Optimal Functionality Placement for Multiplay Service Provider Architectures

Ioannis Papapanagiotou, *Student Member, IEEE*, Matthias Falkner, *Member, IEEE*, and Michael Devetsikiotis, *Fellow, IEEE*

***Abstract*—The proliferation of multiplay services is creating design dilemmas for service providers, related to where certain key networking functionality should be placed. For example, service providers need to know whether to distribute more network intelligence closer to the subscriber or cluster it in a central location. In view of this, we quantify the cost differences among service provider architectures, identified based on the functionality distribution (centralized vs. distributed, clustered vs. unclustered and single vs. multi edge). For this purpose, we formulate a modular mixed-integer programming model based on a set of close-to-real-case scenarios. Given the complexity of such problems, we propose methodologies that can reduce the number of locations. Our results indicate that distributing the IP intelligence and the video replication is preferable. Moreover, deploying edge systems with faster backplane has little benefit in the aggregation network, and providers should rather invest in faster interfaces.**

***Index Terms*—Aggregation networks, metro Ethernet, Ethernet based DSL, edge systems, triple play.**

I. Introduction

THE well established 80/20 rule for client-server versus local traffic has driven network designers to address problems with pure L2 switches acting as hubs, and L3 routers performing the routing towards different paths. However, the rise of on demand video streaming (e.g., Netflix, Hulu), Peer-to-Peer (P2P) television over the internet (e.g., TVUPlayer, PPLive, QQLive, PPStream), and the advent of mobile traffic has triggered a number of actions from broadband providers who are trying to achieve flexibility, scalability and efficiency [21]. According to several studies, the annual global traffic will be doubling in volume every two years, while a quarter of this is projected to be video traffic [2], [3]. Current 4G mobile service provider architectures are bandwidth constrained, and the cost to build out new networks to increase the available bandwidth is prohibitively expensive.

Therefore, it is of vital importance for the Internet Service Providers (ISPs) to engineer their infrastructure to meet these challenges. In this paper, we focus on answering the question of *where to place certain functionalities* in the aggregation network. We model this through a Network Design Problem (NDP) that the SP may use to achieve the maximum efficiency with the minimum deployment capital cost.

In the past, NDPs have been used to address the problem of cost optimization when developing *new* infrastructures. Nonetheless, most ISPs already possess an access infrastructure. In order to address this issue, we propose an optimization formulation that uses the current infrastructure (in terms of design), but defines the optimal distribution of *functionalities* instead. Another important aspect is that devices used in the Ethernet aggregation, and edge part of the network, are not anymore purely layer 2 switches (in the NDP literature they are sometimes called "aggregators" [26]) and layer 3 routers. Most of the vendors have engineered multipurpose edge "systems" that incorporate different functionalities and network intelligence in modular "sub-systems" e.g., high-end backplanes in which multiple interface and functionality line cards are added. Therefore, design problems for such architectures need to be revisited.

Considering the above challenges, in this paper we classify the aggregation architectures based on the location of the functionalities and network intelligence, and then proceed to build a modular optimization model that is able to achieve the minimum deployment cost. Our contribution is summarized as follows:

- We define the possible aggregation architectures and group them into three main categories: 1. *Centralized Single-Edge*; 2. *Centralized Multi-Edge*; and 3. *Distributed Single-Edge*. On top of these categories, another subcategory is added, whether to cluster the edge systems or keep the functionalities in a single box.
- We develop a network design model that goes beyond the well investigated *location* and *dimensioning* problems. Our modeling approach is applied to both design an aggregation architecture, as well as upgrading the current infrastructure.
- We propose models based on edge "systems", rather than network elements. Edge systems may support different types of functionalities either on their own, or as attached "sub-systems" (line cards).
- We formulate constraints that account for physical characteristics (line card capacity, port capacity), bandwidth (device and port bandwidth), and layer 2 versus layer 3 functionalities (such as switching and routing, VLAN/IP termination capacity, or business functions).

Manuscript received March 5, 2011; revised November 28, 2011, and March 15 and May 29, 2012. The associate editor coordinating the review of this paper and approving it for publication was A. Vasilakos.

I. Papapanagiotou and M. Devetsikiotis are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh NC, 27606 USA (e-mail: {ipapapa, mdevets}@ncsu.edu).

M. Falkner is with the Edge Router Business Unit of Cisco Systems, Ottawa, Canada (e-mail: mfalkner@cisco.com).

This work was supported in part by the Institute for Next Generation IT Systems, Research Grant 09-06, and by the Cisco University Research Program, Gift #2008-04555 (3696).

Digital Object Identifier 10.1109/TNSM.2012.061212.110032

- We propose two novel heuristics for scaling down the problem and for decreasing the execution time of the problem.
- We evaluate our model with two close-to-real-case scenarios and with multiple traffic profiles. We show that, with the current trends, the SPs will need to re-engineer their aggregation infrastructure.

The paper is structured as follows: In the next section, we provide an overview of the current state of the art in NDPs and aggregation design methodologies. In section III, we define and explain the architectures according to the distribution of the edge systems. In section IV, we describe the proposed modeling methodology and assumptions. Given the plurality of symbols and notations, all of them are summarized in the tables of section IV. In section V, we explain the modeling approach for the traffic flows and in Section VI we deploy the cost optimization model. In section VII, we explain the heuristics that have been used to scale-down the problem, and in section VIII we showcase an evaluation of two multi-service scenarios over metro area networks based on architectural values from EU ISPs and actual system values from vendors. Section IX, includes the discussion and further remarks, and we conclude with the final section.

II. Background Work

During the Internet era the extensive need for bandwidth became a crucial issue for Internet SPs. This led the research community to focus on Network Design Problems that require advanced optimization techniques to be solved. These problems can be divided into two main categories: *Locationing problems*, which are pure topological design problems and where the demand volumes are not taken into consideration; *Dimensioning problems*, which incorporate the demand volumes. Both classes of problems have been extensively analyzed in [20], [26], [27], [28].

A sub-category of the dimensioning problems, are the two-level hierarchical network problems [10], [15]. In [7], the authors investigate an hierarchical network problem for fast moving users. These problems assume a set of candidate locations, capabilities of concentrators and routers, and determine the optimal location placement. However, more and more vendors are developing network systems that support on the same backplane (or chassis) multiple functionalities (line cards or blades). Moreover, the high demand for video traffic affects the way IPTV functionalities are distributed in Next Generation Networks [17].

Hence, on the basis of a *dimensioning problem*, our proposed approach is different from previous work in the following points: (a) We propose a modeling approach were micro-optimization is required in each location to determine the appropriate intelligence. (b) The proposed modeling approach may be applied to network designs as well as network upgrades. For example, a network architect can use the model to determine the cost of distributing the functionality closer to the subscriber. (c) The proposed design is not limited to a single hierarchical design (e.g. an IP termination functionality can be placed before or after the switch). (d) Each service may have different characteristics. For example video traffic for IPTV is usually multicast [13]; P2P applications tend to generate multiple flows, some of which are geographically concentrated or distributed over long distances [5]; Internet traffic is usually based on a client-server behavior [22].

In [29], the authors have solved a Network Utility Maximization (NUM) problem for triple play services. The authors propose three utility functions for each offered service. However, NUM problems address the issue of fairness and require the a priori knowledge of the utility function, which in many cases is hard to compute. In our work, we propose three different flows, which have directional characteristics and are quantifiable by performing Deep Packet Inspection (DPI) or flow inspection and classification [6].

In [32], the authors considered an ADSL access network consisting of subscribers, Digital Subscriber Line Access Multiplexers (DSLAMs), metro Ethernet switches and a Broadband Remote Access Aggregation Server (BRAS), and have developed analytical expressions for dimensioning the access network in the upstream direction. Moreover, in [11], the authors performed a cost investigation of transport architectures based only on demand and physical layer capabilities. In our work, we use an Ethernet based next generation aggregation network in which the BRAS has been replaced by multi-purpose Broadband Network Gateways (BNGs), Multi-Service Edge Routers (MSEs) and Video BNGs [4], [12]. We also propose multiple architectures based on the distribution and clustering of the edge functionalities. Our approach is based on the TR-101 broadband forum's report [1], which standardizes the Ethernet based aggregation design.

In [24] we presented a comparison of centralized vs distributed unclustered topologies, through two optimization problems. In this paper, we combine the problems into a single cost optimization model and extend the above work to include most of the possible ISP architectures. Finally, we evaluate our methodology with architecture designs provided by three major European service providers.

III. Architectural Comparison

Since each edge system (BNG, MSE, Video BNG) may support differential and modular functionalities, various topologies may be implemented even when the edge router's backplane remains the same. For example a BNG (into a single chassis) that incorporates both IP termination and video functionalities. Another approach that is followed by some vendors is to add several L3 functionalities on L2 devices (e.g., IP termination) [4]. We propose the following architectural combinations for an ISP aggregation network: Centralized or Distributed based on the intelligence; Single or Multi edge based on the services per location; Clustered and Unclustered based on the functionalities.

A. Network Elements in an Aggregation Network

L2 Aggregation devices: are high capacity L2 aggregation switches used as a second level of aggregation, and perform low cost VLAN policy enforcement. They usually support Gigabit Ethernet ports and multiple filtering polices per VLAN. They are the simplest aggregation devices.

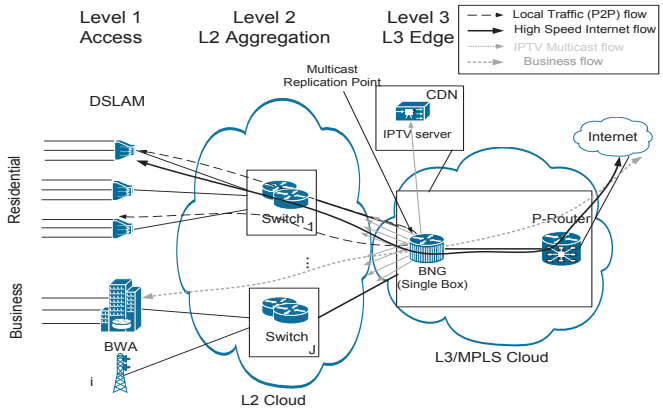

Fig. 1. Centralized single edge overlay architecture.

