IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 28, 2018, accepted January 9, 2019, date of publication January 23, 2019, date of current version February 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894396

Resource Provisioning for a Multi-Layered Network

CHANG XING[1], RONALD G. ADDIE[2], (Member, IEEE), YU PENG[1], RONGPING LIN[3], FAN LI[1], WENJIE HU[1], VYACHESLAV M. ABRAMOV[4], AND MOSHE ZUKERMAN[1], (Fellow, IEEE)

[1]Department of Electronic Engineering, City University of Hong Kong, Hong Kong
[2]Department of Mathematics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[3]School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[4]School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia

Corresponding author: Chang Xing (c.xing@my.cityu.edu.hk)

This work was supported by grants from the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China, under Project CityU 123012 and Project CityU 11200417. Most of the contribution of V. M. Abramov to this paper was made during his service at the City University of Hong Kong.

ABSTRACT Given the growth, complexity, and size of the Internet, new methodologies are needed to support cost-effective resource provisioning. This paper provides a cost-based polynomial-time heuristic algorithm for resource provisioning optimization called multi-layered market algorithm (MMA). The MMA is solvable for multi-layered, multi-technology, and practical-sized networks, where the traffic is modeled as a combination of constant bit-rate and variable bit-rate (VBR) traffic streams. A VBR stream is modeled either by a Gaussian process or by a Poisson Pareto burst process (PPBP) which under certain parameter values is long-range dependent – a known characteristic of the Internet traffic streams. The consideration of VBR traffic models in a multi-layered network optimization is a key novel aspect of MMA. The MMA considers a range of transport technologies operating in layers and traffic sharing schemes. The MMA implements flow-size-based routing where flows according to their sizes are routed independently. As routing affects resource requirement, such considerations are important for resource provisioning by the given cost models. The complexity resulting from these considerations, including layering and PPBP traffic, requires a simplified design philosophy which in this paper, is based on adopting the shortest path routing in each layer. This is achieved by MMA which is based on an iterative algorithm, and resource provisioning that is performed link-by-link in all layers. As a benchmark for MMA, we provide an integer linear programming (ILP) formulation for a multi-layered network optimization problem with fixed end-to-end demands. The MMA is validated by comparing its solutions to those ILP results in different variants of a six-node network, and its software is verified using double-entry bookkeeping – a method commonly used in accounting systems. The MMA runs on a platform called network mark-up language, which enables visualization and further validation of the results.

INDEX TERMS Integer linear programming, multi-layered network, optimization, Poisson Pareto burst process.

I. INTRODUCTION

The increasing complexity of the Internet imposes challenges on network designs and operations. This Internet complexity is a result of its multi-layered multi-technology growing architecture that faces the explosion of multimedia services and applications serving billions of human and non-human users. These users that are connected to the Internet impose exponentially increasing traffic demands, the nature of which is highly variable. To meet quality of service (QoS) requirements of customers for the various services and to cope with increasing demand, service providers often "throw capacity at the problem" which has proven inefficient and drives CAPEX and OPEX up [1]. There is a need for a methodology based on realistic traffic and network modeling that leads to an efficient and scalable design algorithm that achieves a cost-effective resource provisioning for multi-layered networks. In addition to the explosive growth of traffic and services, the service providers also find it difficult to keep up with the complexity associated with frequent technology developments. Such developments are likely to lead to significant cost savings if service providers could carefully choose the technologies for future network.

Lower layers have lower cost per bit, but this cost saving is achieved only if resources are sufficiently utilized at the lower layers. This cost saving is achieved only if resources are sufficiently utilized by the lower layers. For example, in the case of the WDM layer, using a 40 Gb/s wavelength for a 10 Mb/s stream may not save costs. Nevertheless, traffic can be aggregated in higher layers and efficiently utilized in the lower layers. In addition, the modularity and clear interface between independent layers have significant benefit in technology development evolution as one layer can be upgraded, modified or even eliminated independently of the other layers.

An illustration of a potential evolutional trend of the structure of a layered Internet core network has been provided in Fig. 1 and Fig. 2. In particular, these figures illustrate an evolution from the traditional layered layout of core nodes based on IP, ATM, SONET/SDH and WDM to the more flat ''IP focused architecture'' with only IP over optical transport network (OTN) over WDM directly [2]–[4]. A further evolutional step to an ''Optical bypass architecture'' shown in Fig. 1 illustrates a range of options for each transmission technology to different technology options.

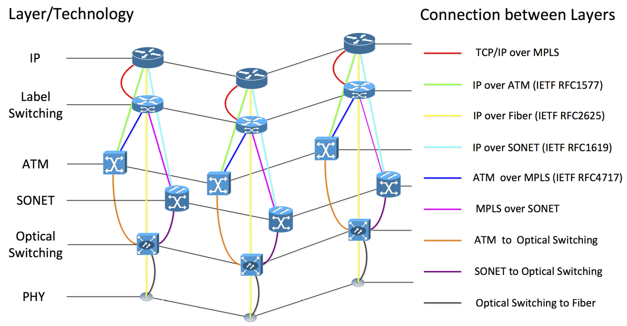

FIGURE 1. Illustration of connectivity in a layered communication network.

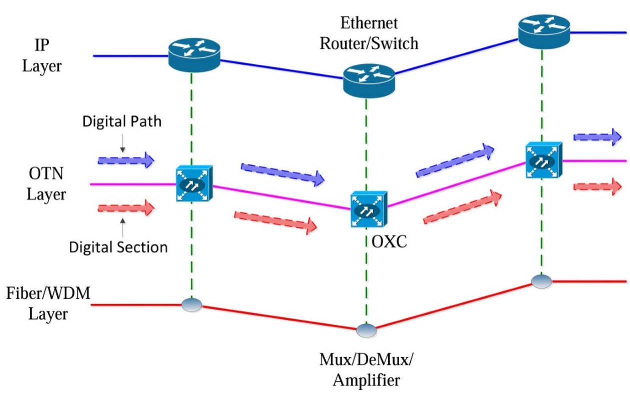

FIGURE 2. OTN based layered network architecture.

From all these developments, given the benefits of layered architectures, we can expect that layered architectures of the Internet are here to stay, and the problems of how to provision resources in layered architectures will remain important for the foreseeable future. Given the growth, complexity and size of the Internet, new methodologies are needed to support cost-effective resource provisioning for multi-layered networks.

A. OUR RESOURCE PROVISIONING ALGORITHM

In this paper, we describe a polynomial-time algorithm for core network resource provisioning, which we call multi-layered market algorithm (MMA). MMA considers a layered architecture of multi-technologies network, and the traffic is modelled by either constant bit-rate (CBR) traffic streams or variable bit-rate (VBR) traffic streams, where the VBR traffic is modelled by a combination (or mixture) of long and short range dependent processes.

Since every network resource is used by several traffic streams, MMA also considers different sharing properties of different switching technologies to allow these traffic streams to share the capacity and the cost of the resources they use. The different sharing schemes will be further discussed in detail in Section II-B. The cost of transmission of the data considering the particular sharing schemes associated with each technology used in each layer as key parameters of the algorithm. MMA aims to optimize the routing choices for each traffic stream between source destination (SD) pairs, and resource provisioning is based on given cost models that considers different traffic sharing allocation schemes in various layers. The solution derived by MMA will achieve cost efficiency of transporting a certain amount of data between SD pairs under the availability of sets of routes and transmission technologies.

Although MMA aims to optimize traffic management including the choice of layers and routing for every stream for the purpose of resource provisioning, it does not provide means to implement such optimization in real time in a real network. This will potentially be provided by currently evolving technologies including software-defined networking (SDN) [5], [6]. Nevertheless, MMA provides scalable resource provisioning that is based on efficient nearly optimal traffic management. Such provisioning is needed given current network evolutionary trends.

MMA is run on a publicly available platform called Netml (network mark-up language). It is currently available at <netml.usq.edu.au>, and we plan to make it available on netml.org [7] during 2019. MMA is validated in three ways. Firstly, we compare the cost-effectiveness of a range of MMA variants to equivalent integer linear programming (ILP) variants for small networks. By careful choice of the various ILP benchmarks, we are able to demonstrate that the polynomial-time MMA indeed achieves efficient design for the complex problem of multi-layered network optimization under realistic traffic models. Secondly, to guarantee the correctness of MMA, we provide a methodological procedure for software verification. As in [8] and [9], we use the term *module* to represent a ''modular capacity unit'' because ''in real networks, capacities can be installed only in modular units, e.g., T1, E1, OC-3'' [8]. In this paper, we allocated capacity to a link based on an integer number of modules per link multiple the module bit rate.

We provide a detailed report which validates the prices attributed to the traffic streams which must equivalent to the total cost based on the cost of the modules in each link [10]. This verification process is equivalent to double-entry bookkeeping (widely used for error detection in the accounting field for centuries) [11]. The third way is based on visualization of the optimized result network where all the details of the network design can also be observed.

