs. Additionally, it supports the inclusion of cloud providers within the federations, supporting on-the-fly allocation of computational and storage resources. This allows the automatic deployment and configuration of proxy caches along the delivery paths, which potentially reduce delivery costs and increase delivered quality.

I. INTRODUCTION

The increasing availability of bandwidth has made modern communication networks into popular platforms for the delivery of multimedia services, such as time-shifted television and video on demand (VoD). Many Internet Service Providers (ISPs) have started offering these IPTV-based services (e.g., Verizon, Comcast). Additionally, so called over-the-top (OTT) content providers have started offering Video-on-Demand and live television services to users across the Internet (e.g., Hulu, Netflix). Both approaches have their distinct advantages.

As ISPs offer their multimedia services to users within their own network, they are able to give guarantees concerning the delivered Quality of Service (QoS). However, the operator needs to compose its own content catalogue, acquiring licenses from copyright owners. Moreover, it can serve this content only to a limited number of potential customers. Therefore, the operator will have to choose between limiting costs by offering only recent and popular content, or satisfying all its customers by providing an extensive catalogue that caters to all tastes.

On the other hand, OTT content providers can offer their services to a potentially huge number of users across the Internet. However, as the current Internet is a best-effort network, they cannot offer any QoS guarantees. Additionally, the content is delivered over the ISP's infrastructure, without it being involved in the control or distribution of the content. This causes the ISP network to become more heavily loaded, while it does not share in the revenues.

In this paper, we propose an approach to overcome the shortcomings of current multimedia content delivery methods. It facilitates the collaboration between content providers and ISPs, in order to guarantee end-to-end QoS. More specifically, we propose a framework for setting up end-to-end network federations (also called alliances). A federation is defined as a collaboration between network domains, in order to deliver a combined service, or set of services [1]. Federations are set up between a content provider, a set of intermediary core Internet domains and an ISP. This allows the content provider to offer its multimedia content across the Internet, while guaranteeing end-to-end QoS to the end-users. The intermediary core Internet domains and ISPs share in the revenues in return for offering bandwidth- and QoS-guaranteed paths through their networks. Additionally, in contrast to existing work, our approach allows cloud providers to join the federation. Their on-demand computational and storage resources can be employed to dynamically deploy proxy caches along the delivery paths in order to reduce resource consumption.

The proposed framework guides the interactions between the stakeholders involved in the envisioned content delivery process. Their obligations and requirements are captured in Service Level Agreements (SLAs). This includes the QoS requirements of the ISPs, and the price paid to the content provider. Through SLA negotiation protocols, the terms of the SLA can be negotiated. An important aspect in these negotiations, which is often neglected in existing work, is the bounds within which the different parties are willing to negotiate. To address this concern, we propose a mathematical model to calculate the total cost for delivering content from the content provider to the ISP. This allows the content provider to calculate an initial price offer to be used in the negotiations, as well as the limit to which it is willing to lower its price.

This work presented in this paper is based on our earlier work [2]. This previous work focussed on an algorithm to solve a static version of the cost model presented here. In contrast, this paper focusses more on the architectural components of our framework, such as the negotiation of SLAs and the interactions between stakeholders. Additionally, the cost model presented here is an extension of our earlier work, as it is more generic and supports dynamic paths through the Internet core.

The remainder of this paper is structured as follows. Section II describes related work in the context of end-to-end QoS-aware path selection algorithms and SLA negotiation. A detailed description of our envisioned framework is provided in Section III. It identifies and defines the stakeholders involved in the content delivery federations, as well as the interactions that take place between them. Subsequently, Section IV goes into more details about setting up the end-to-end delivery paths, and presents a mathematical model for calculating the delivery costs for the content provider. This cost can be used as a basis for the negotiation of the SLA between the provider and ISPs, which is further described in Section V. Section VI provides some details about our prototype implementation. Finally, Section VII concludes the paper.

II. RELATED WORK

Providing end-to-end QoS guarantees for the delivery of services across the Internet, has been an important topic of study for several years. Recently, interest has shifted from intra-domain QoS-based service provisioning towards inter-domain QoS-aware federations. The latter focusses on the collaboration across network domains to achieve end-to-end QoS guarantees. So far, work has mostly focussed on the negotiation of Service Level Agreements (SLAs) and QoS between federation partners. Several evolutionary approaches have been proposed to support end-to-end QoS on top of the current best-effort Internet. Kumar *et al.* presented the Alliance network model [3]. It allows interconnected Autonomous Systems (AS) to form an alliance, which enables optimal inter-domain path selection and QoS guarantees. Additionally, it is compatible with the Border Gateway Protocol (BGP) and can thus co-exist with the current best-effort Internet. Xiangjiang *et al.* presented a similar approach, also based on BGP [4]. More closely related to the work presented in this paper, Pouyllau and Douville proposed an algorithm for negotiating a QoS-guaranteed path across the Internet through setting up network federations [5], [6]. They use a game theory approach combined with Q-learning in order to find paths that satisfy the QoS requirements, while maximizing long-term objectives.