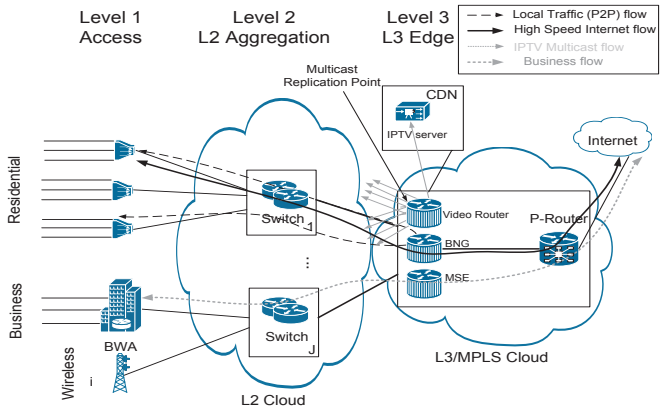

Fig. 2. Centralized multi edge overlay architecture.

BNG: Broadband Network Gateways are IP devices terminating the layer 2 access and route over IP/MPLS with support of a full set of MPLS and IP routing protocols, including multicast routing (PIM-SM/IGMP [8]). They enforce sophisticated IP QoS per service and per-content/source differentiation. They usually terminate PPP sessions or IP tunnels and can support up to hundred of thousands of subscribers and Gbps capacities. Edge routers, according to [1], can support additional functionalities related to Authentication, Authorization and Accounting (AAA) and are dynamic devices that mainly focus on residential customers. Moreover several next generation BNGs have the capability to support mobile blades to provide mobile termination functionalities.

MSE: Multi-Service Edge Routers are responsible for routing business traffic, which is usually dedicated bandwidth. There is no need to authenticate the business users as these are leased lines, hence MSEs support less intelligence.

Video BNG: Video Broadband Network Gateways are dedicated routers that have been introduced to handle the increasing demand for video traffic (e.g. IPTV, Netflix etc.). A video BNG does not implement subscriber management functions (e.g., PPP termination, per user QoS), since these functions are likely to be performed by the other network elements. In fact the video BNG may be the point of insertion of an IPTV flow and/or the replication point (L3 Multicast).

B. Centralized Single-Edge Architecture

This type of design was very important in the early evolution of aggregation networks. In this type of architecture, the L2 Metro Ethernet aggregates the traffic from multiple access points and delivers the Virtual Local Area Network (VLAN) to the IP Edge network, as shown on Fig. 1. Some of the characteristics of this architecture are: 1) all types of traffic are backhauled to the BNGs and then to a single P-Router location, which is connected to the ISP backbone (P-Router is part of the backbone and sometimes called PoP); 2) Subscriber termination functionality, multicast replication and IP QoS policies are executed in the BNG deeper in the network; 3) Multicast traffic for broadcast video is transmitted from the edge router over L2 multicast VLANs to all customer premises (Wireless BS, DSLAM).

C. Centralized Multi-Edge Architecture

In the multi-edge architecture there are different types of edge routers (BNG, Video BNG and MSE) that handle separate classes of subscriber and different types of traffic, as shown in Fig. 2. Residential subscribers are usually terminated in the BNG, whereas the Business VLANs or leased lines are routed through the MSE. Some providers, in order to enforce specific QoS policies on their video channels (IPTV), implement a separate 'Video BNG' [1]. Therefore, the Centralized Multi-Edge architecture benefits from incumbency, since it is easier to evolve from the existing architecture.

D. Distributed Single-Edge Design

A distributed IP Edge approach is being considered by many SPs as an alternative architecture to satisfy the bandwidth requirements for future applications. As shown in Fig. 3, the edge network is comprised by both L2/L3 routers. Video and HSI are backhauled over separate VLANs to the Edge Routers, where services and access to the IP network is controlled. The scalability is increased, since the amount of state information in the BNG is decreased (less subscribers are terminated per BNG) and IP QoS policies are enforced closer to the last mile. IP multicast routing is used across the L2/L3 Carrier Ethernet network for delivery of broadcast video services. In the single edge case, all services flow through a single device distributed closer to the subscriber.

Furthermore, based on the allocation of subscribers the aforementioned architectures can be further divided into:

- *Clustered*: Allocating the subscribers to a particular service over many systems located in the same PoP.
- *Unclustered*: Allocating all subscribers for a particular service to one system.

With the current centralized design, the increasing demand of video channels over IP networks leads to unavoidable bandwidth problem in the aggregation networks. Distributing the replication functionalities in multiple aggregation levels may result in less congestion of traffic bandwidth and VLANs, since a single unicast flow is required from the CDN to the distributed Video BNG. Moreover, the evolution of P2P IPTV sets new standards on how users interact. Distributing the IP intelligence can deal more effectively with traffic flows that

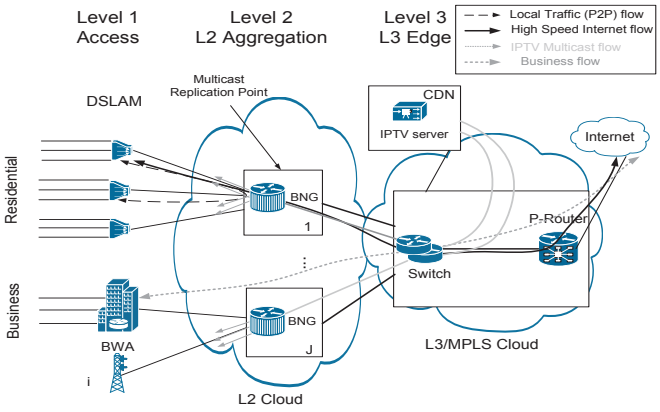

Fig. 3. Distributed single edge overlay architecture.

tend to be "local", since traffic is terminated and routed closer to the subscribers. In addition, several wireless providers are already facing issues on how to backhaul the increasing demand of mobile traffic. For this reason our work focuses on determining which architecture is the optimal one.

Finally, the distributed multi-edge architecture has been abandoned due to the increasing complexity of distributing multiple type of boxes over the aggregation network. Hence, we exclude it from the potential solutions.

IV. Methodology and Assumptions

A. Overview

The main objective of our work is to create a single modular and portable model that is able to provide the cost optimal solution among different architectural candidates. We model this through a mixed integer programming model. We divide the variables in qualitative (0-1) and quantitative. The qualitative variables are used to select the appropriate architectural approach (Centralized vs Distributed), whereas the quantitative to determine the number of systems per location. The model is a non linear mixed integer; the non-linearity appears in the constraints. Such problems can be a challenging and daring venture, because they combine all difficulties of their subclasses: the combinatorial nature of integer programs (MIP) and the difficulty in solving non-convex nonlinear programs (NLP). Hence, we have transformed the problem to a linear one in the expense of extra constraints. In addition, we decreased the search space by defining upper bounds for the number of elements and interfaces. By using these heuristics the optimization problem was solvable in a finite amount of time.

In this model, we assume a tree topology (bipartite) with multiple aggregation levels. In fact, most of the ISPs usually implement small and medium aggregation sites and several core sites. The number of locations varies based on the size of the ISP. Small locations may vary from 10-10,000, the medium locations from 10-1,000 and the core location from 1-100. Every core site may support different kinds of areas (rural or non-rural) with various customer distributions. Thus, we are modeling a single three-level aggregation network. The results can then be extrapolated to include the whole ISP aggregation infrastructure (multiple aggregation networks). In other words, solving the problem for the whole ISP network would require running the model multiple times for each aggregation network with different input parameters (e.g., subscribers, levels of aggregation).

TABLE I
Modeling Input Parameters

	Description	Symbol
General	Access Location	$a = [1, ..., A]$
	1st Level Aggregation Locations	$i = [1, ..., I]$
	2nd Level Aggregation Locations	$j = [1, ..., J]$
	Core Locations	$k = [1, ..., K]$
	Number of subscribers	SUB
	Subscribers per Residential Location a	res_a
	Subscribers per Business Location a	bus_a
	DSLAM subscriber capacity a	S_a
	Number of IPTV Channels from CDN	Ch
	% of concurrent subscribers watching IPTV	w_i
	Capacity of a Port	$c = \{1G, 10G\}$
Agg. Switch	Cost ($)	co^{L2}
	Bandwidth (Gbps)	C^{L2}
	VLAN capacity	$vlan$
	Number of slots for Ports	$P^{L2,c}$
	Cost ($) of switching ports	$co^{L2,c}$
Edge Systems	Cost ($) for Type t	$co^{L3}t$
	Bandwidth (Gbps) for Type t	C_t
	IP Termination Capacity for Type t	sub_t^{L3}
	Number of slots for Ports for Type t	$P_t^{L3,c}$
	Cost ($) of routing ports	$co^{L3,c}$

B. Definitions

On the first level there may exist A access locations. In each access location $a = [1, ..., A]$, the L2 access devices are able to handle a finite number of subscribers. On the remaining levels the model is going to define how to distribute the higher layer functionalities and how to handle different kinds of service flows.

1) Architecture: The first aggregation level is comprised of I locations ($i = [1, ..., I]$), the second level of J locations ($j = [1, ..., J]$) and the last level of $K = 1$ single location (single core site since we are modeling one aggregation area). The core site incorporates the P-Router which is responsible to route the non local traffic to the backbone or traffic from Tier-2/3 SPs (LAC/LNS functionality). However, it was not included as a variable in the model as it does not play any role on the functionality distribution. The insertion point of the video traffic from the CDN is assumed to be the core location. From the results, we noticed that if the insertion point was the Video BNG, the solution would have a very small variation.

2) Links: In order to model the links between each aggregation level, we have used boolean constants $u'_{n-1,n}$ for every aggregation level $n \in \{i, j, k\}$. If $u'_{n-1,n} = 1$, then a lower level location $n-1$ is connected with a higher level location n. However, it may be the case that the problem identifies that less aggregation levels are needed. Thus, in a similar manner $u'_{n-1,n+1}$ represents the connection with links between two different layers.