With the ever-increasing growth of the Internet, its adverse effect on the environment in terms of its contribution to global warming also increases [12]. The combination of potential energy tax and social responsibility of network providers motivate the design of greener networks. MMA can be used for the design of such networks by amplifying energy costs in the optimization. This will allow weighting in favor of ''greener'' solutions, and the solution by our algorithm will potentially contribute to a cost efficient as well as an energy efficient network.

B. RELATED WORK

Multi-layered network design of traffic routing and capacity assignment assuming deterministic traffic in a two-layer or three-layer network has been addressed previously in various publications (e.g. [2], [8], [9], [13]–[31]). A comprehensive survey of approaches and methods for multi-layered network optimization and design is given in [20]. Heuristics based on different routing algorithms for design of multi-layer and multi-technologies network are discussed in [9] and [13]–[15]. The research in [2], [8], [21], [22], and [24], tackles optimization of CAPEX or the number of network resources for a multilayer network by selection of network components and determination of corresponding prices as specified in the cost models to achieve cost-effective architectures.

Koster *et al.* [27] proposed a single-plane cut method to tackle the minimum cost problem in two-layer SDH/WDM telecommunication network designing. The work in [21] shows that to minimize the total network cost, traffic in the IP layer is groomed for low loaded network and high loads of transit traffic are relocated into the optical layer. A classical approach is based on the use of iterations that alternate between optimizing traffic routing for solving multi-commodity problems and resource provisioning, for example, using dynamic programming [31]. Aparicio-Pardo *et al.* [29] proposed a heuristic MILP model which is based on multi-commodity flows for routing optimization and CAPEX for IP/MPLS over WDM network according to time-varies traffic patterns. Lopez *et al.* [17] proposed a use case for application-awareness in network planning, where resources provisioning in multilayer optimization is dynamically performed.

Traffic traces show that Internet traffic exhibits long-range dependence and self-similarity behavior (e.g., [32]), which is mainly because of high variability in the sizes of Internet flows (or their bandwidth demands) that can be characterized by heavy tailed distributions [33]–[37]. It is well known that it has been considered difficult to obtain exact analytical results for performance of systems fed by long-range dependent processes for the entire range of the parameter values. A popular traffic model is the so-called Poisson Pareto burst process (PPBP) (a.k.a M/G/∞ process for the special case Pareto distribution is used for the 'G'). This process consists of a Poisson process of arriving flows which are assumed to have independent Pareto distributed lengths. The Pareto distribution is heavy tailed for certain choices of parameter values. The assumption of Poisson arrivals of flows can be justified considering the very large number of independent users that generate these flows. In addition, the assumption of the flow sizes being heavy tail distributed is supported by the significant data and research in this area [33]–[36]. Existing studies of performance evaluation of PPBP queues include asymptotic analysis [38], [39] and approximations that are applicable to the full range of parameter values [40]–[42]. However, no exact analytical results are available for such queues.

Whitt [43] extended the earlier work by Kuehn [44], and provided an analysis of a queueing network assuming of renewal arrivals to each individual queue in the network. The analysis in both [43] and [44] was based on the decomposition of the network into individual queues, and approximating the statistical characteristics of each individual queue. This approach, described also in [45], is called *parametric decomposition method*. It led to a software design tool of AT&T called queueing network analyser (QNA). While QNA generalized existing solutions that were based on pure chance traffic (Poisson processes) by considering renewal processes, MMA further extends QNA by considering even more realistic traffic that comprises CBR, PPBP and Gaussian traffic streams.

Another important Internet traffic model is the Gaussian process [46], [47]. In particular, we use the short range dependent Ornstein-Uhlenbeck process as a traffic model [48], [49]. The Ornstein-Uhlenbeck model was proposed in [46] as a model for a superposition of a large number of identical, independent processes. An approximation for the bandwidth requirement of multiplexed Gaussian distribution bit rate streams is given in [50], which is equal to mean aggregate bit rate plus several standard deviations of the aggregate bit rate, where the number of standard deviations depends on the grade of service (GoS).

In addressing comprehensively the optimization of networks, it is common to consider the problem of developing a mathematical formulation that captures all the key realistic features of networks that includes all its layers and potential technology alternatives. This includes understanding of how network technologies impact the network design of resource provisioning and network operation by considering how the traffic demand variability affects routing. For this purpose, we need algorithms that provide optimal or near to optimal designs for a realistically large-scale network.

Most of the optimization problems associated with realistic size networks cannot be solved in a practically sufficient short time with existing algorithms and available software tools [8]. When we consider all the aspects, it becomes harder to find a comprehensive solution for all relevant cases. As far as we are aware, there is no existing algorithm that provides optimal solution for a realistic size multi-layered network with both CBR and VBR traffic streams which can be used in a design tool to achieve realistic and cost-effective solutions in polynomial time. Such a tool is needed for future network design.

We have published several conference papers that described earlier versions of this work in [7], [10], and [51]–[54], and in the PhD thesis of the third author [55]. We have developed a web-based online optimization tool platform called Netml [7]. The Netml system includes features that allow us to visualize capacity assignment as well as utilization for each physical and virtual links and provides data for all analyses of MMA. In [51], we introduced the concept of flow-size dependent routing in multi-layered network optimization. Next, in [52], we also clarified that MMA can also help network designers decide the choice of technologies for future network in addition to resource provisioning. In [53], we further expanded the analysis by considering different link types, which are, simple, virtual permanent and dynamic. These links are dynamically generated by MMA to achieve minimum cost. The general idea of MMA was firstly illustrated in [10] and one example network with only PPBP traffic was validated in [54]. More information of previous work can be found in [55].

C. NOVELTY AND CONTRIBUTION

This paper includes a range of MMA related important novel contributions that have not been published earlier. A list of the novel contributions of this paper is provided as follows:

1) Although resource sharing was briefly mentioned in [10] and [55], it was not implemented there as part of MMA. In this paper, for the first time, we report implementation, within MMA, of different traffic sharing allocation schemes of the various transport technologies in the different layers, and cost optimizations by two given cost models between different combinations of these schemes. The inclusion of these sharing schemes in a multi-layered network optimization can significantly affect network costs, leading to more efficient resource provisioning; this is a novel contribution in this paper.
2) For the first time, this paper provides a demonstration based on extensive empirical results that MMA is solvable in polynomial time which makes it applicable to realistic size networks.
3) None of the earlier publications included the formulation of non-linear multi-layered network optimization provided in Section II-C. This formulation rigorously defines the problem solved by MMA.
4) Although our earlier publications on MMA [10], [52], [54] considered CBR and PPBP traffic streams, none of them considered Gaussian traffic streams. The inclusion of Gaussian traffic in MMA further generalizes its traffic modeling and therefore enhances the applicability of MMA to real-life networks.
5) The comprehensive results for MMA validation based on ILP benchmarking presented in this paper are mostly new. We note that ILP formulation presented here was provided in PhD thesis [55], and was used in certain variants for validation of MMA, and some numerical results based on these variants were also published in [54]. In this paper, we consider a wider range of variants that further enhance our understanding of the quality of results achieved by MMA. Also new is the use of visualization for extra validation of MMA results applicable to larger networks. In fact, all the numerical results presented in Section V, are new, and are based on thoroughly verified software using double-entry bookkeeping.

D. ORGANIZATION

The remainder of this paper is organized as follows. In Section II, we provide network and cost models and a formal description of the multi-layered network optimization problem considering CBR and VBR (Gaussian and PPBP) traffic. Section III provides a detailed description of MMA. In Section IV, we present an ILP formulation for a multi-layered network optimization problem considering only CBR traffic. Section V presents extensive numerical results for MMA validation by a range of ILP variants, for evaluation of MMA running time, as well as results that demonstrate MMA scalability, the effect of different sharing schemes, and flow-size dependent routing. Finally, Section VI concludes the whole paper.

II. NETWORK MODEL AND PROBLEM DESCRIPTION

We consider an N-node multi-layered communication network, in which traffic streams between each SD pair are modelled as CBR or VBR, and the VBR traffic is modelled by a combination (or mixture) of PPBPs and Gaussian processes. The size of modules in each layer, unit costs of each module, and physical topology of the network are given. Different transport technologies with different traffic sharing characteristics are available in the various layers. We aim to find the optimal resource provisioning considering flow-size dependent routing that minimizes the total network cost with different models of traffic subject to meeting GoS requirements.