Recently, Balasubramaniam *et al.* proposed an integrated architecture combining service lifecycle management and dynamic routing [7]. The proposed management framework is inspired by biological systems, and aims to autonomously configure services and the underlying routing system in order to guarantee the ever-changing end-to-end QoS requirements of a large number of heterogeneous services. Although their research has similar goals to ours, they focus on service

lifecycle management and dynamic, distributed routing. In contrast, our work centers around the negotiation of SLAs and long-term cost minimisation through intelligent resource reservations.

In addition to identifying and configuring the resources that are necessary to deliver services under specific quality constraints, the terms of the agreement should be negotiated before a federation can be successfully set up. To achieve this, SLA negotiation protocols can be employed. Several frameworks and architectures have been proposed to support SLA negotiation between federation partners. Yuanming *et al.* developed a framework for the negotiation of SLAs between service providers, network operators and content providers [8]. More recently, the SLA-based SERVICEable Metacomputing Environment (SERVME) was proposed [9]. It consists of a framework and accompanying SLA model, which guide the SLA negotiation process, match providers based on QoS requirements and perform on-demand resource provisioning. In addition to frameworks that support the negotiation and management of SLAs, several protocols that allow the actual SLA terms to be negotiated have been proposed. The Web Services Agreement Specification (WS-Agreement) is a Web Services protocol for establishing agreement between two parties, such as a service provider and consumer [10]. It consists of an XML-based language for specifying SLAs, as well as a protocol for negotiating its terms [11]. Hasselmeyer *et al.* proposed a Discrete-Offer-Protocol that allows the service provider to make a single offer, which the consumer can accept or reject [12]. More recently they proposed a more elaborate protocol [13]. It supports multi-round negotiations, as well as re-negotiating the terms of an SLA already in place as the requirements of participants change.

Our work differs from existing work on end-to-end QoS provisioning in that our aim is to support more complex federations. In addition to federations with network domains that provide network resources, our framework supports the inclusion of clouds, which provide storage and computational resources. The actual SLA negotiation protocols are outside the scope of our work. Instead we focus on determining the bounds within which the terms, such as price, should be negotiated.

III. FEDERATED CONTENT DELIVERY FRAMEWORK

The goal of the envisioned framework is to facilitate the end-to-end delivery of multimedia content across the Internet. To achieve this, it allows content providers and access ISPs to set up federations with the transit ISPs and cloud providers along the end-to-end path connecting them. Cooperating in a federation holds advantages for all involved stakeholders. On one hand, it allows the content provider to guarantee QoS across the Internet. On the other, it entitles the intermediary network domains to a share of the content provider's revenue. The remainder of this section identifies the different stakeholders involved in these federations and describes how they interact to set them up. Sections IV and V further elaborate on the algorithmic details of the federation set-up process.

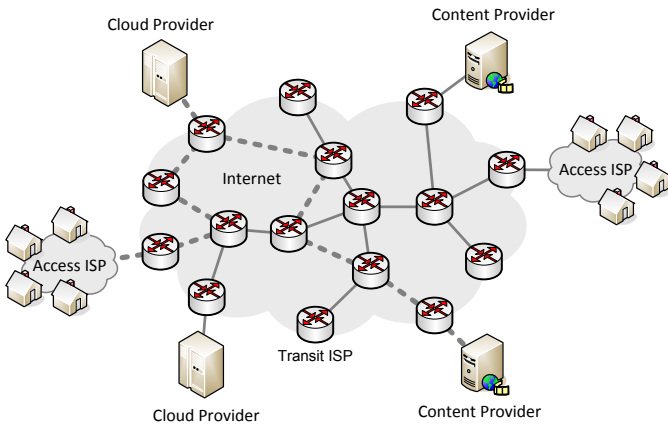


Fig. 1. An overview of the stakeholders involved in the end-to-end content delivery process: content providers, cloud providers, access ISPs and transit ISPs; the dotted line represents an example delivery path