3) Parameters: Table I summarizes the parameters of the model. Our model is based on a network that has at most three aggregation levels. Our approach is able to determine, whether all three are needed or the ISP may decide to implement fewer

TABLE II
TYPES OF EDGE FUNCTIONALITIES t

Type	Functionality	Edge Design	Combinations
A	All	Single Edge	None
B	HSI and Video	Single/Multi Edge	only type E
C	HSI and Business	Single/Multi Edge	only type F
D	HSI	Single/Multi Edge	type E,F
E	Business (MSE)	Multi Edge	type D,F or type B
F	Video	Multi Edge	type D,E or type C

levels. In the same table, we also present the properties of each of the devices. Note that edge systems have the subscript t, which corresponds to different types. In regards to the port number, we make the following simplification: Edge system backplanes contain a finite amount of slots. In each slot, line cards are added that have a finite number of ports. Hence, instead of having so many variables, we have decided to associate the number of ports per device as the one to vary P_t^{L3}. Each port is also associated with a cost $co^{L3,c}$.

4) Edge System Functionality: We also assume six different types of edge routers $t \in T = \{A, B, C, D, E, F\}$, based on the functionality, the characteristics and the services that they support. We define the following edge system functionalities:

- High Speed Internet (HSI): $T_1 = \{A, B, C, D\} \subset T$ because they require IP termination;
- Business: $T_2 = \{A, C, E\} \subset T$ for business subscribers; since they lease VLANs they do not require any termination.
- Video replication (Multicast): $T_3 = \{A, B, F\} \subset T$.

Table II summarizes the types, the functionalities, the edge design that they form, and the potential combinations with other edge systems at the same location. More specifically, in order to avoid redundancy of functionalities and therefore unnecessary cost, not all types of edge routers should be used in (a) a single location and (b) in any path from access node to P-router. For instance, if a BNG (type A) is used in a location i, then it could perform everything in a single box, thus no other BNG should be installed in the same path (or location) towards the P-Router.

C. Variables

The objective function of the optimization problems is, given the appropriate locations, corresponding links and traffic demands, to optimally allocate the network elements and interfaces, and identify where to place the routing functionalities. For this, there are two different types of variables, qualitative and quantitative[1]:

1) Qualitative Variables (0-1):

- $u_{n,t}^{L3}$ is a boolean variable that specifies which type of system should be used at location n. If the edge router of type t is chosen at location n, then it is equal to 1; otherwise it is 0.
- u_n^{L2} is a boolean variable that specifies if a switch is going to be placed at location n or not. If it is, then it takes the value 1; otherwise it is 0.
- u_n^0 is a boolean that takes the value 1, if no device is installed at all. That is $u_{n,t}^{L3} = 0, \forall t \in T$ and $u_n^{L2} = 0$. Evidently $u_{n,t}^{L3} + u_n^{L2} + u_n^0 = 1$.
- u_n^{1G} is a boolean variable that takes the value 1, if 1Ge interfaces (line cards) are chosen to be installed at location n; otherwise it is 0.
- u_n^{10G} is a boolean variable that takes the value 1, if 10Ge interfaces (line cards) are chosen to be installed at location n; otherwise it is 0.

Thus if $u_{n,t}^{L3} = 0$, then either a switch or another type of edge system or nothing is installed. If $u_{k,t}^{L3} = 1$ for any $\forall t \in T$, then the optimal architecture is the centralized one (according to the definition of the previous section). Whereas, if $u_{n,t}^{L3} = 1$, for $n \neq k$, the optimal architecture is the distributed. For the centralized case, if $t = A$, then the programming model will have selected the single-edge architecture.

2) Quantitative Variables: As quantitative variables we define those variables that are natural numbers, $\mathbb{N}_0$. In this set there can be: "real" variables, those associated with a specific characteristic and affect the objective function; "dummy" variables, those that are used for other purposes and are not part of the objective function. The latter ones are used in the locations that the provider does not need to open or use. More specifically:

- $Y_{n,n+1}^{x,c,up}$ and $Y_{n-1,n}^{x,c,dn}$ are the integer variables specifying the number of interfaces of capacity c attached to the network element of layer x and located in location n. Those interfaces are attached to links that connect, either the uplink with one level higher $(n+1)$, or the downlink with one level lower $(n-1)$.
- $Y_{n,t}^{x}$ is the integer variable specifying the number of network elements x at location n and of type t. If $Y_{n,t}^{L3} > 1$ then it means that the location n needs its routing functionalities to be clustered.
- Y_n^0 is the integer dummy variable that is used whenever neither routing nor switching functionality is installed in location n. If all the n locations are in the same level have $Y_n^0 > 0$, then this means that it is cost optimal to have less aggregation levels.
- $Y_{n,n+1,dn}^0$ is the integer dummy variable that is used to depict a fictional number of interfaces, whenever no device is determined to have been installed at the $n+1$ location, $Y_{n+1}^0 = 0$.

V. MODELING THE TRAFFIC FLOWS

In our modeling approach, we assume that the beginning of the traffic is the first aggregation point, or else any access location a. This was done for several reasons: a) the number of subscribers may range to millions, making the problem hard to be solved with that amount of variables and constants; b) network architects usually use mean (or some percentage) traffic demands over subscribers of a bigger geographical area [3], [18]; c) according to the VLAN policy implemented, traffic subscriber VLANs are bundled into a bigger VLAN per service from the access location (e.g. DLSAM in wired and nodeB in 4G wireless) until the edge location (BNG in ethernet aggregation or RNC in 4G wireless); d) the service providers do not tend to incorporate L3 type capabilities and

[1]The variables are without an accent. Lower-case u is used for qualitative variables that take values 0 or 1, and upper-case Y is used for quantitative variables that are natural numbers

TABLE III
VOLUMES OF TRAFFIC

Traffic Volume per subscriber	
HSI traffic (Mbps) for residential customers	HSI_a^{res}
HSI traffic (Mbps) for business customers	HSI_a^{bus}
Bidirectional P2P (Local traffic) in Mbps	$P2P_a$
IPTV Bandwidth (Mbps) of each channel ch	$iptv_{ch}$
Internet (Non local Traffic) flowing	
through the location n	x_n^{HSI}
IPTV Traffic flowing	
through the location n	x_n^{IPTV}
from the CDN to the first IP replication point	x_n^{CDN}
P2P (Local Traffic) flowing	
through the location n	x_n^{P2P}
from one level below	$x_{n,dn,in}^{P2P}$
from one level above	$x_{n,up,in}^{P2P}$
to one level below	$x_{n,dn,out}^{P2P}$
to one level above	$x_{n,up,out}^{P2P}$

management functionalities that close to the subscribers. For example HSI_a^{res} (on table III is calculated by taking into account the harmonic mean of the bandwidth consumed per subscriber that is served by the access location a. However, an ISP may properly configure the per subscriber usage in order to assume a worst case (or some percentile) and possibly overprovision the network [19]. In the evaluation, we are investigating a variety of network throughput and showcase the effect of the traffic demand into the optimal solution.

The traffic classes and service types are generally available to the ISPs, since BNGs incorporate DPI or flow classification functionalities [6]. For modeling reasons, we have grouped the traffic flows into three main classes, i.e. IPTV, Internet (Non Local) and P2P (Local). The traffic demands are shown on Table III.

We categorize the traffic flows according to the effect (utilization, direction etc.) that they have on each aggregation level $n \in \{i, j, k\}$. Hence we define three main categories of traffic class:

A. Internet (non local) Traffic Class

Internet traffic class is any type of residential or business traffic that will flow through all the levels of the aggregation network. It is usually a client-server type of traffic and its major portion is usually web traffic [18]. In terms of network optimization, this type of traffic satisfies the conservation theorem, unless redundancy elimination devices are included in any level of the network [23]. The Internet traffic per aggregation level is represented by the following equations.

$$\begin{aligned} x_i^{HSI} &= \sum_a u'_{a,i}(x_a^{res} + x_a^{bus}) \\ x_j^{HSI} &= \sum_i u'_{i,j} x_i^{HSI} \\ x_k^{HSI} &= \sum_a (x_a^{res} + x_a^{bus}) \end{aligned} \tag{1}$$

The non local traffic that flows through an access location a is based on the number of subscribers and the corresponding network usage per subscriber. Thus $x_a^{res} = res_a \cdot HSI_a^{res}$ and $x_a^{bus} = bus_a \cdot HSI_a^{bus}$, indicate the bundled traffic that arrives in an access location, for residential and business subscribers respectively. The reason that we use separate volumes for each type of subscriber is because the network usage is different and their connections towards the "edge systems" are different.

B. P2P (Local) Traffic Class

P2P traffic class includes those applications that exhibit locality in their behavior. Those traffic flows are distributed across the aggregation network over several directions and may either remain local in the same geographical area (e.g., P2P or a business unit that has a local server), or in the same routing domain, and depend on how peers are interconnected. In P2P systems the files are split in smaller chunks and are downloaded from various sources. The system is therefore initiating various flows to various users. The P2P flow will have different performance based on the placement of the routing functionalities. Let us assume that users in the same access location are downloading P2P traffic with a certain throughput $P2P_a$ Mbps from sources that are either in the same edge network or outside the aggregation area. Therefore, the total P2P traffic that flows through an access location is $x_a^{P2P} = sub_a P2P_a$.

The distribution of functionalities is also affected by the overlay P2P network. If the user is downloading a file from peers in the same first aggregation level then we denote this probability with p_1. Similarly if the two peers coexist at the same second aggregation level location, their probability is p_2, and p_3 for the third aggregation level location. Since the P2P files are usually divided in chunks, the above probabilities can be regarded as equal to the portion of traffic that is coming from users at the same aggregation level. Practically those probabilities are related to the direction of the traffic, and can be calculated by the ISP through performing flow identification (such functionality can be found in both L2/L3 devices). Equivalently for the flow that goes outside of the network $1-\sum_{l=1}^{3} p_l$. The question now is how this probability may affect the cost of the architecture.