The term *traffic stream* in this paper is defined as a statistical aggregation of flows between an SD pair, and its statistical characteristics can be parameterized by the combined set of parameters of its Gaussian and PPBP components. The term *flow* represents traffic generated between an SD pair as a result of, for example, a click on a website, making a phone call or sending an email. Under our PPBP traffic stream model, we assume that the sizes of such flows are Pareto distributed. For Gaussian process or PPBP, the terms *traffic stream* and *demand* are used interchangeably. A traffic stream/demand represents a collection of flows between an SD pair, and is defined by its SD pair and its statistical characteristics. A CBR traffic stream is a permanent demand such as a leased line which stays constant for its entire lifetime. A CBR flow places a constant load, in bits/s, during the life of the flow. The flows within a PPBP traffic stream are all CBR.

A. PRELIMINARIES AND NOTATION

We adopt the framework of [8] which provides a sufficiently general framework for multi-layered network design optimization. The number of layers is L, that are labeled $1, 2, 3, \ldots, L$, where Layer 1 is the bottom physical transmission layer and L is the top layer. Layer $L+1$ is supposed to be a demand layer that contains all statistical information of traffic streams between each SD pair. Layers $2, 3, \ldots L+1$, represent virtual topologies. The problem formulation does not limit the number of layers, however, the number of layers must be fixed throughout each experiment.

In this paper, as discussed, in addition to CBR traffic streams, we also assume VBR traffic streams. Although CBR traffic streams, e.g., for private networks, are important, most customers generate VBR traffic streams, so network providers are obliged to make use of statistical multiplexing to save network resources. This highlights the importance of considering VBR traffic streams, as part of resource provisioning of multi-layered network optimization.

We note that, since VBR traffic streams are made up of Pareto distributed flow sizes, the network design algorithms we consider, allow the route for a flow to vary depending on its size – this is called *flow-size dependent routing* [52]. In general, K types of flows are defined according to their sizes, denoted by their *size indicators* $k = 1, 2, \ldots K$. For example, $K = 2$ means that two types of flows are considered, namely, mice and elephants, which had been considered in various publications (e.g. [56], [57]). In PPBP, the random variable d representing the flow-size follows Pareto distribution given as:

$$P\left(d > x\right) = \begin{cases} \left(\frac{x}{\delta}\right)^{-\gamma}, & \textit{if } x \geq \delta, \\ 1, & \textit{otherwise}, \end{cases}$$

where $\delta > 0$ is the scale parameter which is the lower bound for the random variable d, and $\gamma > 0$ is the shape parameter of the Pareto distribution of the flow size d. For $1 < \gamma < 2$, we have the mean $\mu(d) = \frac{\gamma\delta}{\gamma-1}$ and the variance is infinite. Since routing may cause traffic flows to be split according to the size of their flows, we also sometimes need to use the truncated Pareto distribution to describe the flow size distribution of traffic streams. As discussed in [51], the mean and variance of truncated Pareto distribution d with shape parameter γ, minimum flow-size of truncated Pareto flows δ and maximum flow-size of truncated Pareto flows Δ is given as:

$$\mu(d) = \frac{\gamma\left(\delta^{1-\gamma} - \Delta^{1-\gamma}\right)}{(\gamma-1)\left(\delta^{-\gamma} - \Delta^{-\gamma}\right)},$$

$$\sigma^2(d) = \frac{\gamma\delta\left(\delta^{2-\gamma} - \Delta^{2-\gamma}\right)}{(\gamma-2)\left(\delta^{1-\gamma} - \Delta^{1-\gamma}\right)}.$$

Traffic with an un-truncated Pareto distribution contains, in general, both mice and elephants, whereas traffic with a truncated Pareto distribution might contain only mice if Δ is small. Also, if the lower limit δ of the Pareto distribution is large, a traffic stream might contain only elephants.

B. NETWORK AND COST MODEL

In this section, we describe in detail our network model and how costs are assigned to its components. Our network model aims to be as simple as possible while retaining the capacity to include all the key aspects of real network structures that are either directly relevant to cost, or to operations that affect both performance and cost.

a: Layered Link Management

Naturally, the model includes *nodes* and *links*. Both of which have a capacity, that is measured in either bits/sec, or packets/sec, which are inter-convertible by the average packet size. Two four-node networks are shown in Figs. 3 and 4. All the nodes have the same structure, and include capability to switch at all three layers of the network. However, links must occur in Layer l whenever they occur in Layer $l-1$, but not the other way around. We consider a graph of nodes and links in the network that may consist of the following types of links.

1) *Simple links* A simple link in Layer l is formed between two nodes by providing the transport service between these two nodes in Layer l and using a path between these two nodes at Layer $l-1$. Accordingly, a simple

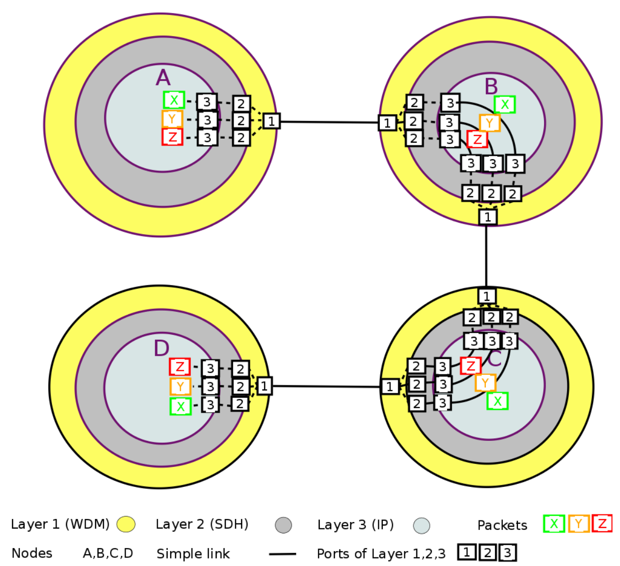

FIGURE 3. IP over SDH over WDM.

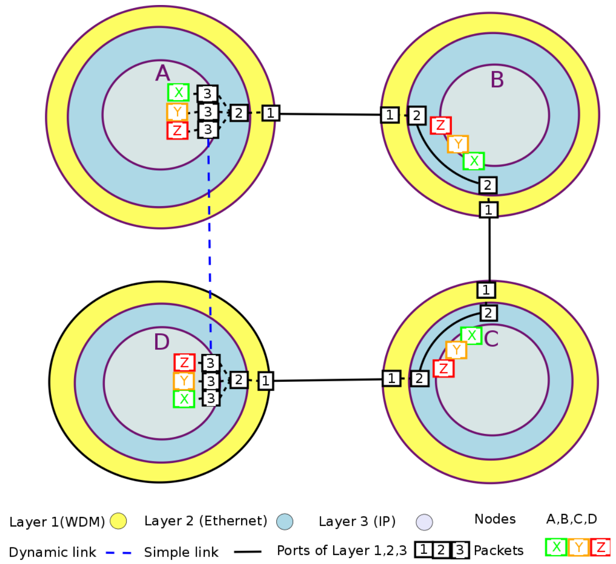

FIGURE 4. IP over Ethernet over WDM.

link in Layer L will use a path in Layer $L-1$ and then each link in that path will be a simple link that uses a path in Layer $L-2$, etc. until a path in the physical layer is used.

2) *Dynamic links* are formed between two nodes by setting up a path between these two nodes in the layer below. The choice of the path is based on the size of flows. In the layers where dynamic links are provided, these links are assumed to be universally available between all the pairs of nodes. However, we also assume that a flow set up cost is incurred for setting up a dynamic link.
3) *Virtual permanent links* are also formed between two nodes by setting up a path between these two nodes which is permanently active, hence, there is no setup cost for each flow. Virtual permanent links, however, incur an additional *permanent* cost due to the requirement for entries in MPLS tables, for example. This cost is assumed to be the same for each virtual permanent link, irrespective of any of its characteristics, in particular its rate, although it might be different in each layer.

b: Traffic Sharing Over Layers

In some cases, switching technologies have modular capacity, nevertheless, they allow the carried traffic streams to share all the capacity of allocated modules. For example, because Ethernet is a packet-switching technology, it allows all the IP traffic being carried to share the total capacity of each Ethernet link. The same is true of ATM, but it is not true for an SDH layer because although SDH has switches, they switch synchronous channels rather than packets.

As sharing schemes are applied, since each network resource is commonly shared by multiple traffic streams, these streams must share the cost of the resources that they use. Therefore, a clear plan is needed to subdivide costs in a cost model. Assigning costs for modules to streams, or to be more precise, to individual packets that make up a flow, is challenging because modules may be shared in several ways by the streams. In MMA [7], the different schemes of sharing of resources, associated with each layer of a layered network are as follows (see also in [10]):

1) *Non-traffic-sharing allocation scheme* In MMA, the most basic allocation scheme is termed, *Non-traffic-sharing*. This scheme is an option for any layer. The resource is only used by the specific stream assigned to the module. Since the unused module of the links cannot be further used, the total cost of the module should be borne by the stream that uses it. This is the appropriate sharing scheme to use for *SDH*.
2) *Traffic-sharing allocation scheme* In this situation, traffic can be routed on a link without having exclusive access to the network resource. This is called *Traffic-sharing*. For example, the unused resource of a link could be shared by any traffic and not necessarily traffic between the two end-nodes. If a link has been allocated in multiple modules and the streams within these modules do not fully utilize them, the unused capacity of these modules is preferably shared by other streams. This is the appropriate sharing scheme to use in *IP, Ethernet*.