A. Stakeholders

Figure 1 positions the different stakeholders throughout the network. There are four types of stakeholders involved in the content delivery federations: content providers, access ISPs, transit ISPs, and cloud providers. The content providers, access ISPs and cloud providers are positioned at the edge of the Internet, connected to the rest of the network through a single transit ISP. The *content provider* locally hosts a set of multimedia content items. It aims to sell these to interested access ISPs. Traditionally, the *access ISP* provided Internet access to a set of end-users. Additionally, they have started offering novel multimedia services, including time-shifted television and Video on Demand. Our framework builds on this and considers the access ISPs the direct customers of the content provider. In line with current advances, the access ISP is assumed to offer multimedia services to the end-users. However, in contrast to current approaches, it does not manage its own content catalogue, but rather obtains content via one or more content providers, over the Internet. The Internet core consists of many *transit ISPs*, together forming a network of networks. They connect the different edge domains to the Internet, and are responsible for routing network traffic from source to target. In the current Internet, end-to-end routing is static and best-effort. However, future Internet research has advanced the idea of on-demand end-to-end QoS provisioning within the Internet core [3], [4], [6]. As such, transit ISPs could be included in the end-to-end federations, providing QoS guarantees in return for a share of the content provider’s revenue. The cloud computing paradigm promises the on-demand and elastic provisioning of storage and computing resources. This allows content providers to deploy proxy caches on-the-fly across the Internet, by dynamically leasing storage resources from *cloud providers*. The use of proxy caches can significantly reduce bandwidth consumption and improve QoS, by storing popular content closer to the end-users [14]. The dotted lines in Figure 1 represent an example end-to-end delivery path, which includes a single cloud provider.

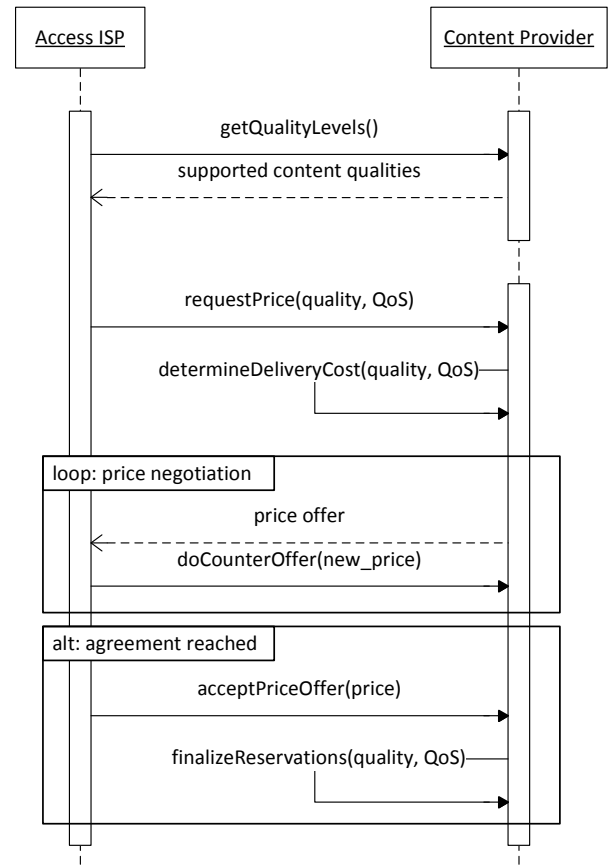


Fig. 2. A sequence diagram detailing the negotiation process between the content provider and its customer, the access ISP

The presented framework thus combines characteristics and advantages of several existing multimedia content delivery approaches. First, it allows content providers to offer their content to end-users across the Internet, as is the case in the OTT scenario. However, in contrast to OTT content delivery, access ISPs are still involved in the delivery process and can share in the revenue. Thus allowing access ISPs to offer a plethora of multimedia services, as they currently often do, without needing to maintain their own content catalogue. Additionally, by including transit ISPs and cloud providers in the delivery federations, QoS guarantees can be offered, further reducing disadvantages of OTT content delivery.

B. Interactions

The stakeholders involved in the content delivery process interact in several ways in order to set up QoS-guaranteed end-to-end paths across the Internet. The remainder of this section describes these interactions in more detail.