The following equations show how the P2P traffic flows over the networks. Effectively the first equations bundles the P2P traffic into a VLAN that will be offered to the first aggregation layer i. The existence of a router of type t at any level k, as defined by the variable $u_{k,t}^{L3}$ will affect the direction of the flow. For example, if a router that performs IP termination ($t \in T_1$) is placed at the first aggregation layer ($u_{i,t}^{L3} = 1$), all traffic that would remain local at the first aggregation layer would not have to flow through the rest of the layers. Similarly for the rest of layers

$$\begin{aligned} x_i^{P2P} &= \sum_a u'_{a,i} x_a^{P2P} \\ x_j^{P2P} &= \sum_{t \in T_1} [u_{k,t}^{L3} x_k^{P2P} + \sum_i u'_{i,j} x_i^{P2P} (u_{j,t}^{L3} + u_{i,t}^{L3} \sum_{k=2}^{3} p_k)] \\ x_k^{P2P} &= \{(1 - \sum_{t \in T_1} u_{k,t}^{L3})(1 - \sum_{k=1}^{2} p_k) + \sum_{t \in T_1} u_{k,t}^{L3}\} \sum_j x_j^{P2P} \end{aligned} \tag{2}$$

C. Multicast Traffic Class

Multicast is required in order to make efficient use of network resources when delivering broadcast content. Mainly, the desired goal is to support multicast optimization by controlling

the flooding of Ethernet multicast frames. The IGMP/PIM-SM agents can locally adjust replication filters on the device, such that packets are replicated only on those ports that have specifically requested to be part of the multicast group [8]. However, the location of the multicast functionality (either as IGMP snooping on L2 devices, or as IP multicast on L3 devices) is based on the VLAN architecture. In the following, we assume that the VLAN policy implemented by the ISP requires the replication to happen in the edge system.

A unicast IPTV traffic flow is assumed to be initiated from the CDN(s) and is replicated at the Edge Router, where multicast is implemented. The streaming bit rate is usually between $iptv_{ch}^{SD}$ =1Mbps for Standard Definition (SD) and $iptv_{ch}^{HD}$ =10Mbps for HD channels and the user is assumed to choose either the channel in one definition or in the other [30]. Several studies have shown that the selection of IPTV channels $ch \in \{1, ..., Ch\}$, follows a Zipf Law distribution $p_{ch} = \alpha/ch$, given that the channels are arranged by channel popularity [31] (α is a constant). It is also assumed that the percentage of residential users per a access locations (e.g., DSLAM, Wireless Base station) that have selected the IPTV service is w_a.

In order to calculate the required bandwidth for the unicast transmission of IPTV channels, from the CDN to the first replication point, the number of requested channels, from the residential subscriber population, needs to be determined. The probability that at least one user is watching a channel c, and belongs at the subtree formed by the location n, can be determined as $1 - P\{\text{none watching channel ch}\}(t) = 1-(1-p_{ch})^{\sum_a w_a \cdot res_a \cdot u'_{a,n}}$. Therefore, regarding the unicast flow from the CDN to the first IP replication point, the bandwidth can be defined by the following equation

$$x_n^{CDN} = \frac{iptv_{ch}^{HD} + iptv_{ch}^{SD}}{2} \cdot \sum_{c=1}^{Ch} 1-(1-p_c)^{\sum_a w_a \cdot res_a \cdot u'_{a,n}} \tag{3}$$

After the flow reaches the Video router all flows are distributed to the subscribers through multicast. However, the distribution of video routing functionality plays a role on the distribution of IPTV flows over the network. Note that if a multicast routing functionality has been determined to exist in that location, then the flow conservation theorem does not hold (since the video router will replicate the IPTV flow to the corresponding IPTV VLANs). For example, if a single BNG is placed in the first aggregation level, then the multicast is going to happen closer to the subscribers. Therefore the IPTV flow that every access location is demanding is $x_a^{IPTV} = w_a \cdot res_a \cdot IPTV_{ch}$ (percentage of users w_a from the population of sub_a connected to access location a), then the flows are going to be distributed in the network as follows

$$\begin{aligned} x_i^{IPTV} &= \sum_a u'_{a,i} x_a^{IPTV} \\ x_j^{IPTV} &= \sum_i u'_{i,j}[\sum_{t\in T_3} u_{i,t}^{L3} x_j^{CDN} + (1 - \sum_{t\in T_3} u_{i,t}^{L3}) x_i^{IPTV}] \\ x_k^{IPTV} &= \sum_a x_a^{IPTV} \sum_{t\in T_3} u_{k,t}^{L3} + x_k^{CDN}(1 - \sum_{t\in T_3} u_{k,t}^{L3}) \end{aligned} \tag{4}$$

For the first definition of the above equation, x_i^{IPTV} is equal to the bandwidth that all IPTV VLANs generate. For the x_j^{IPTV}, there are two possibilities. If a type T_3 router is placed at any of the dependent first aggregation levels ($u_{i,t}^{L3} = 1$), then the bandwidth of the IPTV flows is equal to $\sum_i x_i^{IPTV}$. If this is not the case, then the bandwidth is equal to the eq. (1). The third equation takes into account the property $\sum_{t\in T_3}(u_{i,t}^{L3} + u_{j,t}^{L3} + u_{k,t}^{L3})$. Thus, if a router that includes IPTV functionality is included at the central location k, then the IPTV bandwidth is $\sum_a x_a^{IPTV}$. Otherwise it means that an IPTV capable router has been placed in either the 1st or the 2nd aggregation level and the bandwidth for the flow is $x_k^{CDN} \simeq \frac{iptv_{ch}^{HD} + iptv_{ch}^{SD}}{2}$. From the above, it is readily seen that if users tend to watch similar programs, less bandwidth is used, decreasing the infrastructure cost. In addition, integrating the multicast functionality (i.e. $t \in T_3$), closer to the access network, decreases the bandwidth too.

Therefore the total amount of traffic that arrives in every aggregation location is $x_n = x_n^{IPTV} + x_n^{P2P} + x_n^{HSI}$.

VI. Edge Design Model

The objective function of the optimization problem is to determine the optimum deployment cost by optimally distributing the functionalities over the aggregation network

$$\min \sum_{n\in\{i,j,k\}} [\sum_c \sum_x (co^{x,c} Y_n^{x,c}) + \sum_t co_t^{L3} Y_{n,t}^x + co^{L2} Y_n^{L2}] \tag{5}$$

The above objective function includes the cost of the interfaces based on the capacity $c = [1G, 10G]$, that will be implemented in a node of layer $x = [L2, L3]$ at location n, as well as the cost of the network elements that will be placed at the same location.

A. Constraints

1) Non Linear to Linear Transformation: The problem has two sets of variables per location, $Y_{n,t}^{L3}$ and $u_{n,t}^{L3}$. The multiplication of those variables results in a non linear problem. Hence, we try to transform the model from a non-linear to a linear, by adding some extra constraints along with two big constants $BIGD_n$ (for the devices) and $BIGI_n$ (for the interfaces). For example, devices per location may range between 10-100, therefore $BIGD_n$ can be as high as 1000. Any value higher than that may result in an unnecessary increase of the search space. In the next section, we provide a heuristic to determine an approximate upper bound. The extra constraints are expressed as follows:

$$\begin{aligned} Y_{n,t}^{L3} &\le BIGD_n \\ Y_n^{L2} \le BIGD_n(1 - \sum_{t\in T_1} u_{n,t}^{L3}),\ n &\in \{i,j,k\} \\ \{t \in T_1 | n = i,j\}\ or\ \{t \in T | n = k\} & \end{aligned} \tag{6}$$

The above constraints specify that (a) the $Y_{n,t}^{L3}$ must be bigger than a big constant, and (b) in every location n there can be only an aggregation switch or an edge system. In terms of functionalities, the two sets define that for the first and second aggregation levels $n = \{i,j\}$, only IP termination functionalities can be distributed ($t \in T_1$). The last set effectively excludes the multi-edge architecture from the possible

solutions. We follow a similar approach for the interfaces.

$$\begin{aligned} Y_n^{L3,c} &\leq BIGI_n \sum_{t\in T_1} u_{n,t}^{L3}, \\ \{t \in T_1 | n = i,j\}, \{t \in T | n = k\} & \\ Y_n^{L2,c} &\leq BIGI_n(1 - \sum_{t\in T_1} u_{n,t}^{L3}),\ n \in \{i,j,k\} \\ Y_n^{L3,c} + Y_n^{L2,c} &\leq BIGI_n u_n^c,\ n \in \{i,j,k\} \end{aligned} \tag{7}$$

The first constraint, of the above set, defines that an edge system interface will be chosen if at least one edge system is installed in that location. The second constraint shows that an aggregation switch at location n must not be installed if an edge router is installed at this location. The last constraint indicates that either a 1Ge or 10Ge interface can be installed at a network device.

2) Avoiding redundancy of functionality: In order to avoid the redundancy of functionality in a path, only a certain combination of routers can be used per path. The path starts from the access node and ends at the P-Router, and can be represented by the multiplication of the following binary constants $u'_{i,j}u'_{j,k}$. In that path, each functionality must not be repeated (which is represented by the sum of the binary variables in the path must not be greater than one) and certain functionality must not be used.

$$\begin{aligned} u'_{i,j}u'_{j,k} \sum_{t\in T_1} u_{i,t}^{L3} + u_{j,t}^{L3} + u_{k,t}^{L3} &\leq 1 \\ u'_{i,j}u'_{j,k}(\sum_{t=\{A,C\}} (u_{i,t}^{L3} + u_{j,t}^{L3}) + \sum_{t\in T_2} u_{k,t}^{L3}) &\leq 1 \\ u'_{i,j}u'_{j,k}(\sum_{t=\{A,B\}} (u_{i,t}^{L3} + u_{j,t}^{L3}) + \sum_{t\in T_3} u_{k,t}^{L3}) &\leq 1 \\ u'_{i,j}u'_{j,k} \leq \sum_{t\in T_1} u_{i,t}^{L3} + u_{j,t}^{L3} + u_{k,t}^{L3} & \\ u'_{i,j}u'_{j,k} \leq \sum_{t=\{A,C\}} (u_{i,t}^{L3} + u_{j,t}^{L3}) + \sum_{t\in T_2} u_{k,t}^{L3} & \\ u'_{i,j}u'_{j,k} \leq \sum_{t=\{A,B\}} (u_{i,t}^{L3} + u_{j,t}^{L3}) + \sum_{t\in T_3} u_{k,t}^{L3} & \end{aligned} \tag{8}$$