An important technology which is also relevant to this paper is the OTN architecture as shown in Fig. 2. One important advantage of OTN over ATM/SDH/WDM, is that it allows using traffic-sharing. In particular, the new transmission architecture, based on carrier Ethernet with capacities of 1 Gb/s and 10 Gb/s, and in future 100 Gb/s, is capable of traffic-sharing. SDH, on the other hand, normally is not capable of traffic-sharing because it has to be carried in SDH switches which are based on a TDM architecture. This helps to explain why the OTN architecture provides transmission capacity more cost-effectively, because, in OTN, the capacity in the link is not dedicated, so unutilized capacity can be further used for other traffic. Ethernet is also a technology that has this property. Hence, such technologies in various layers should be modelled as traffic-sharing.

c: Cost Model

In addition to nodes and links, the model includes the concept of *ports*, which are required at the physical or virtual interface of links. A link in Layer L requires a Layer L port. As shown in Fig. 3, at a terminal node (Node D for example) of a link in layer L, there must be ports for layers 1 up to L. Fig. 4 shows an example of the ports used by each link in each layer when a dynamic link or a virtual permanent link is used between the SD nodes at layer L. At intermediate nodes (e.g. Node B and Node C), a link might require only ports in layers 1 to $k < L$. In a node, packets are transferred through links between layers and switched from input port to output port by switching fabrics in the node. The packets can only be transferred to another node by a physical link in Layer 1 between these two nodes.

Consider an example of setting up dynamic link of an Ethernet path through a network under traffic-sharing scheme. The setup process is mainly handled by the IP router which is in the layer above, not by the Ethernet switches. When dynamic links are used, additional costs which are expressed as overheads are introduced: *flow set-up cost* which is expressed as a percentage overhead relative to the layer and assigned to this layer when costs are added up.

The flow set-up cost is born in two layers, the layer where the links of the path exist, and the layer above, in which the path appears as a dynamic link. For simplicity, in the experiments, we ignore the cost incurred in the lower of these two layers. When flow set-up cost is 0, dynamic links are always cheaper than the two hops of static simple links, because the switching cost in the middle hop is not incurred, and there are no costs in using dynamic links.

Costs are assigned as *module costs*, which for each layer, represent the cost of the ports which implement the particular technology, i.e. the cost of the equipment resources required for this layer. The cost of a network is determined by the capacity of the ports associated with each link or node. The capacity of ports is determined by the traffic streams using this link or node and the technology (traffic sharing scheme) used in this layer. Generally, capacities of links are allocated in discrete modules. For example, in multiples of 1 Gbit/s, multiples of 10 Gbit/s, etc. This sometimes causes a waste of capacity (due to an underutilized module) that propagates to lower layers. This is due to the fact that, usually, the entire capacity (all the modules) of Layer l must be carried by Layer $l-1$. This will usually be larger than the aggregated traffic of Layer l. If only a small proportion of a large module is used, on a certain link, the cost of using this link will be high, and the design algorithm will tend to reduce the use of such a link, possibly even to the point where it is not used at all.

In MMA, the different cost models of resources, associated with each layer of a layered network are as follows:

1) *Port-accounting cost model* In this model, capacity is associated with ports of links. For example, in IP layer, each modular cost includes CAPEX and OPEX of routing traffic streams in the link using that module. The way we model the cost of resources is based on capacity of links in the network. We allow the traffic streams to share/not share the resources provided by *links*. This is termed *Port-accounting*; it is appropriate for use in all the layers.
2) *Node-accounting cost model* In this model, capacity is not associated with links but calculated in nodes at top layer. The cost of node resources is determined by a weighted sum of the originating, transit, and terminating traffic at the node. This cost model is only appropriate to the top layer, e.g. the IP layer. This is termed *Node-accounting*.

Generally, the design of multi-layered network involves more than two layers with different technologies. Therefore, the layers can apply different combinations of traffic sharing schemes. Different sharing scheme could be used in different technologies and may involve different costs. Let us consider an example in a 3-layer network combined with different schemes of sharing, with and without dynamic links (at the top layer). A simple network and traffic model for use here is the *non-traffic-sharing scheme with port-accounting* for each layer. In this model, the original demands determine the capacity of links at top layer and then the capacity of the links are served in turn as the demands to be routed in the second layer. Thus, we need to determine the capacity of links in the second layer based on routing in this layer which in turn served as the demands to be routed in the bottom layer.

Node-accounting is useful if we include dimensioning of IP routers. Node-accounting is important if we wish to model and estimate the cost saving due to the use of dynamic links. In real networks, dynamic links will only be used because the top-layer switching (routing) cost can be avoided by using them. We model the cost of nodes by using node-accounting. When node-accounting is used, the cost of that layer is based on the choice of nodes, not on the links in that layer. The links, in fact, appear to have no cost. Or to be more precise, the cost of the links is passed on to the nodes, and to the layers below. When *traffic-sharing scheme with node-accounting* is used, the setting up of dynamic links at Layer L, which uses the resources in the layer below, can be interpreted as making a choice of the preferred technology, for example, bypassing a certain switching technology which is available at Layer L. Therefore, when MMA chooses a certain route, this can be viewed as the choice of a preferred technology.

d: Examples

Fig. 3 shows a network in which IP, SDH and WDM technologies are used in Layers 3, 2, and 1, respectively. Three traffic streams from different IP ports (which are in different colors) are transmitted through SDH ports and further aggregated at WDM layer in source Node A. The aggregated traffic is transported by physical links to destination Node D via intermediate Node B and Node C. The ports of Node B and Node C switch the traffic streams from an input port to an output port using a switching capability in the IP layer. When Ethernet is applied at Layer 2 instead of SDH as shown in Fig. 4, in source Node A, the sharing of capacity happens in both Layer 1 and 2. Three traffic streams from different IP ports aggregated at Ethernet ports at layer 2 and further share its capacity in WDM at Layer 1. The aggregated traffic then is transported by physical links to destination Node D going through intermediate Node B and Node C. Since two dynamic links are set up in Layer 3 by adding a path between these two nodes at Layer 2, a link might require only ports in Layers 1 and 2 at intermediate nodes. In Fig. 4, the traffic streams only use the ports of Layers 1 and 2 in the intermediate Node B.

We also provide a diagram directly from the Netml system which displays networks designed by MMA under different settings. There are three PPBP traffic streams from Node A to Node D with same means and standard deviations in a four-node network. Three layers with different

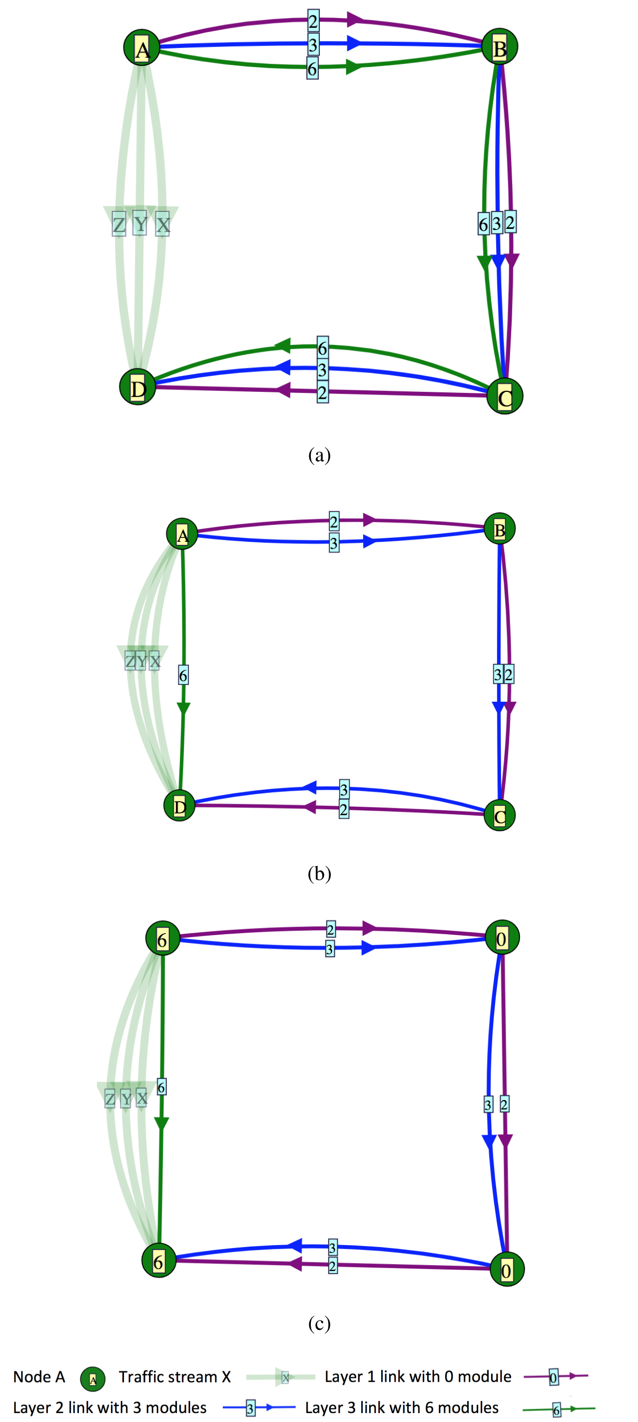

FIGURE 5. **MMA Solutions with different sharing schemes viewed in Netml. (a) Non-traffic-sharing scheme with port-accounting. (b) Traffic-sharing scheme with port-accounting. (c) Traffic-sharing scheme with node-accounting.**