The federation set-up process is initiated by the access ISP when it decides it wants to offer the content of a specific content provider to its end-users. Figure 2 depicts the negotiation process between the content provider and an access ISP in detail. First, the ISP requests the list of quality levels offered by the content provider. The content provider replies with information about each quality level. Depending

on the type of multimedia content the offered information may differ. It could include for example spatial resolution, temporal resolution, and bit-rate. The access ISP can use this plethora of information to determine which quality levels its end-users will be interested in (based on their device capabilities and expectations). From the content provider’s perspective the bit-rate is the important attribute, as it determines the amount of bandwidth resources that need to be reserved in order to deliver the content and consequently the cost associated with delivering the content. Subsequently, the access ISP requests an initial price offer for delivering content with a specific quality and specific QoS requirements (e.g., delay, jitter, packet loss). The content provider uses its price models and associated algorithms to determine the costs associated with delivering its content to the access ISP. To calculate this cost, it additionally needs to find a cost-minimizing end-to-end path and identify the network domains along this path. This is a complex problem in itself that will be further discussed in Section IV. The content provider can then formulate an initial price offer to the ISP, based on its calculated delivery costs and the expected profit. This offer triggers an optional negotiation protocol, which allows the ISP to do a counter offer. More details about this protocol are supplied in Section V. This negotiation can end in either agreement or disagreement. If an agreement is reached, the SLA negotiation process is finalized and the ISP can start offering the content provider’s multimedia content to its end-users. This finalized SLA contains the negotiated price, content quality, QoS requirements and the time-frame over which it is valid. The negotiated SLA is only valid for a certain period of time to take into account changes in delivery costs. After the time-frame has expired, a new SLA should be negotiated taking into account current parameters and costs. If no agreement is reached during the negotiation procedure, the federation is not finalized. It is then up to the access ISP to start again, either by contacting another content provider or requesting less stringent QoS.

As previously stated, the content provider needs to determine the costs associated with delivering its content across the Internet towards its federated access ISPs. Additionally, it needs to reserve the necessary resources in order to satisfy its obligations captured in the negotiated SLA. This introduces a set of additional interactions between the content provider, the transit ISPs and the cloud providers along the end-to-end path to the access ISP. Figure 3 depicts the interactions involved in finding an end-to-end delivery path between the content provider and an access ISP. In order to calculate the costs involved in delivering its content to a specific access ISP, the content provider first needs to identify the network domains the content will pass through. These domains can either be transit ISPs or cloud providers. In case of the former, the content provider requests the cost per bandwidth unit for reserving a path through the domain with the requested QoS. In case of the latter, it asks the cost for reserving storage resources. This cost information is then employed by the content provider to calculate the cost-minimizing end-to-end path to the access ISP that satisfies the requested QoS. Finally,

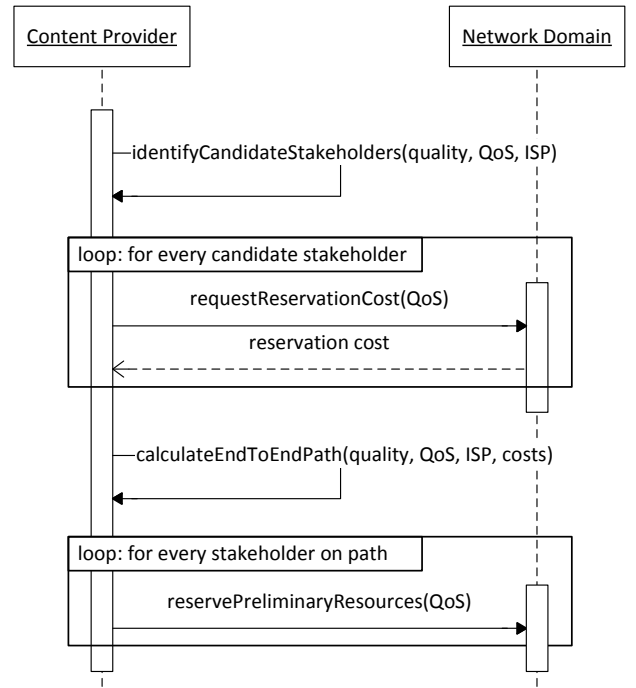


Fig. 3. A sequence diagram detailing the interactions involved in finding an end-to-end delivery path between the content provider and an access ISP

it performs a preliminary lock on the necessary resources within the domains along this path, to make sure it can satisfy the offer that will be made to the access ISP. These preliminary reservations are finalized as soon as the content provider and access ISP come to an agreement. If they do not, the locks are released.

IV. SETTING UP END-TO-END DELIVERY PATHS

Section III gave an overview of the stakeholders involved in the envisioned content delivery approach, as well as how they interact and collaborate. This section further explores the algorithmic details, which were previously omitted. The problem of finding an end-to-end delivery path between a content provider and an access ISP is mathematically modelled. Solving this model provides an upper bound to the cost of delivering the content from the content provider to its access ISP customers.