3) Capacity Constraints: The capacity constraints define the amount of traffic that either an aggregation switch or an edge system can support. This is effectively the backplane capacity. For the case of a distributed router $u_{n,t}^{L3} = 1$ selected at location $n = \{i,j\}$, the following two constraints are defined

$$u_{n,t}^{L3}(x_n^{CDN} + x_n^{res} + x_n^{bus} + x_n^{P2P}) \leq Y_{n,t}^{L3} C_t^{L3}, \quad \forall n = \{i,j\}, t = \{A,B\} \tag{9}$$

$$u_{n,t}^{L3}(x_n^{IPTV} + x_n^{res} + x_n^{bus} + x_n^{P2P}) \leq Y_{n,t}^{L3} C_t^{L3}, \quad \forall n = \{i,j\}, t = \{C,D\} \tag{10}$$

Type A and B do have multicast functionality, therefore they will replicate the traffic. Hence, only the flows that correspond to the requested channels need to be routed x_n^{CDN}. Nonetheless, if type C or D are selected to be distributed, then the replication of the channels is done in the core location, and they will need to route the bundled flows for the multicast traffic. Starting from the top, type A router handles all traffic, therefore the throughput at a central k must be less than the total throughput of the routers of type A. Note that type A handles all types of traffic, thus it serves all the traffic $x_k^{CDN} + x_k^{bus} + x_k^{res} + x_k^{P2P}$. Similarly for the rest, with the main difference that only a subset of functionality are support per edge system.

$$\begin{aligned} u_{k,A}^{L3}(x_k^{CDN} + x_k^{bus} + x_k^{res} + x_k^{P2P}) &\leq Y_{k,A}^{L3} C_A \\ u_{k,B}^{L3}(x_k^{bus} + x_k^{res} + x_k^{P2P}) &\leq Y_{k,B}^{L3} C_B \\ u_{k,C}^{L3}(x_k^{res} + x_k^{bus} + x_k^{P2P}) &\leq Y_{k,C}^{L3} C_C \\ u_{k,D}^{L3}(x_k^{res} + x_k^{P2P}) &\leq Y_{k,D}^{L3} C_D \\ u_{k,E}^{L3} x_k^{bus} &\leq Y_{k,E}^{L3} C_E \\ u_{k,F}^{L3} x_k^{CDN} &\leq Y_{k,F}^{L3} C_F \end{aligned} \tag{11}$$

For the switching capacity things are more complicated as the capacity is affected by the distribution of the edge system functionality. For example, the closer the multicast functionality is to the subscribers, the closest the replication takes place. Equation (VI-A3) indicates the switching capacity constraints. The first is related to the locations in the first aggregation level i. This constraint indicates that if an edge system is installed $u_{i,t}^{L3} = 1$ then there is no need to care about the switching capacity. The second constraint is related to the locations in the second aggregation level j. If multicast replication takes place at lower level $u_{i,A}^{L3} + u_{i,B}^{L3} = 1$, then only a single flow per channel needs to be switched x_{CDN}. If it is taking place at the core location, $u_{i,C}^{L3} + u_{i,D}^{L3} = 1$ or $\sum_{t\in T_3} u_{k,t}^{L3} = 1$ (T_3 is the set of the routers that support multicast), then a separate flow per subscriber for the IPTV needs to be switched. Finally, if a router is placed at the second aggregation level $u_{j,t}^{L3} = 1$, then there is no need to switch any traffic. If a switch is placed in a core location and also an MSE or Video BNG is placed at the same location, then they will route business and IPTV traffic and therefore those two portions need to be excluded from the switching capacity $x_k^{bus}(1-\sum_{t\in T_2} u_{k,t}^{L3}) + x_k^{CDN}(1-\sum_{t\in T_3} u_{k,t}^{L3})$. Those constraints can be found in the inequality set (VI-A3) at the top of the next page.

4) Subscriber Termination/VLAN Capacity: The following equalities determine the number of VLANs per location. A distinction is made between the residential and the business VLANs.

$$\begin{aligned} S_i^{res} = \sum_a u'_{a,i} S_a^{res}, S_i^{bus} &= \sum_a u'_{a,i} S_a^{bus} \\ S_j^{res} &= \sum_i u'_{i,j}(1 - \sum_{t\in T_1} u_{i,t}^{L3}) S_i^{res} \\ S_j^{bus} &= \sum_i u'_{i,j}(1 - \sum_{t=\{A,C\}} u_{i,t}^{L3}) S_i^{bus} \\ S_k^{res} &= \sum_j (1 - \sum_{T_1} u_{j,t}^{L3}) S_j^{res} \\ S_k^{bus} &= \sum_j (1 - \sum_{t=\{A,C\}} u_{j,t}^{L3}) S_j^{bus} \end{aligned} \tag{13}$$

For routers of type A, B, C and D which may exist at any of the levels $n = \{i,j,k\}$ the subscriber termination constants are similar to the capacity ones. They are expressed as follow $u_{n,t}^{L3} S_n \leq Y_{n,t}^{L3}$, where $S_n = S_n^{res} + S_n^{bus}$ for Type A/C, $S_n = S_n^{res}$ for B/D. $S_n = sum_a BUS_a$ for Type E,

$$(1-\sum_{t\in T_1} u_{i,t}^{L3})(x_i^{IPTV}+x_i^{res}+x_i^{bus}+x_i^{P2P}) \le Y_i^{L2} C_{ags}$$

$$\sum_i (u_{i,j}(u_{i,A}^{L3}+u_{i,B}^{L3})x_k^{CDN} + (u_{i,C}^{L3}+u_{i,D}^{L3})x_i^{IPTV}) + \sum_{t\in T_3} u_{k,t}^{L3} x_j^{IPTV} + (x_j^{res}+x_j^{bus}+x_j^{P2P})(1-\sum_{t\in T_1} u_{j,t}^{L3}) \le Y_j^{L2} C_{ags} \quad (12)$$

$$(1-\sum_{t\in T_1} u_{k,t}^{L3})(x_k^{CDN}+x_k^{res}+x_k^{bus}+x_k^{P2P}) + x_k^{bus}(1-\sum_{t\in T_2} u_{k,t}^{L3}) + x_k^{CDN}(1-\sum_{t\in T_3} u_{k,t}^{L3}) \le Y_k^{L2} C_{ags}$$

and $S_n = \sum_a w_a RES_a$ for Type F. The latter is expressed lie that as all these must be terminated at the Video BNG, because only a portion w_a of subscribers are watching a channel and the total number of IPTV VLANs are determined by $\sum_a w_a res_a$

For VLANs that need to be switched, the case is a bit more complicated. For the second level of aggregation and if a customer has been already been terminated in the previous level, then $J-1$ VLANs will be required to switch traffic among the same level of switches and one more that will send the traffic to the P-router. If however only the residential subscribers have been terminated in the previous level ($u_{i,B}^{L3}+u_{i,D}^{L3}=1$) then the business VLANs will still remain ($S_i^{bus}+J$). The constraints for the VLAN capacity are mentioned in the next page, inequality (14)

5) Link and Interface Constraints: Each access location is assumed to have N_a devices and each location has both uplink and downlink interfaces. The minimum number of downlink interfaces from the first aggregation level is equal to the number of access devices that are connected to this location, i.e. $u'_{a,i}N_a$. In every of the first and second level aggregation locations either a router or a switch or nothing is placed. Therefore, two of the $Y_{n-1,n}^{L3,c,dn}$, $Y_{n-1,n}^{L2,c,dn}$ and $Y_{i,j}^{0,dn}$ are going to be zero.

$$u'_{a,i}N_a \le \sum_c (Y_{a,i}^{L3,c,dn} + Y_{a,i}^{L2,c,dn})$$

$$u'_{i,j}(\sum_{t\in T_1} Y_{i,t}^{L3} + Y_i^{L2}) \le Y_{i,j}^{0,dn} + \sum_c Y_{i,j}^{L3,c,dn} + Y_{i,j}^{L2,c,dn}$$

$$\sum_{t\in T_1} Y_{j,t}^{L3} + Y_j^{L2} + \sum_i Y_{i,j}^{0,dn} \le \sum_c Y_{j,k}^{L3,c,dn} + Y_{j,k}^{L2,c,dn} \quad (15)$$

The network elements also have line card capacity. Each line card is a separate module that has a specific capacity of ports and functionalities. We assume two types of line cards, one for switches and one for routers. Each element may take ports of the same type, 1Ge and 10Ge, and the port capacity is determined by multiplying the number of line card slots to the number of ports per line card. Since the number of ports per line card is affected by the bandwidth of the interfaces, the same will hold for the total number of ports per network element. The constraints are expressed as follows.

$$\sum_a u'_{a,i} Y_{a,i}^{L2,c,dn} + \sum_j u'_{i,j} Y_{i,j}^{L2,c,up} \le Y_i^{L2} P_i^{L2,c}$$

$$\sum_a u'_{a,i} Y_{a,i}^{L3,c,dn} + \sum_j u'_{i,j} Y_{i,j}^{L3,c,up} \le \sum_{t\in T_1} Y_{i,t}^{L3} P_{i,t}^{L3,c}$$

$$\sum_i u'_{i,j} Y_{i,j}^{L2,c,dn} + Y_k^{L2,c,up} \le Y_j^{L2} P_j^{L2,c} \quad (16)$$

$$\sum_i u'_{i,j} Y_{i,j}^{L2,c,dn} + Y_k^{L2,c,up} \le \sum_{t\in T} Y_{j,t}^{L3} P_{i,t}^{L3,c}$$

As we can see above, the first constraint makes sure that there is at least one link between the access node and the first aggregation level. However if the solution determines that no device is going to be installed, then $d \to \infty$ and $Y_i^0 = 1$.