module sizes for each layer are pre-defined in the settings. The Figs. 5(a)-(c) describe the resource provisioning for each link (node) with different sharing schemes in the Layer 3. The settings in other layers are the same.

The colors of links represent the generated links in different layers and the number on the link represents the number of modules used on that link (node) which imply the costs of resources. As shown in Fig. 5(a), the resource is allocated to links at to all the hops in each layer where the traffic travels which is the same principle as Fig. 3. When the dynamic link is enabled and traffic-sharing scheme is used, the results are shown in Figs. 5(b)-(c). One dynamic link is generated by MMA between Node A and Node D which saves the switching cost of Layer 3. In the case of traffic-sharing scheme with port-accounting cost model as Fig. 5(b), the total cost of Layer 3 is the resource used by the dynamic link which is 6 units and this saves up 12 units compared to non-traffic sharing scheme. This scheme is the same principle as illustrated in Fig. 4. In the case of node-accounting cost model as Fig. 5(c), the cost of the top layer is calculated in nodes which the traffic originating or terminating. Therefore, there is no cost involved in Node B and Node C which saves up to 6 modules compared to Fig. 5(a).

e: Discussion

Defining network and cost models which adequately explain the functionality and cost of modern networking technology, and developing a network design tool which full reflects these costs, can help us to explore different combinations of alternative technologies, (e.g. IP, SDH, Ethernet/OTN, WDM) which may be operating in one network. To compare the different combinations of technologies in our experiments, we specify the technologies for use in different layers and allow the tool to optimize network cost. When investigating whether to choose one technology or another, the design tool must be run multiple times: once for each alternative combination of technologies considered. When comparing technologies, the presence or absence of traffic sharing is expected to be a feature which features prominently. This will help service providers compare between multiple competing technologies.

C. PROBLEM FORMULATION

The general network design problem which was previously described informally will now be set out more formally. The goal is to determine C_e^l which represents the capacity of each link in each layer which takes a discrete value representing multiples of modules, which depends on the layer.

Two distinct types of dimensioning algorithm may be selected, one type suited to access networks, and the other suited to core networks. When access type dimensioning is selected, the maximum utilization of each link is fixed. In this case, therefore, the mean of the aggregate traffic on a link is used to determine its minimum required capacity. When the core type algorithm is selected, the minimum required capacity is determined to suit the target GoS and for this purpose, the mean and standard deviation of the traffic on a link is taken into account. In this paper, it is assumed that the core type dimensioning algorithm is selected on all links.

The function $Q_e^{\{l\}}(\cdot, \cdot)$ in the following is the GoS function which guarantees the minimum capacity requirement

TABLE 1. Table of notation.

Symbol	Description
L	Number of layers
$C(l)$	Cost of Layer l, $l = 1, \ldots, L$
$E^{\{l\}}$	Total number of links in Layer l
$c_e^{\{l\}}$	Capacity of link e
$M_{e,t}^{\{l\}}$	Mean of traffic stream t on the link e
$T_t^{\{l\}}$	Total number of traffic streams
$S_{e,t}^{\{l\}}$	Standard deviation of traffic stream t on the link e
$\lambda_{e,k}^{\{l\}}$	The arrival rate of PPBP traffic k carried on the link e
$r_{e,k}^{\{l\}}$	The bit rate of PPBP traffic k carried on link e
$\delta^{\{l\}}$	Lower bound of the size of truncated Pareto flows
$\Delta^{\{l\}}$	Upper bound of the size of truncated Pareto flows
$\gamma^{\{l\}}$	Shape parameter of truncated Pareto flows
$\mu(\gamma^{\{l\}}, \delta^{\{l\}}, \Delta^{\{l\}})$	The mean of the truncated Pareto traffic k as the equation in II-A
$\xi_e^{\{l\}}$	Modular cost of capacity on the link e
$f^{\{l\}}$	Flow set up cost for dynamic links d
$v_p^{\{l\}}$	Virtual permanent link cost for link p
$Q_e^{\{l\}}(\cdot,\cdot)$	GoS function for the link e
$\mathcal{D}^{\{l\}}$	Total number of dynamic links d
$\mathcal{P}^{\{l\}}$	Total number of virtual permanent link p
The superscript $(\cdot)^{\{l\}}$ uniformly denotes the layer	

for acceptable blocking probability. When the dynamic links and virtual permanent link are not considered, the objective function and constraints are quite simple because in such case we only consider costs associated with the mean and the standard deviation of the bit-rate of traffic t on any link e given by $M_{e,t}^{l}$ and $S_{e,t}^{l}$. Our objective is to minimize $C(l)$, which is the total cost of Layer l as given by:

$$C(l) = \sum_{\mathrm{e}=1}^{E^{\{l\}}} \left(\xi_{\mathrm{e}}^{\{l\}} c_{\mathrm{e}}^{\{l\}}\right),$$

subject to, for each Layer, $l = 1, \ldots, L$, where transmission capacity must be allocated, and each link e, in Layer l:

$$Q_e^{\{l\}}\left(\sum_{t=1}^{T(t)} M_{e,t}^{\{l\}}, \sum_{t=1}^{T(t)} S_{e,t}^{\{l\}^2}\right) \leq c_e^{\{l\}}.$$

As discussed, traffic streams between their sources and destinations follow either Gaussian or CBR processes, or PPBPs. Note that a CBR traffic stream can be viewed as a Gaussian stream with zero variance. When a traffic stream is modelled as Gaussian [58], the capacity of a link e in Layer l is determined by the mean and variance of the traffic on this link subject to GoS requirements which forced by the $Q_e^{\{l\}}(\cdot,\cdot)$ function.

The objective function is more complicated for PPBP traffic streams, since different types of Pareto distributed flow are considered, and we allow a possibility that the traffic stream can be split according to flow-size dependent routing by setting up a dynamic link between every two end nodes. The objective function should be modified to allow a dynamic link in Layer l which involves a flow set-up cost $f^{\{l\}}$ to be identified with a path in a layer *below* if dynamic links are enabled. If virtual permanent link is enabled, the objective function should also consider a virtual permanent link cost which is assumed to be the same for each virtual permanent link. Our final objective is given by:

$$C(l) = \sum_{\mathrm{e}=1}^{E^{\{l\}}} \xi_{\mathrm{e}}^{\{l\}} c_{\mathrm{e}}^{\{l\}} + \sum_{\mathrm{d}=1}^{D^{\{l\}}} f_d^{\{l\}} c_{\mathrm{d}}^{\{l\}} + \sum_{\mathrm{p}=1}^{P^{\{l\}}} v_{\mathrm{p}}^{\{l\}},$$

where for each layer, $l = 1, \ldots, L$, and each link, in Layer l are subject to the previous GoS function.