A. Notations and assumptions

The Internet consists of a set of ISPs \mathcal{I} , interconnected in a hierarchical graph structure. Two types of ISPs are considered: transit ISPs \mathcal{T} and access ISPs \mathcal{A} . Additionally, there are different types of edge network domains, of which two are involved in our envisioned framework: content providers \mathcal{C} and cloud providers \mathcal{P} . Every transit ISP $t \in \mathcal{T}$ is directly connected to a set of neighbouring transit ISPs $\mathcal{N}_t \subseteq \mathcal{T}$. The content providers, cloud providers, and access ISPs together form the set of edge domains $\mathcal{E} = \mathcal{C} \cup \mathcal{P} \cup \mathcal{A}$. Every edge domain $e \in \mathcal{E}$ is connected to the Internet through a single transit ISP $t_e \in \mathcal{T}$.

In the considered scenario, it is possible to reserve QoS-guaranteed paths through the Internet core. To this end, every transit ISP $t \in \mathcal{T}$ offers a set of QoS levels L_t . Every $l \in L_t$ maps a set of QoS parameters Q_l to a set of guarantees V_l (e.g., maximum delay, minimum availability) and has an associated transmission cost θ_l per unit of content. The guarantee of $l \in L_t$ associated with parameter $q \in Q_q$ is denoted as $v_{l,q} \in V_l$. Cloud providers support on-demand provisioning of computational and storage resources, allowing content providers to dynamically deploy proxy caches. As such, there is a cost σ_p associated with every cloud provider $p \in \mathcal{P}$, for storing a single unit of content.

Every content provider $c \in \mathcal{C}$ offers a content catalogue. It offers its content at a number of different qualities, characterised by a set of bit rates B_c . The average size of its content items of quality $b \in B_c$ is defined as $s_{c,b}$ units. A cumulative popularity distribution $D_c(x)$ is associated with every content provider c . It is a function that takes as argument an amount x and returns as value the percentage of times one of the x most popular content items of c will be requested. In order to set up a dynamic federation, the access ISP $a \in \mathcal{A}$ sends a request to a content provider $c \in \mathcal{C}$. Such a request r consists of a mapping between QoS parameters Q_r and expected guarantees V_r . The expected QoS guarantee of request r associated with parameter $q \in Q_r$ is defined as $v_{r,q}$. Additionally, it mentions the expected quality $\beta_r \in B_c$.

Different QoS parameters are aggregated in different ways. For example, the total end-to-end delay is the sum of the individual link delays, while the total throughput is the minimum of all link throughputs. As such, we define the end-to-end QoS aggregation operator of a QoS parameter q as \otimes_q . It calculates the aggregated value of the QoS parameter based on the individual link values. Additionally, the relationship to determine if a QoS value satisfies a requirement depends on the QoS parameter. For example, packet loss is satisfied if it is below the required value, while throughput is satisfied if it is above. As such, we define the satisfaction relationship of a QoS parameter q as \vdash_q .

B. Problem formulation

The goal, given a request r , is to find an end-to-end path from a content provider $c \in \mathcal{C}$ to an access ISP $a \in \mathcal{A}$ that satisfies the requirements specified in r . Additionally, it should minimize the total delivery costs, allowing the content provider to maximize its profit. To achieve this, the content provider finds a cost minimizing path through the Internet core. Optionally, proxy caches can be deployed along this path in order to reduce transmission costs. It is obvious that sharing a single proxy cache between multiple access ISPs is more cost efficient, as it reduces transmission costs for all of them, while generating the storage costs only once. As such, we consider a more generic version of the presented problem, which finds a delivery tree, with the content provider at the root and its access ISP customers at the leaves. We assume that different qualities of the same content item are totally separate files. As such, we can construct a tree for every quality level offered

by the content provider containing as leaves the access ISPs requesting that quality level. This can be done without loss of generality as cache sharing across quality levels does not introduce performance gains anyway.

The solution consists of a path $\pi_{c \rightarrow a} = [d_1, \dots, d_n]$ between the content provider $c \in \mathcal{C}$ and all of its access ISPs $a \in \mathcal{A}_c$. The path consists of 1 content provider, 1 access ISP, a set of 0 or more cloud providers and 1 or more transit ISPs. Additionally, it satisfies the following constraints:

$$d_1 = c \quad (1)$$

$$d_n = a \quad (2)$$

$$\forall e \in \pi_{c \rightarrow a} \cap (\mathcal{C} \cup \mathcal{P}), \quad e_{\succ} = t_e \quad (3)$$

$$\forall e \in \pi_{c \rightarrow a} \cap (\mathcal{P} \cup \mathcal{A}), \quad e_{\prec} = t_e \quad (4)$$

$$\forall t \in \pi_{c \rightarrow a} \cap \mathcal{T}, \quad t_{\succ} \in \mathcal{N}_t \cup \mathcal{P} \cup \mathcal{A} \quad (5)$$

With i_{\succ} and i_{\prec} defined as the successor and predecessor of i in the path. Equations 1 and 2 state that the first and last domain in the path should be the content provider and access ISP. Equation 3 states that the successor of the content provider and every cloud provider in the path should be its gateway. Similarly, this is defined for the predecessor of every cloud provider and the access ISP in Equation 4. Finally, Equation 5 states that the successor of every transit ISP in the path should be one of its neighbours, or alternatively a cloud provider or the access ISP.