6) Interface Capacity Constraints: The interface capacity constraints are the most complicated, since they must be determined without any knowledge of the placement of the devices. In every device there are going to be interfaces that connect both the uplink and the downlink.

$$Y_n^{x,c} = Y_{n-1,n}^{x,c,dn} + Y_{n,n+1}^{x,c,up} \quad (17)$$

gives the total number of interfaces of capacity c per location n. Moreover, in all cases the uplink capacity is equal to the downlink capacity of the higher layer. For instance for the link between the access location a and first, the constraint is as follows $u'_{a,i}(x_a^{IPTV}+x_a^{res}+x_a^{bus}+x_a^{P2P}) \le \sum_c \sum_x Y_{a,i}^{x,c,dn} C^c$. For the links between i and j, The bandwidth of the uplink of the first aggregation level is calculated to be: a) the internet and business traffic (non local flows); b) if a multicast functionality is placed at the first level, $u_{i,A}^{L3}+u_{i,B}^{L3}=1$, the flow of the channels from the second aggregation level x_j^{CDN}; c) if there is not a multicast functionality $1-\sum_{t\in T_1} u_{i,t}^{L3}=1$ on the first level, the IPTV flow of each subscriber; d) if there is routing functionality $\sum_{t\in T_1} u_{i,t}^{L3}=1$ the P2P portion that is not first level local $(1-p_1)x_i^{P2P}$. For simplicity we mention only the downlink constraints (the uplink i to j are the same). Similarly the remaining interface capacity constraints are calculated for the other two aggregation levels.

VII. Model Implementation

As a mixed integer programming model, the problem is considered to be NP-hard (polynomial time hard). We applied two heuristics to decrease the number of variables. After this phase, we used the branch-cut-algorithm from the CPLEX 11 engine. Finally, we used the CPLEX presolver to eliminate redundant rows and columns. In the following we describe the two heuristics:

$$(1-\sum_{t\in T_1} u_{i,t}^{L3})S_i \le Y_i^{L2} sub_{ags}$$

$$\sum_i u'_{i.j}((u_{i,A}^{L3}+u_{i,C}^{L3})J+(u_{i,B}^{L3}+u_{i,D}^{L3})(S_i^{bus}+J))+\sum_{t\in T_1} u_{k,t}^{L3}S_k \le Y_i^{L2} sub_{ags} \quad (14)$$

$$u'_{i,j}((u_{i,A}^{L3}+u_{i,B}^{L3})x_j^{CDN}+\sum_{t\in T_1} u_{i,t}^{L3}(1-p_1)x_i^{P2P}+(1-\sum_{t\in T_1} u_{i,t}^{L3})x_i^{IPTV}+x_i^{P2P})+x_i^{HSI}+x_i^{BUS} \le \sum_c\sum_x Y_{i,j}^{x,c,dn}C^c \quad (18)$$

$$x_j^{HSI}+x_j^{BUS}+(1-\sum_{t\in T_1} u_{j,t}^{L3})(1-p_1-p_2)x_j^{P2P}+(1-\sum_{t\in T_3} u_{j,t}^{L3})x_k^{CDN}+\sum_{t\in T_3} u_{j,t}^{L3}x_j^{IPTV} \le \sum_c\sum_x Y_{i,k}^{x,c,dn}C^c \quad (19)$$

$$x_k^{HSI}+(1-\sum_{k=1}^{3} p_k)x_k^{P2P}+x_k^{CDN}+x_k^{BUS} \le \sum_c\sum_x Y_k^{x,c,up}C^c \quad (20)$$

TABLE IV
INPUT PARAMETERS FOR AGGREGATION NETWORK CHARACTERISTICS AND INTERFACE PROPERTIES

Symbol	Descr/Units	Small SP	Big SP
A	Locations	1000	2000
I	Locations	6	50
J	Locations	3	5
K	Locations	1	
SUB	Subscribers	200K	400K
res_a	Subscribers	160K (80%)	320K (80%)
bus_a	Subscribers	40K (20%)	80K (20%)
S_a	Subscribers	200	
Ch	Channels	100	
w_i	% of viewers	50%	
C_{1G}^{L2}/C_{10G}^{L2}	Bytes	1G/10G	
C_{1G}^{L3}/C_{10G}^{L3}	Bytes	1G/10G	
$co^{L2,1G}$	$	1K	
$co^{L3,1G}$	$	2K	
$co^{L2,10G}$	$	2K	
$co^{L3,10G}$	$	4K	

Heuristic 1: Minimizing number of variables. Generally the number of access locations A varies from hundreds to thousands making the problem hard to be solved. For this, we implemented a clustering algorithm to determine the number of access locations. The algorithm starts with a single cluster $cl = 1$, increases the number of clusters in every iteration, and terminates when both the time to convergence is smaller than a predefined value t' and the optimality tolerance is also smaller than a predefined value ac'. This means that the number of access locations will be $A = cl * I$, where cl is the number of clusters. The divisive clustering procedure is as follows

procedure CLUSTER(I, t', ac')
  $cl \leftarrow 1$
  $A \leftarrow I$
  Run model to determine t and ac
  while $t \le t'$ and $ac \le ac'$ **do**
    $cl \leftarrow cl + 1$
    Perform K-means clustering
    $A \leftarrow cl * I$
    Run model to determine t and ac
  end while
  return cl
end procedure

In every iteration, the K-means instance [16] is called with attributes related to the technical characteristics of the access devices. In our case we used as attributesthe L2 subscriber capacity of the access technologies. After running the optimization model several times, we determined that if the model does not converge to a solution in less than $t' = 120sec$, it will hardly manage to converge (soft threshold). Finally the optimality tolerance was set to 10^{-6}. In any case, a potential error in terms of Kbps will not affect the solution of the model (since all system capacities are at the order of thousands of Gbps).

Each access location is assumed to incorporate the same access device. However, the definition of the "location" in our model is not strictly a geographical location. Therefore, if in the same geographic location different access technologies are implemented, then these are treated as two separate locations.

Heuristic 2: Minimizing execution time. The optimization has integer variables that are either binary (qualitative) or non-binary (quantitative). In order to transform the problem to integer programming model, the values of $BIGI_n^c$ and $BIGD_n$ were introduced. If these values are selected to be very big, then the solver would require a significant amount of time to converge to an optimal solution. If the values are selected to be small, the solution would not be the optimal (or the problem could be infeasible). Thus the values were selected by taking into account the amount of traffic that flows and the number of VLANs in the corresponding location. A logical upper bound for the number of devices per aggregation location can be derived from

$$BIGD_n = max\{\frac{X_n}{min\{C_t^{L3}, t\in T\}}, \frac{sub_n}{min\{sub_t^{L3}, t\in T\}}\} \quad (21)$$

For an upper bound on the number of interfaces, we used the above upper bound multiplied by the most numbers of ports

$$BIGI_n^c = BIGD_n * max\{P_t^{c,L3}, t\in T\},\ c\in\{1G, 10G\} \quad (22)$$

Using the above values the solver was determining the solution in less than 1min on an Intel Core2 Duo T7700 with 4GB RAM.

VIII. EVALUATION AND RESULTS

In this section, we use two exemplar architectures to evaluate the aforementioned edge design model. The first

TABLE V
NETWORK ELEMENT FEATURES

Edge System	Cost $co_{L3,t}$(K$)	Capacity C_t(Gbps)	IP Term. $subL3,t$	1Ge $P^{1G}_{L3,t}$	10Ge $P^{10G}_{L3,t}$
All	300	160	64000	192	24
HSI/VideoBNG	600	640	64000	480	64
HSI/MSE	220	40	32000	96	12
HSI	340	280	32000	140	28
MSE	180	20	4000	96	12
Video BNG	200	280	10000	140	28
	Cost K$	Switching C_t(Gbps)	VLAN $vlan$	1Ge P^{1G}_{L2}	10Ge P^{10G}_{L2}
Agg. switch	270	280	64000	140	28

architecture, namely "small SP", is based on an EU ISP with 200K subscribers. The second architecture, namely "big SP" is based on another EU ISP with twice the number of subscribers, and ten times more first level aggregation locations. We use several traffic scenarios per subscriber (very low up to very high). We performed this investigation such that a future traffic growth is incorporated in our modeling. In table IV, the details of the architectures are given. In table V the values for each edge system are showcased. These values were provided by vendor and are representative of high-end systems at the time of submission[2]. The number of interfaces in the table is derived by multiplying the number of ports per line card with the capacity of line cards per edge system. In addition the local (P2P) traffic does not exceed 20% of the total internet traffic [18] and P2P locality probabilities p_k get values from a uniform distribution. The questions that we try to address are as follows

- Which of the functionalities should the provider distribute closer to the subscriber?
- Does the size of the ISP affect the choice of the optimal functionality placement?
- Should the providers choose faster (but more costly) line cards or a faster backplane in an Edge Router?
- What is the usage (or how much redundancy exists) per edge system for throughput, termination capabilities and port availability?
- How does a change in the number of subscribers affect the intelligence distribution?

In Fig. 4 the cost of multiple architectures is shown for the big service provider. The centralized and 2nd level distributed architectures are investigated. The 1st level aggregation is the first level above the access network and similarly for the rest. Due to the number of 1st level aggregation locations, we found that the cost of distributing the functionalities at the 1st level was higher than in the other two. For this reason, the centralized and the 2nd level distributed architectures were chosen to be plotted. In the case of a small amount of multicast flows, the centralized unclustered seems to be the cost optimal architecture. This is because most of the traffic goes through the central locations to the backbone. However, an increase of multicast traffic leads to significant increase in the cost of the centralized architecture.

[2]The prices of those systems may vary over time and offer. However those values are representative at the time of submission through private communication with at least one vendor. An incremental change of those values does not affect the formulation of the problem.

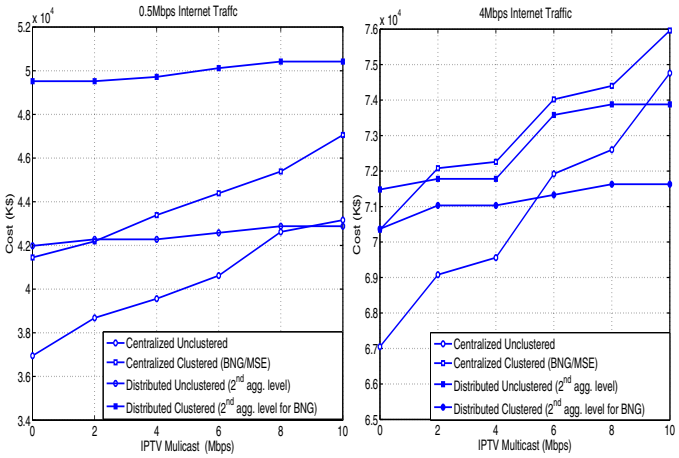

Fig. 4. Cost comparison for the big service provider per aggregation network for 0.5Mbps and 4Mbps of local and non local traffic (Internet).