Recall that the PPBP traffic streams are split by flow-size dependent routing. Therefore, when considering both Gaussian and PPBP traffic, in this case, we define traffic t represents a Gaussian traffic stream and traffic k represents a PPBP traffic stream Suppose a Poisson stream of truncated Pareto flow k on link e has arrival rate $\lambda_{\mathrm{e},k}$, traffic bit rate $r_{e,k}^{\{l\}}$ with a range of flow sizes from $\delta^{\{l\}}$ to $\Delta^{\{l\}}$. The mean of traffic k is given by $M_{e,k}^{\{l\}}$ which denotes the mean bit-rate of the PPBP traffic k transported on link e through layer l with flow-size smaller than $\Delta^{\{l\}}$ and larger than $\delta^{\{l\}}$. According to [51]:

$$M_{e,k}^{\{l\}} = \lambda_{\mathrm{e},k}^{\{l\}} r_{e,k}^{\{l\}} \mu(\gamma^{\{l\}}, \delta^{\{l\}}, \Delta^{\{l\}}),$$

The mean bit-rate of PPBP traffic k transported on link e through layer l with un-truncated Pareto flow-size larger than $\delta^{\{l\}}$ is [40]:

$$M_{e,k}^{\{l\}} = \frac{\lambda_{\mathrm{e},k}^{\{l\}} r_{e,k}^{\{l\}} \delta^{\{l\}} \gamma^{\{l\}}}{(\gamma^{\{l\}} - 1)},$$

The total mean bit rate of the link e on Layer l is the summation of mean of all the traffic on this link, i. e.

$$M_e^{\{l\}} = \sum_{(k,t)=(1,1)}^{T(t),K(k)} (M_{e,t}^{\{l\}} + M_{e,k}^{\{l\}}).$$

The PPBP traffic variance $S_e^{\{l\}^2}$ on link e through Layer l is given by the variance of a Poisson distribution with mean given by the number of active flows, as explained in Section II-A and in [51].

III. THE MULTI-LAYERED MARKET ALGORITHM

MMA is an iterative algorithm, where at each iteration, traffic is routed in each layer, and links are dimensioned to carry it. Layer $L+1$ represents the demands from customers, and each Layer l, $l = 1 \ldots L$ can be viewed as a business that must meet the demand itself, or outsource it by using a dynamic or virtual permanent link associated with a path in the layer below. In each layer, MMA iteratively assigns link capacities to satisfy the GoS constraints which is defined by $Q(\cdot,\cdot)$ function in Section II-C. The assigned capacity for each link in each layer is based on routing, and results in a set of costs of links which are calculated at each layer based on the number of modules of the links. At the end of each iteration, the cost of using each link is determined from the routing and link capacity assignment which is current at that time. The process repeats itself until there are insignificant differences between successive iterations. It should be mentioned that a certain number of dynamic or virtual permanent links are applied in routing to save total network cost, while unused links are deleted if they are unused for a sequence of iterations. Pseudo-code for MMA is shown in Algorithm 1 and the terminology is in Table 2.

Algorithm 1 Pseudo Code for MMA

```
traffic stream id[x] ← 0
flowsize[i] ← d[i]
while it < minIts| (it < maxIts) & (c ≥ Δ|c′ ≥ Δ) do
  for l = L, ..., 1 do
    Set up simple links in Layer[l]
    according to the physical topology;
  end for
  for l = L, ..., 2 do
    if d[i] ≠ 0 then
      Set up dynamic and virtual
      permanent links in Layer[l];
      Split streams based on their
      size;
      Route split streams by Floyd's
      algorithm in Layer[l] based on
      traffic sharing scheme;
    end if
    Merge demands from Layer[l] and the
    split streams in Layer[l − 1];
    Calculate capacity in Layer[l − 1];
    if x < Max then
      traffic stream id[x]++
    end if
    Remove unused simple, dynamic and
    virtual permanent links;
    Calculate costs based on cost
    models;
    Update costs;
  end for
end while
```

TABLE 2. Terminology for Algorithm 1.

term	Definition
$id[x]$	Traffic stream ID number x
it	the Number of current iteration
minIts	The minimum number of iterations the algorithm should run
maxIts	The maximum number of iterations the algorithm could run
flowsize $d[i]$	Flow-size type i
L	Total number of the layers
l	Layer number $(0, \ldots, L-1)$
Max	Total number of the traffic streams
c	Difference of cost value from one iteration to next
c'	Value of c at one iteration
Δ	Stopping criterion based on c and c'

IV. ILP FORMULATION

ILP formulations are widely used for modeling multi-layered network optimization problem and the fundamental ideas can be found in the book [8]. The objective function of the problems is to minimize the total network cost in terms of reduction of CAPEX and OPEX. In this section, we provide an ILP formulation for a multi-layered network optimization problem where the objective function is the minimization of network cost in terms of capacity assignment for links. To be specific, in our ILP, every traffic stream has a mean and a variance (or standard deviation) of the bit-rate. There will be multiple simultaneously transmitted streams between each SD pair. The capacity requirement of a traffic stream k from node i to node j is defined as x_{ij}^k. Since ILP is only solvable for deterministic traffic, the capacity required is calculated as the mean bit-rate of the traffic stream plus N_σ times the standard deviation of the traffic bit-rate to guarantee the blocking performance [50]. Different approaches for accommodating VBR traffic in the ILP framework are explored in Section V. The value of N_σ is set according to the GoS requirement. The capacity of each link is calculated as the number of modules multiply by the module capacity, which should not smaller than the summation of the carried capacity requirements of each traffic streams on that link.

The ILP cannot guarantee that an optimal solution is found if the routing choices for traffic streams are limited. If routing choices are not limited, better solutions may be obtained. Traffic streams may be split and can choose a realistic subset of paths instead of being restricted to specific routing choices, e.g., shortest path routing. Each traffic stream may choose the same routing path or split to different routing paths. Multiple traffic streams between the same SD pair are routed independently in the top (Lth) layer. These traffic streams are merged in the lower layers which is similar to traffic grooming which is the strategy which is applied in current networks.

Note that the ILP formulation does not include the complication of VBR traffic that we considered in the previous section, however even without such considerations, the algorithmic complexity of ILP must rule out applications to large networks. The ILP optimal results provide traffic routing as well as capacity assignments which are used as benchmark to validate MMA in the case of small networks.

FORMAL STATEMENT

Indices:

$P_{mn,l}$: a binary number, which donates a link between nodes m and n. $P_{mn,l} = 1$ if link between m and n is connected, which m is original node and n is destination node at layer l, otherwise is equal to 0.

L: number of layers of the network.

l: the lth layer of the network, $l = 1$ is the bottom physical layer, $l = L + 1$ is the demand layer.

C_l: module size of lth layer.

K_l: cost per module of lth layer, where $K_l \gg K_{l-1}$, $l = 0, 2, \ldots, L - 1$.

x_{ij}^{k}: denotes the load of traffic stream k from node i to node j.

T: transmission cost in the physical layer for one module in one unit length.

U: a large integer.

i, j, m, n, t: nodes.

Variables:

$F_{mn,L}^{ij,k}$: an integer variable, which denotes the amount of carried traffic on link (m, n) of Layer L of traffic stream k from node i to node j in Layer L.

$F_{mn,l}^{ij}$: an integer variable, which denotes the amount of carried traffic on link (m, n) of Layer l for the link (i, j) at Layer $(l + 1)$.

$y_{mn,l}$: total amount of traffic on link (m, n) in layer l.

$M_{mn,l}$: an integer variable, which denotes the number of modules that link (m, n) uses in Layer l.

Objective:

$$Minimize: \sum_{l}\sum_{mn} K_l \cdot M_{mn,l} + \sum_{m,n} T \cdot P_{mn,1} \cdot M_{mn,1} \tag{1}$$

The objective function is: minimize total network cost.

Constraints:

$$\sum_{n} F_{in,L}^{ij,k} = \sum_{m} F_{mj,L}^{ij,k} = x_{ij}^{k} \quad \forall i, j, k \tag{2}$$

$$\sum_{m} F_{mt,L}^{ij,k} = \sum_{n} F_{tn,L}^{ij,k} \quad \forall i, j, k, t \neq i, t \neq j \tag{3}$$

$$\sum_{i,j,k} F_{mn,L}^{ij,k} = y_{mn,L} \quad \forall m, n \tag{4}$$

$$\sum_{n} F_{in,l}^{ij} = \sum_{m} F_{mj,l}^{ij} = y_{ij,l+1} \quad \forall l = 1, 2, \ldots, L-1, i, j \tag{5}$$

$$\sum_{m} F_{mt,l}^{ij} = \sum_{n} F_{tn,l}^{ij} \quad \forall l = 1, 2, \ldots, L-1, i, j, t \neq i, t \neq j \tag{6}$$

$$\sum_{i,j} F_{mn,l}^{ij} = y_{mn,l} \quad \forall l = 1, 2, \ldots, L-1, m, n \tag{7}$$

$$y_{mn,l} \leq C_l \cdot M_{mn,l} \quad \forall l = 1, 2, \ldots, L, m, n \tag{8}$$

$$y_{mn,l} \leq U \cdot P_{mn,l} \quad \forall l = 1, 2, \ldots, L, m, n \tag{9}$$

Equations (2)–(4) are flow conservation constraints on the top Layer L, which flow conservations are on each source, transit and destination node. For a specific traffic stream in the Layer L, Equation (2) ensures that outgoing demand of a traffic stream from source node i is equal to the incoming demand of this traffic stream to destination node j. Both of them are equivalent to the demand of this traffic stream assigned to this SD pair. Equation (3) guarantees that for transit nodes, the demand of traffic outgoing from the node and incoming to the node are the same. Equation (4) ensures each link in Layer L that the total traffic load is the sum of the traffic streams carried on this link.