Within every transit ISP along the end-to-end path between c and a , one of the available QoS levels should be chosen. The selected QoS level in domain t for request r is denoted as $l_{r,t} \in L_t$. The aggregate end-to-end value of every QoS parameter should satisfy the limit specified in r . This is translated into the following set of constraints:

$$\forall q \in Q_r : \quad \otimes_{t \in \pi_{c \rightarrow a} \cap \mathcal{T}} v_{l_{r,t},q} \vdash_q v_{r,q} \quad (6)$$

Subsequently, proxy caches need to be deployed within the cloud provider domains of the path $\pi_{c \rightarrow a}$. Such a proxy cache temporarily stores the most popular subset of the requested content in order to reduce bandwidth consumption of the incoming links. The size of the cache deployed at cloud provider $p \in \pi_{c \rightarrow a} \cap \mathcal{P}$ is defined as γ_p .

The goal is then, for a given content provider $c \in \mathcal{C}$ and content quality $b \in B_c$, to find a delivery tree towards its access ISP customers requesting quality b , select suitable QoS levels l_t in every transit ISP $t \in \mathcal{T}$, and find optimal cache sizes γ_p for every cloud provider $p \in \mathcal{P}$ with the global minimal cost, while satisfying the previously defined constraints. The cost consists of two parts: transmission and storage. The storage cost per unit of time is easily calculated as follows:

$$\sum_{p \in \pi_{c \rightarrow a} \cap \mathcal{P}} \gamma_p \times \sigma_p \times s_{c,b} \quad (7)$$

In contrast, the transmission cost is more difficult to calculate, as it is influenced by the intermediary caches. The actual transmission cost is difficult to model, as a cache alters the

popularity distribution on the incoming path. For a single cache of size γ , this is not much of a problem, as the bandwidth consumption on the incoming link is reduced by $D_c(\gamma)$. However, the problem becomes more complex in a hierarchical scenario, where the leaf caches have different sizes. To alleviate this problem, we propose a formula to calculate an upper bound on transmission cost, by choosing the minimum size of all direct child caches, instead of their actual sizes. First, we calculate the percentage of requested content that traverses every transit ISP in $\pi_{c \rightarrow a}$. As this percentage is the same for all transit ISPs forming the sub-path of $\pi_{c \rightarrow a}$ between an edge domain $e \in \mathcal{P} \cup \mathcal{A}$ and its predecessor edge domain $e_{\leftarrow} \in \mathcal{C} \cup \mathcal{P}$, it is calculated only for all $e \in \mathcal{P} \cup \mathcal{A}$. First, we calculate the aggregate cache size at domain e , which is defined as the lower bound of the number of content items that are cached at e , or in one of the downstream caches:

$$\gamma_e^{\text{aggr}} = \gamma_e + \min_{c \in e_{\downarrow}} \gamma_c^{\text{aggr}} \quad (8)$$

With e_{\downarrow} the children of e in the delivery tree. The lower bound on the bandwidth reduction in the transit ISPs between e_{\leftarrow} and e is then calculated as $D_c(\gamma_e^{\text{aggr}})$. This allows us to finally calculate an upper bound of the transmission cost for a single request per unit of time:

$$c^{\text{tran}} = \sum_{e \in \pi_{c \rightarrow a} \cap (\mathcal{A} \cup \mathcal{P})} \sum_{t \in \pi_{e_{\leftarrow} \rightarrow e} \cap \mathcal{T}} (1 - D_c(\gamma_e^{\text{aggr}})) \times \theta_{t_t} \times b \quad (9)$$

Equations 7 and 9 define the storage and transmission cost respectively. However, the storage cost is independent of the number of delivered content items, while the transmission cost does depend on it. As such, both costs cannot be trivially combined in the total delivery cost per delivered item (which is what the access ISP expects). However, it is possible to calculate the total cost per delivered item, if we define it as a function of the average number of simultaneously delivered content items d . As the cache is shared among multiple access ISPs, the storage cost should additionally be divided by the total number of access ISPs that employ it. We define the total number of access ISPs that employ the cache at cloud provider p , as a_p . This results in the following combined cost function:

$$c(d) = \left(\sum_{p \in \pi_{c \rightarrow a} \cap \mathcal{P}} \frac{\gamma_p}{a_p} \times \sigma_p \times s_{c,b} \right) + d \times c^{\text{tran}} \quad (10)$$