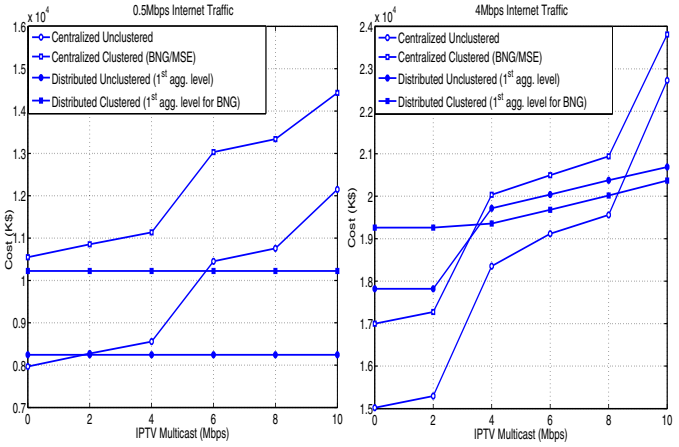

Fig. 5. Cost comparison for the small service provider per aggregation network for 0.5Mbps and 4Mbps of local and non local traffic (Internet).

On the other hand, the rate of cost increase for the distributed is smaller, and after some point it becomes cheaper. The point at which the graphs meet is important. It effectively shows at which point the SP will need to distribute the functionalities. For small amounts of internet traffic (0.5Mbps per user) the meeting point is around 8 Mbps of multicast traffic, and for a much larger internet traffic (4Mbps per user) it is around 4Mbps. In such large amounts of traffic, either the capacity constraints of the elements or the port constraints are met, leading to more devices/interfaces. Therefore, *the distribution of IP intelligence is preferable, because (a) it alleviates the bottleneck due to P2P traffic that in any other case would need to go the central location and (b) the multicast traffic is replicated closer to the access network*. Finally, clustering the devices does not have a positive impact (compared to the unclustered case) to the cost. In the Big SP there is enough "capacity" (by "capacity" here we mean the shadow price of the subscriber, traffic and port constraints) to support all different services.

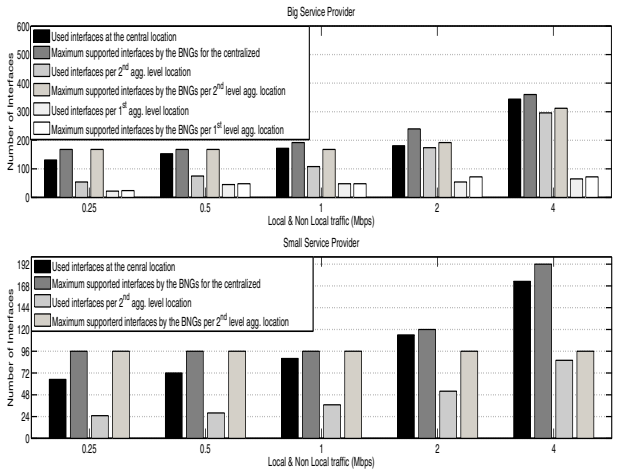

Fig. 6. Used ports vs available ports in an aggregation location with variable Internet and constant IPTV=6Mbps traffic.

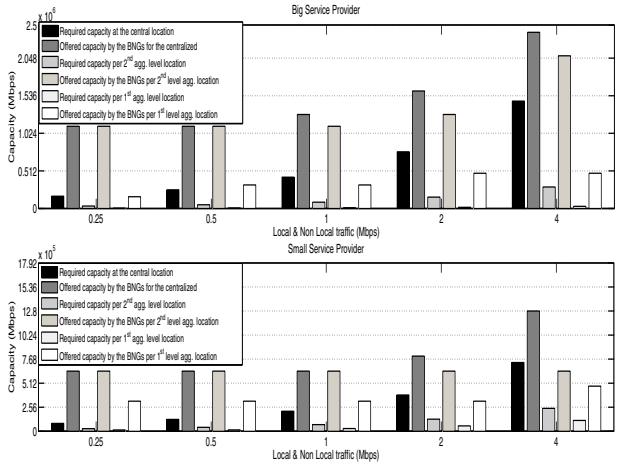

Fig. 7. Traffic capacity and bandwidth of edge locations plotted with variable Internet traffic and constant IPTV=6Mbps traffic.

In Fig. 5 we performed the same comparison for a smaller provider. In this architecture the differences among the implementations become more apparent. For low values of multicast traffic, centralized is optimal. But the increase of multicast traffic affects significantly the cost of a centralized architecture. Moreover assuming a constant internet bandwidth, an increase of multicast traffic would not affect the cost of a distributed architecture. Getting deeper into this issue, the IPTV flows are unicast, until they reach a router, in which case they become multicast. The number of unicast flows are at most equal to the number of channels the SP provides. For the worst case scenario of 100 different channels to be watched at HD (10Mbps), the total traffic will not exceed the capacity of an interface (1Gbps or 10Gbps). A cost breakdown showed that it is optimal to use 10Gbps interfaces, which leaves plenty of bandwidth for even more channels to be offered without having to increase the number of interfaces.

Since the replication functionality is closer to the access networks, the increase of multicast traffic will only affect the downlink ports of the routers. A single 1Gbps port, that is connected to an access location, will be able to handle 100 HD IPTV flows. Since every DSLAM is having a subscriber capacity of 200 customers and 50% of customers are assumed to be registered to watch TV, then there is no effect on the number of ports. Combining those results with the sensitivity analysis from Fig. 6, it looks that there are enough slots to handle an even bigger amount of multicast flows (therefore interfaces). The optimal cost will only be affected by the cost of the interfaces (shadow price). Thus, *distributing the replication functionality seems to be cost efficient*. In similar traffic scenarios for the small SP, the distributed architecture costs two times less than the centralized.

In Fig. 6, sensitivity analysis was performed to determine if and how much the number of ports is affecting the solution of the problem. In this figure the unclustered case is plotted for variable internet traffic. The bars are presented in sets; the first bar is the number of used ports at an aggregation location and the second bar is the sum of the number of available interfaces at the same location. Thus, the second bar must be always larger than the first bar, in order for a constraint to be satisfied. The steps of the y-axis are equal to the 10G interface capacity of a BNG (24 ports).

It is clearly shown that in the centralized unclustered topology, and for both a small SP and a big SP, the number of required interfaces determines the number of edge systems to be used (for example for 0.5, 1 and 4Mbps if one less BNG was used the port constraint would have been violated). A similar conclusion can be derived for the distributed unclustered case at the first aggregation level. Therefore, *increasing the interfaces bandwidth such that they handle more traffic or increasing the capacity of the chassis for line cards would decrease the number of required edge systems (and therefore the cost)*.

In Fig. 7 the amount of traffic that flows through a specific location and the amount of traffic that the edge systems can handle at the same location is presented. Similarly to the above case, the steps of the y-axis are equal to the backplane capacity of a BNG (160Gbps). In most cases, it is clearly shown that the bandwidth of a router does not have an effect on the number of edge systems. For example, even if half of the edge systems are being used, the set of the bandwidth constraints would still not be violated. Therefore, *an increase in the capacity of the edge systems does not have a significant impact on the aggregation networks.*

In Fig. 8, it is shown that for the Big SP with low volumes of internet traffic per subscriber and centralized topology, the number of edge systems seems to be affected by the VLAN capacity. However, as the traffic increases, the interface capacity constraint becomes the one that requires the addition of more edge systems. A similar pattern is followed by the small SP. An interesting point is that *the subscriber termination capabilities have no effect on the distributed architectures*.

Finally, in table IV, a breakdown of costs is shown for 1Mbps HSI and 6Mbps multicast traffic. For both the small and the big SP and when a distributed topology is implemented, the edge router cost becomes dominant. Due to the fact that the traffic is distributed over the access locations, when distributing the functionalities closer to the subscribers,

TABLE VI
COST BREAKDOWN IN THOUSANDS OF DOLLARS FOR IPTV=6MBPS AND HSI=1MBPS

	Agg. Switch (L2)	L2 1G	L2 10G	Edge Routers (L3)	L3 1G	L3 10G
Centr. Small SP	6480	1884	276	1200	0	276
Distr. Small SP (2nd agg. level)	5400	1884	42	3600	0	444
Distr. Small SP (1st agg. level)	1350	0	168	3600	2784	0
Centr. Big SP	31050	0	5230	2100	0	460
Distr. Big SP (2nd agg. level)	27540	0	4570	10500	0	1080
Distr. Big SP (1st agg. level)	3240	0	480	15000	7100	0

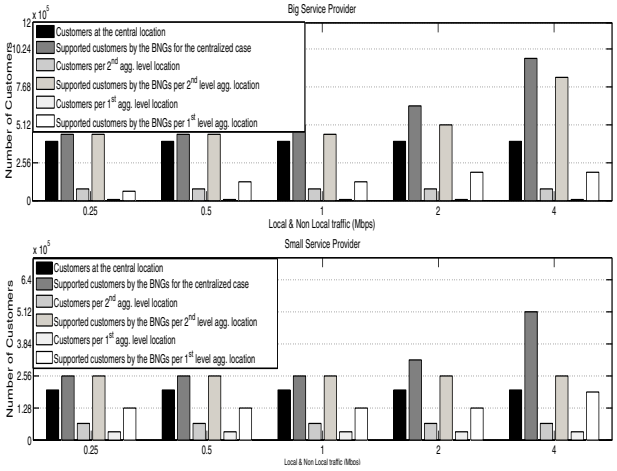

Fig. 8. Number of subscribers that are terminated per location with variable Internet traffic and constant IPTV=6Mbps traffic.

it is cost optimal to use 1G line cards. In case of a centralized topology, the optimal solution consists of 10G line cards for the edge systems.

IX. DISCUSSION

Our results showed that re-engineering the edge infrastructure is of importance in order to accommodate intelligent aggregation. More specifically:

1) For a small SP the distributed architecture is cheaper, since only 2-3 BNGs are required per 1st aggregation level. For the big SP the distributed architecture is also less expensive, but the difference with the centralized is smaller.
2) For a small SP the number of BNGs is mainly affected by the number of ports. Yet, if the SP wants to add another edge system in order to have spare ports for resiliency or/and service flexibility, the total cost would not surpass the cost of a centralized architecture.
3) For the big SP, the number of free interfaces are less and the subscriber termination constraints have a small shadow price, leading to smaller flexibility.
4) Distributing the functionalities simplifies the operational flexibility since the network administrator only has to deal with the interface constraints. Moreover, for both the small and the big SP, the cost of the distributed architecture does not increase when the number of IPTV channels increase.
5) Centralized single-edge architectures have been proven to reach scalability limits, whereas clustered, multi-edge or distributed architectures offer better architectural scalability, since smaller number of subscribers are terminated per system.