Flow conservation constraints also apply to Equations (5)–(7). It should be noted that for optimal routing, flows can be split among multiple paths on the top layer and the flows are groomed together to share a link. The routing is provided for the merged flows in the lower layer. Equation (8) ensures that the capacity offered by the number of modules multiple by the module size, used by any link in each layer, must be larger than the traffic load on that link in that layer. Equation (9) ensures that only two nodes which a have direct connection between them can have direct traffic between them in that layer.

V. IMPLEMENTATION AND NUMERICAL RESULTS

In this section, we present numerical results to validate MMA and illustrate its capabilities. We first present a validation of MMA using ILP variants in the case of a six-node network. Then, we provide results for application of MMA to larger networks. Next, we demonstrate that MMA has polynomial running time based on results of a variety of sizes of networks. Finally, further verification by double-entry bookkeeping and validation by visualization of MMA are discussed.

A. SIX-NODE NETWORK RESULTS

To solve the ILP optimization for the objective function and constraints described in Section IV, we use a commercial ILP solver – CPLEX [59]. A series of experiments in which ILP and MMA are both applied to the same network and the results compared have been undertaken and the results are reported in this section. The ILP optimization results are used as the benchmark for heuristic algorithm MMA. The network used in all examples has the topology shown in Fig. 6 which has 30 directional SD pairs. We assume there are three layers in all the experiments, and the module sizes for Layers 3, 2, and 1 are assumed to be 0.5 Gb/s, 1 Gb/s and 2 Gb/s, respectively. The modular port cost is normalized as 10, 5 and 1 for Layers 3, 2, and 1, respectively. The cost of physical transmission is 2 for a module (mapping to a port in the physical layer). We also assume that there are multiple traffic streams from different applications between each SD pair. In our experiments, the traffic streams are routed independently at Layer 3 and traffic grooming is applied in the lower layers to share the network resources.

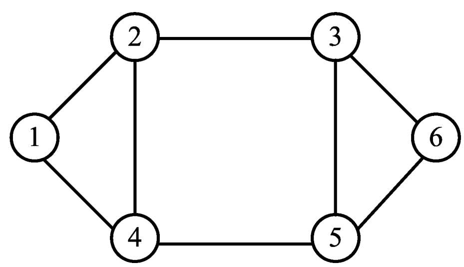

FIGURE 6. A six-node network.

Traffic streams are assumed to be either CBR or VBR. The VBR traffic streams are modelled either as PPBP or as Gaussian. Two statistical characteristics that are relevant to all the traffic streams are the mean μ and standard deviation σ. As mentioned previously, in the case of CBR traffic streams $\sigma = 0$. PPBP traffic streams are characterized by four parameters: Poisson arrival rate λ, the constant rate of flow r, the minimum allowable flow-size δ and the rate of decay of the Pareto tail γ. Equations that provide the mean μ and variance σ^2 and the Hurst parameter H as functions of above four parameters are provided in [40]. The reader is reminded that these traffic streams, in certain cases, may be split and routed over multiple routes. In such a case, for PPBP, such splitting will result in truncated Pareto processes for which we have provided expressions for mean and variance in Section II-C.

Each individual experiment is repeated 10 times with a different randomly selected mean and standard deviation for each traffic stream. For all the numerical results presented here, we assume that the mean of each (VBR and CBR) traffic streams is randomly and uniformly distributed between 10Mb/s and 400 Mb/s, and the standard deviation of its bit-rate is either set to 0, or randomly and uniformly distributed between 40 Mb/s and 120 Mb/s.

In Figs. 7–9, we present numerical results for all the ILP and MMA variants based on the six-node network of Fig. 6. The value plotted in these figures is the average value of total network cost over ten experiments in which the traffic stream means and standard deviations have been varied randomly. The X-axis is the percentage of VBR traffic streams of the total traffic streams where the traffic is modelled by Gaussian or PPBP according to the relevant ILP or MMA variant. The Y-axis is the total cost of the network. To fully explore the benefit of dynamic and virtual permanent links, we suppress one when the other is enabled.

1) TERMINOLOGY

In each experiment, as well as adopting one of the seven different optimization algorithms, the proportion of VBR vs CBR traffic has been varied, and the results are plotted in Figs. 7–9, with the proportion of VBR traffic used as the x-axis. These proportions dictate in each case the proportion of traffic streams for which the standard deviation is set to 0. The capacity allocation for both ILP and MMA is carried out in accordance with the formula $C = \mu + N_\sigma \times \sigma$. The results demonstrated are all based on $N_\sigma = 3$ [60] in this section.

To thoroughly compare ILP and MMA, the optimal cost under three variants of ILP have been compared against the cost obtained by four variants of MMA. The variants of the ILP and MMA network design algorithms are as follows:

1) *ILP with Peak Rate Allocation (ILP + PRA)*:
In this variant, we set, conservatively, for every end-to-end stream, the capacity allocation $C = \mu + 3 \times \sigma$. The routing and capacity allocation required to carry these capacities are then found by using CPLEX to solve the ILP problem, as described in Section IV. The solution found in this way will carry these CBR traffic streams, without loss, from the sources to the destinations. This is somewhat similar to Peak Rate Allocation used in the development of B-ISDN in the 1990s and converted bandwidth requirements guarantee the traffic having good enough blocking performances [60]. This approach has the weakness that it wastes too much capacity because it does not consider the capacity saving benefit of statistical multiplexing of VBR streams. Nevertheless, since the ILP is optimal, it still provides a benchmark.
2) *ILP with Statistical Multiplexing End-to-End (ILP + SM_EtE)*:
This variant partially overcomes the weakness of the first variant by considering the statistical multiplexing benefit of all the streams for each SD pair. In other words, for every SD pair, we consider a traffic stream comprised of all the individual streams between this SD pair, and we calculate the capacity allocation $C = \mu + 3 \times \sigma$, where now the μ is equal to the sum of means of the different traffic streams between this SD pair, and σ is the square root of the sum of the variances of the different traffic streams between this SD pair. Notice that we assume here that the different traffic streams between a given SD pair are independent.
3) *ILP with Statistical Multiplexing (ILP + SM)*:
In the third variant, we use the routing and capacity allocation results obtained from the second variant, and then for each link, we allocate capacity required $C = \mu + 3 \times \sigma$ of the total multiplexed traffic on that link according to GoS.

Four variants that are based on MMA are used for comparison with the above mentioned three ILP variants. The first two of these four variants correspond one-to-one with first two above-mentioned ILP variants. The first two variants involve initially replacing the traffic by CBR traffic with one or other type of peak allocation. This enables us to demonstrate the accuracy of MMA relative to the optimal, but not scalable ILP solutions. In the third and fourth variants, mixtures of VBR and CBR traffic streams are directly supplied to MMA without conversion to CBR traffic by any type of peak-rate or statistical allocation scheme first.

1) *MMA with Peak Rate Allocation (MMA + PRA)*:
This is equivalent to the first ILP variant, namely, *ILP + PRA*, where we set conservatively, for every end-to-end stream, the capacity allocation $C = \mu + 3 \times \sigma$. In this way, all traffic streams become CBR.
2) *MMA with Statistical Multiplexing End- to-End (MMA + SM_EtE)*:
This is equivalent to the second ILP variant, namely, *ILP + SM_EtE*. Traffic between each SD pair is converted to CBR streams with throughput $C = \mu + 3 \times \sigma$, where μ and σ are the mean and standard deviation of the aggregate traffic.
3) *MMA with Statistical Multiplexing (MMA + SM)*:
This is equivalent to ILP with Statistical Multiplexing. The traffic is modelled by Gaussian with two parameter μ and σ and for each link, we allocate the capacity required on that link by the formula $C = \mu + 3 \times \sigma$, according to GoS.

4) *MMA*: The VBR traffic streams are modelled by PPBP parameters, and for each link in each layer, the capacity assignment meets the QoS constraints using the link capacity determined by previous iterations of routing as Section II. The virtual permanent links and dynamic links could be applied.

In Figs. 7 and 8, we present results for cases where the non-traffic-sharing is chosen in Layer 3 in all relevant MMA variants when dynamic or virtual permanent links are disabled. We also present results for cases where the traffic-sharing is chosen in Layer 3 and 2 and dynamic or virtual permanent links are enabled. We consider two approaches in our optimizations. Firstly, the same physical topology, as described in Fig. 6, is used in all layers, and original topology only is used in every layer for ILP. In ILP variants, which use original topology only, since the ILP model does not allow link costs to vary dynamically, dynamic links are not relevant. Hence, in MMA variants which we compare with ILP, in these cases (where routing is always shortest path), dynamic links are not included as an option. In the second approach, all three ILP variants and MMA with dynamic or virtual permanent links choose the fully-meshed topology at the top layer (Layer 3). In particular, in all three ILP variants, a direct link can be selected between all pairs of nodes at Layer 3.