C. Notes on solving the problem

The problem of finding a path between two edge domains and selecting suitable QoS levels in each intermediary domain along the path has been studied in the past, for example by Pouyllau and Douville [5]. They found the problem to be equivalent to the k-multi-constrained optimal path problem [15], which is NP-Hard. To alleviate this, Pouyllau and Douville proposed a heuristic based on game theory and Q-learning [6]. Additionally, their algorithm optimizes long-term revenue, instead of finding the optimal solution on a per-request basis. The problem presented above, however, is more complex as it adds the dimension of placing proxy caches and determining

the cost-minimizing cache sizes. Due to the advantages of cache sharing, we also consider a delivery tree rather than a path. As such, the presented problem is also NP-Hard and existing algorithms for the path composition problem cannot be trivially adapted for our purposes.

As a solution, an iterative heuristic could be devised. First, it finds the direct cost-minimizing path between the content provider and its access ISP customers, without deploying intermediary caches. This could be done using existing heuristics for the k-multi-constrained optimal path problem, as described above. Iteratively, the algorithm could then identify proxy cache locations that would reduce its costs and include those in the path towards its access ISP customers. The details of this heuristic will be further developed in future work.

V. SLA NEGOTIATION

The SLA specifies the requirements of the service consumer and the assurances by the service provider [10]. In our presented framework, SLAs are used to specify the QoS guarantees offered by the content provider, the prices paid by the access ISP for the delivery of the content, and the time-frame over which the agreement is valid. SLAs are valid only for a fixed period of time. This allows the content provider to periodically alter its prices to take into account cost fluctuations for reserving resources in the Internet core. As previously stated, the content provider and access ISP must agree on a price for delivering content, before the federation can be finalized. Important aspects of this negotiation, which are often neglected, are the initial price offer and the lower price limit. If the content provider estimates this price limit too low, it will not achieve its expected profit. Similarly, if it estimates it too high, the potential customer might be turned away. The presented framework tackles this issue by providing a mathematical model, that allows the content provider to calculate the costs associated with delivering its content to a specific access ISP (cf. Section IV). This cost can be used as a basis for determining the initial price offer as well as the lower limit. In order to be able to successfully negotiate, the access ISP also needs to determine the boundaries within which it can negotiate. However, this is far less complex than calculating the end-to-end delivery cost. It can use as a benchmark, for example, the price the end-users are willing to pay to consume the content, or the price local competitors are charging.

Several protocols have already been proposed in literature for the actual negotiation of the terms specified in the SLA. As such, we consider the actual protocol details to be outside the scope of our work. Existing SLA negotiation protocols can be incorporated in the presented framework [11], [13].

VI. TOWARDS AN IMPLEMENTED PROTOTYPE

We are currently in the process of implementing a prototype, which also incorporates the algorithms presented in prior work [2]. The framework itself is implemented in the Java programming language and uses the CPLEX 12.3 library to solve the mathematical model. The WS-Agreement recommendation is used for the representation of SLAs and negotiating their

terms. SLAs take the form of an XML document, with the following generic structure:

```
<wsag:Agreement AgreementId=xs:string>
  <wsag:Name>xs:string</wsag:Name> ?
  <wsag:Context>
    wsag:AgreementContextType
  </wsag:Context>
  <wsag:Terms>
    wsag:TermCompositorType
  </wsag:Terms>
</wsag:Agreement>
```

The agreement consists of a mandatory context and a collection of agreement terms. The context identifies the involved parties, the services that are being agreed upon and the duration over which it is valid. It takes the following form:

```
<wsag:Context xs:anyAttribute>
  <wsag:AgreementInitiator>
    xs:anyType
  </wsag:AgreementInitiator> ?
  <wsag:AgreementResponder>
    xs:anyType
  </wsag:AgreementResponder> ?
  <wsag:ServiceProvider>
    wsag:AgreementRoleType
  </wsag:ServiceProvider>
  <wsag:ExpirationTime>
    xs:DateTime
  </wsag:ExpirationTime> ?
  <wsag:TemplateId>
    xs:string
  </wsag:TemplateId> ?
  <wsag:TemplateName>
    xs:string
  </wsag:TemplateName> ?
  <xs:any/> *
</wsag:Context>
```

The terms make up the main body of the agreement. They express consensus between or obligations of a party. There are two types of terms: service terms and guarantee terms. The former provide information needed to instantiate or identify a service to which an agreement pertains and to which the guarantees apply. The latter specify the service levels that the parties are agreeing to.