Distributed architectures have some further advantages. Scalability from the system resources point of view is better, since the amount of state information (memory overhead) in the edge system can be substantially less than in a centralized architecture. In fact distributed architectures benefit from policy enforcement close to the subscriber (no need to backhaul traffic that can be dropped).

One other important aspect is related to the operation of the network. More specifically, single-edge architectures allow provisioning of all subscribers and services on a single system, thus facilitating QoS provisioning using hierarchical schedulers, as well as minimal operational expenditures (OPEX). On the contrary, multi-edge architectures lead to distributed provisioning for services destined to the same subscriber, thus requiring intelligent protocols for proper QoS management and generation of significant amount of state information.

One can name more aspects that our study has not included, but due to space limitations our main contribution was to construct a modularized model and provide several insights for the service providers. We plan to extend this modeling approach into a 4G mobile architecture [9], [14], as well as investigate how caching architectures [25] affect the flow distribution.

X. CONCLUSION

This paper focused on a quantitative, cost-optimal hierarchical network model for differentiated services. The model was based on edge "systems", which can receive extra functionalities as attached sub-systems. The modeling approach incorporates physical characteristics, bandwidth and VLAN/IP termination capacity, L2 versus L3, multicast and business functionalities. In order to scale down the problem, and decrease the execution time, we applied a set of heuristics that are based on clustering algorithms.

Our results indicate that the widely implemented centralized Ethernet based single-edge architecture has reached its scalability limits, hence distributing the functionalities closer to the subscriber is far more efficient. A small SP will benefit more from the intelligence distribution than a larger SP and, independently of the size, the SPs should cluster the business functionalities on a different edge system. Finally, we showed that SPs would receive only a marginal benefit from higher capacity edge systems, and should rather focus on the proper allocation of the functionalities in their aggregation network.

REFERENCES

[1] "Migration to Ethernet-based DSL aggregation," TR-101, DSL forum, technical report, Apr. 2006.

[2] "Cisco visual networking index: usage," technical report, 2010.
[3] "Sandvine global Internet phenomena report," technical report, 2010.
[4] P. Arberg, T. Cagenius, O. Tidblad, M. Ullerstig, and P. Winterbottom, "Network infrastructure for IPTV, *Ericsson Review*, vol. 84, no. 3, 2007.
[5] M. Cha, P. Rodriguez, S. Moon, and J. Crowcroft, "On next-generation telco-managed P2P TV architectures," in *Usenix 2008 International Worskop on Peer-to-Peer Systems*.
[6] U. Chaudhary, I. Papapanagiotou, and M. Devetsikiotis, "Flow classification using clustering and association rule mining," in *Proc. 2010 IEEE CAMAD*, pages 1–5.
[7] F. De Greve, F. Van Quickenborne, F. De Turck, I. Moerman, and P. Demeester, "Aggregation network design for offering multimedia services to fast moving users," *Springer Quality of Service in Multiservice IP Networks*, pp. 235–248, 2005.
[8] B. Fenner, M. Handley, H. Holbrook, and I. Kouvelas, "RFC4601 protocol independent multicast - sparse mode (PIM-SM)," protocol specification, IETF, 2006.
[9] Z. Ghebretensaé, J. Harmatos, and K. Gustafsson, "Mobile broadband backhaul network migration from TDM to carrier Ethernet," *IEEE Commun. Mag.*, vol. 48, no. 10, pp. 102–109, 2010.
[10] I. Gódor and G. Magyar, "Cost-optimal topology planning of hierarchical access networks," *Comput. Oper. Res.*, vol. 32, no. 1, pp. 59–86, 2005.
[11] S. Han, S. Lisle, and G. Nehib, "IPTV transport architecture alternatives and economic considerations," *IEEE Commun. Mag.*, vol. 46, no. 2, pp. 70–77, 2008.
[12] C. Hermsmeyer, E. Hernandez-Valencia, D. Stoll, and O. Tamm, "Ethernet aggregation and core network models for effcient and reliable IPTV services," *Bell Labs Technical J.*, vol. 12, no. 1, pp. 57–76, 2007.
[13] Y. Huang, Y. Chen, R. Jana, H. Jiang, M. Rabinovich, A. Reibman, B. Wei, and Z. Xiao, "Capacity analysis of mediagrid: a P2P IPTV platform for fiber to the node (FTTN) networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 1, Jan. 2007.
[14] D. Hunter, A. McGuire, and G. Parsons, "Carrier Ethernet for mobile backhaul [guest editorial]," *IEEE Commun. Mag.*, vol. 48, no. 10, pp. 92–93, 2010.
[15] A. A. V. Ignacio, V. J. M. F. Filho, and R. D. Galv ao, "Lower and upper bounds for a two-level hierarchical location problem in computer networks," *Comput. Oper. Res.*, vol. 35, no. 6, pp. 1982–1998, 2008.
[16] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 881–892, 2002.
[17] G. Lee, C. Lee, W. Rhee, and J. Choi, "Functional architecture for NGN-based personalized IPTV services," *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 329–342, 2009.
[18] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On dominant characteristics of residential broadband Internet traffic," in *Proc. 2009 ACM SIGCOMM Conference on Internet Measurement Conference*, pp. 90–102.
[19] M. Menth, R. Martin, and J. Charzinski, "Capacity overprovisioning for networks with resilience requirements," in *Proc. 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 87–98.
[20] M. Minoux, "Networks synthesis and optimum network design problems: models, solution methods and applications," *Networks*, vol. 19, no. 3, pp. 313–360, 1989.
[21] R. Nagarajan and S. Ooghe, "Next-generation access network architectures for video, voice, interactive gaming, and other emerging applications: challenges and directions," *Bell Labs Technical J.*, vol. 13, pp. 69–86, Spring 2008.
[22] L. Newell and M. Sif, "Optimizing the broadband aggregation network for triple-play services," *Annual Review of Broadband Commun.*, p. 15, 2006.
[23] I. Papapanagiotou, R. Callaway, and M. Devetsikiotis, "Chunk and object level deduplication for web optimization: a hybrid approach," in *2012 International Communications Conference*.
[24] I. Papapanagiotou and M. Devetsikiotis, "Aggregation design methodologies for triple play services," in *2010 IEEE Consumer Communications and Networking Conference*.
[25] I. Papapanagiotou, E. Nahum, and V. Pappas, "Smartphones vs. laptops: comparing web browsing behavior and the implications for caching," in *2012 SIGMETRICS*.
[26] M. Pioro and D. Medhi, *Routing, Flow and Capacity Design in Communication and Computer Networks*. Elsevier Inc. and Morgan Kaufmann Publisher, 2004.
[27] M. Resende and P. Pardalos, *Handbook of Optimization in Telecommunications*. Springer Verlag, 2006.
[28] B. Sanso and P. Soriano, *Telecommunications Network Planning*. Kluwer Academic Publishers, 1999.
[29] L. Shi, C. Liu, and B. Liu, "Network utility maximization for triple-play services," *Computer Commun.*, vol. 31, no. 10, pp. 2257–2269, 2008.
[30] D. Smith, "IP TV bandwidth demand: multicast and channel surfing," in *2007 IEEE INFOCOM*, pp. 2546–2550.
[31] E. Veloso, "A hierchical characterization of a live streaming media worklad," in *2002 ACM SIGCOMM IMW*.
[32] K. Xiong, H. Perros, and S. Blake, "Bandwidth provisioning in ADSL access networks," *Int'l J. Network Management*, vol. 19, no. 5, pp. 427–444, 2009.

Ioannis Papapanagiotou (S05, M12) received the Dipl.Ing. degree in Electrical and Computer Engineering from the University of Patras, Greece in 2006, and the M.Sc. and Ph.D. degrees in Computer Engineering/Operations Research from North Carolina State University in 2009 and 2012, respectively. He has been awarded the best paper awards in IEEE GLOBECOM 2007 and IEEE CAMAD 2010, and he is also the recipient of the IBM PhD Fellowship, Academy of Athens and A. Metzelopoulos scholarships. He is currently an Advisory Engineer in IBM. His interests lie in the areas of network design, data deduplication, and mobile traffic analysis.

Matthias Falkner (S97, M00) is a distinguished Technical Marketing Engineer in Cisco's Services Routing Technology Group (SRTG). He currently focuses on Service Provider architectures, and on next-generation broadband architectures in particular. Matthias works on the evolution of Cisco's midrange router portfolio, specializing on technologies that are particularly relevant for Service Providers, such as High-Availability, Broadband Technologies or Quality of Service. Matthias' research interests are in the area of network architecture evolution and traffic modeling. Matthias holds a Ph.D. in Systems and Computer engineering from Carleton University, Canada, and an MSc in Operations Research & Information Systems from the London School of Economics and Political Science, UK.

Michael Devetsikiotis [S85, M94, SM03, F12] received the Dipl.Ing. degree in electrical engineering from the Aristotle University of Thessaloniki, Greece, in 1988, and the M.Sc. and Ph.D. degrees in electrical engineering from North Carolina State University, Raleigh, in 1990 and 1993, respectively. In 1993 he joined the Dept. of Computer Engineering at Carleton University, Ottawa, Canada as a Postdoc Fellow. He later became an Adjunct Research Professor, Assistant and Associate Professor in 1996 and 1999 respectively. In 2000, he joined the Dept. of Electrical & Computer Engineer at NC State University as an Associate Professor, and in 2006 he became a Professor. Dr. Devetsikiotis is a member of the honor societies of Eta Kappa Nu, Sigma Xi, and Phi Kappa Phi. He served as Chairman of the IEEE Communications Society Technical Committee Communication Systems Integration and Modeling, and as a member of the ComSoc Education Board. He has served as an editor and in the editorial board of various journals, and has chaired several IEEE conferences.