VII. CONCLUSION

This paper presents a novel framework for setting up end-to-end federations between the stakeholders involved in the delivery of multimedia content. More specifically, it guides the negotiation of SLAs between content providers, ISPs and cloud providers. This allows them to overcome the disadvantages associated with current delivery approaches, such as OTT content provisioning and content offered by access ISPs. Our framework has several novel aspects compared to existing works. First, by including cloud providers in the end-to-end delivery paths, proxy caches can be dynamically deployed throughout the network. This introduces an additional complexity to the problem, but allows further optimization of the delivery process. Second, we propose a cost model that can be

used as a basis for the SLA negotiations. It is important for the stakeholders involved in the negotiation process to know the limits within which they can vary their terms. This aspect is often ignored in existing works. Our model allows the content provider to calculate the costs for delivering its content. This can be used to determine his initial price offer, as well as his lower price limit.

ACKNOWLEDGMENT

Steven Latré and Tim Wauters are funded by the Fund for Scientific Research Flanders (FWO). The research leading to these results was partially performed within the context of the FP7 OCEAN project and received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement number 248775.

REFERENCES

- [1] M. Serrano, S. van der Meer, V. Holum, J. Murphy, and J. Strassner, "Federation, a matter of autonomic management in the Future Internet," in *12th IEEE/IFIP Network Operations and Management Symposium (NOMS)*, 2010, pp. 845–849.
- [2] J. Famaey, S. Latré, T. Wauters, and F. De Turck, "FedRR – a federated resource reservation algorithm for multimedia services," in *13th IEEE/IFIP Network Operations and Management Symposium (NOMS)*, 2012.
- [3] N. Kumar and G. Saraph, "End-to-end QoS in interdomain routing," in *2nd International Conference on Networking and Services*, 2006.
- [4] H. Xiangjiang, Z. Peidong, C. Kaiyu, and G. Zhenghu, "AS alliance in inter-domain routing," in *22nd International Conference on Information Networking and Applications – Workshops (AINAW)*, 2008, pp. 151–156.
- [5] H. Pouyllau and R. Douville, "End-to-end QoS negotiation in network federations," in *12th IEEE/IFIP Network Operations and Management Symposium – Workshops (NOMS)*, 2010, pp. 173–176.
- [6] H. Pouyllau and G. Carofiglio, "Inter-carrier SLA negotiation using Q-learning," *Telecommunication Systems*, 2011.
- [7] S. Balasubramaniam, D. Botvich, R. Carroll, J. Mineraud, T. Nakano, T. Suda, and W. Donnelly, "Biologically inspired future service environment," *Computer Networks*, vol. 55, no. 15, pp. 3423–3440, 2011.
- [8] C. Yuanming, W. Wendong, G. Xiangyang, and Q. Xirong, "Initiator-domain-based sla negotiation for inter-domain qos-service provisioning," in *4th International Conference on Networking and Services*, 2008.
- [9] P. Rubach and M. Sobolewski, "Dynamic SLA negotiation in autonomic federated environments," in *On the Move to Meaningful Internet Systems*, 2009.
- [10] A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, T. Nakata, J. Pruyne, J. Rofrano, S. Tuecke, and M. Xu, "Web services agreement specification (WS-Agreement)," 2011, <http://www.ogf.org/documents/GFD.193.pdf>.
- [11] D. Battré, F. Brazier, K. Clark, M. Oey, A. Papispyrou, P. Wieder, and W. Ziegler, "WS-Agreement negotiation version 1.0," 2011, <http://www.ogf.org/documents/GFD.192.pdf>.
- [12] P. Hasselmeyer, H. Mersch, B. Koller, H.-N. Quyen, L. Schubert, and P. Wieder, "Implementing an SLA negotiation framework," in *Expanding the Knowledge Economy: Issues, Applications, Case Studies (eChallenges)*, 2007, pp. 154–161.
- [13] M. Parkin, P. Hasselmeyer, B. Koller, and P. Wieder, "An SLA renegotiation protocol," in *2nd Non Functional Properties and Service Level Agreements in Service Oriented Computing Workshop*, 2008.
- [14] D. De Vleeschauwer and K. Laevens, "Performance of caching algorithms for IPTV on-demand services," *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 491–501, 2009.
- [15] J. Xiao and R. Boutaba, "QoS-aware service composition and adaptation in autonomic communication," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 12, pp. 2344–2360, 2005.