In our experiments, we consider six traffic streams between each SD pair in the network. In MMA + SM and MMA, we vary the proportion of the traffic carried by VBR traffic streams of the total traffic (CBR + VBR). This is done by varying the proportion of the number of SD pairs of the VBR traffic streams of the total number of SD pairs. This enables us to observe the effect of this proportion on the efficiency of the heuristic algorithm (MMA) relative to its ILP benchmark. The total network costs of the ILP and MMA variants are compared under various proportion of VBR traffic streams.

2) RESULTS

Recall that each plotted point in Figs. 7–9 is the averaged outcome from 10 individual experiments, as discussed in Section V-A.1. For each individual experiment, the same network and traffic is used as the problem to solve for each of the seven ILP and MMA variants. Ninety-five percent confidence intervals for the total network cost were calculated, using the Student-t distribution. None of the confidence intervals exceeded 10% of the estimates.

When the same physical topology (see Fig. 6) is used in all layers, the ILP + PRA outperforms the heuristic algorithm MMA + PRA by about 10%. ILP has an advantage over MMA when there are opportunities to route traffic through the spare capacity on links. However, it will be difficult in real networks to use the unutilized capacity in this way by splitting traffic and forcing some of it to use non-shortest path routes.

By using statistical multiplexing for the traffic between the same SD pairs, ILP + SM_EtE can save up to 1/4 of the total network cost compared to ILP + PRA. This may

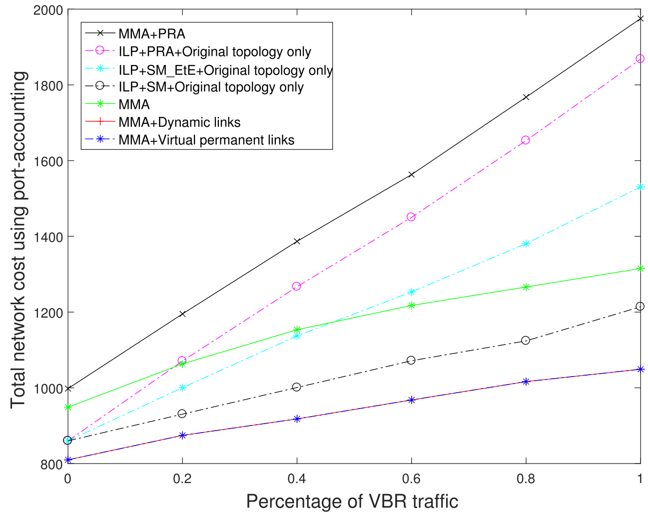

FIGURE 7. Comparison of MMA and ILP with original topology only in Layer 3.

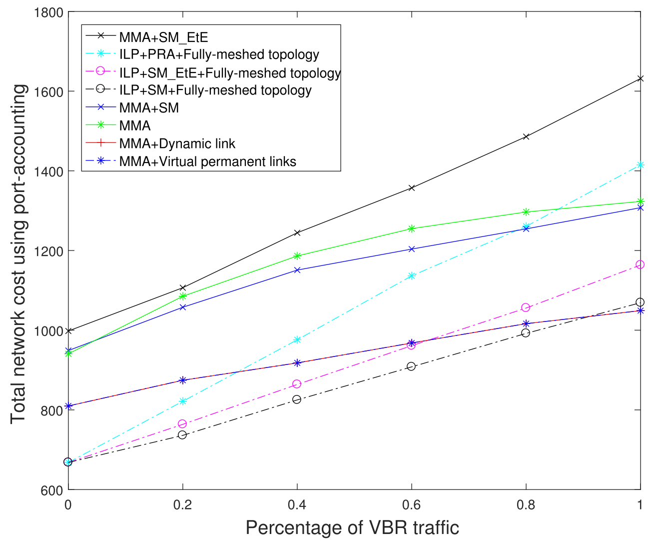

FIGURE 8. Comparison of MMA and ILP with fully-meshed topology in Layer 3.

be explained by the fact that the traffic streams between the same SD pairs can share common switching and transmission resources. By using ILP + SM, the total cost of the network can be further reduced, by nearly 1/6, compared with ILP + SM_EtE. As more traffic streams share common resources carried on each link, the effect of statistical multiplexing is significant in improving link utilization.

In Fig. 8, we present results where fully-meshed topology can be used in Layer 3 in all three ILP variants, so a full mesh of SD links can be used in the top layer. First, notice that savings of up to 1/4 of the total network cost is achieved by ILP + PRA with fully-meshed topology in Fig. 8 compared to ILP + PRA with original topology shown in Fig. 7. Notice that when the original topology is used in every layer, there is limited candidate choice of routing in each layer. When fully-meshed topology is applied in Layer 3, a direct link between all pair of nodes could be set which produces an optimal choice of

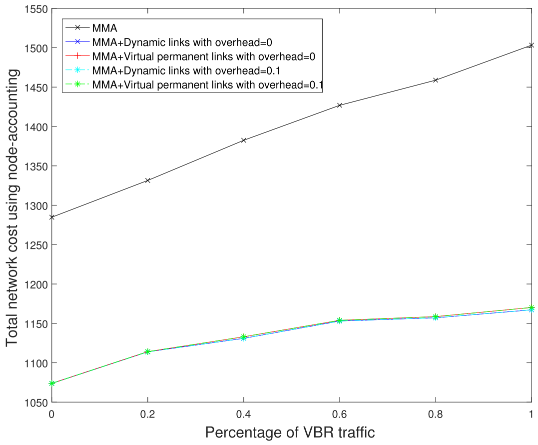

FIGURE 9. **Benefits of dynamic and virtual permanent links.**

topology, i. e. mesh. There are cases in the real network where the benefit of fully-meshed topology can be realized, e.g., when a tunnel label switched paths (LSPs) is used between all the provider edge (PE) routers that participate in a virtual private LAN service (VPLS).

Similar to Fig. 7, when statistical multiplexing is further applied in the variants ILP + SM_EtE and ILP + SM, we observe in Fig. 8, that in the example considered, they further save up to 1/5 and 1/3 of the total network cost, respectively, compared to ILP + PRA with fully-meshed topology. The results show that the use of a full mesh in the top layer is normally optimal according to ILP. The reason is that this topology saves the cost of switching/routing at the top layer.

MMA makes almost no assumptions about the traffic model, and hence it is able to cater for VBR traffic modelled by, for example, a PPBP. MMA involves VBR traffic streams by using the PPBP model, which captures the variability of Internet flows and therefore, it is applicable to real Internet traffic. By considering flow-size independent routing, MMA using PPBP is able to demonstrate the advantages of using dynamic links. In this way, a stream transported end-to-end may use different layers at different parts on its path depending on cost. The short and long flows may be carried on different routes and different layers. This enables us to investigate improvement in the efficiency of routing which can be achieved by flow-size dependent routing.

In Figs. 7 and 8, we also provide results for MMA where VBR traffic streams are modelled by PPBP. The flow set-up cost for dynamic links in Layer 3 and Layer 2 are set to 0 while virtual permanent links are disabled. In this case, setting up a dynamic link between two nodes in Layer 3 will benefit from no set-up cost and also from saving switching cost in Layer 3. This encourages PPBP traffic streams to use an end-to-end dynamic link in Layer 3. The aim here is to estimate the full potential in terms of cost-saving of dynamic links by making the assumption that their set-up cost is negligible. In Fig. 7, we observe that ILP + SM with original topology saves more cost than MMA which achieves less total network cost. The reason for this is, again, if the CBR proportion is high, the ILP utilizes the unused link resource more efficiently while MMA routes each traffic stream individually along one least cost path. As in the previous case, the network cost difference of ILP + SM and MMA is reduced as the proportion of VBR traffic increases because of the increase of statistical multiplexing and the benefit of spare capacity reuse.

In Fig. 7, the non-traffic-sharing is used in Layer 3, 2 and 1 for MMA + PRA and MMA. When dynamic or virtual permanent links are applied in MMA, they enable more efficient utilization of capacity as compared to the utilization achievable without dynamic or permanent links. When VBR is more than 80% of the total traffic, MMA can achieve a further cost reduction of cutting one fourth of the total network cost by taking advantage of dynamic or permanent links. The Layer 3 becomes a nearly fully meshed topology with dynamic links in MMA. The reason is that when more traffic streams are VBR, MMA can save more bandwidths by using dynamic or permanent links because different flows in a PPBP traffic stream may be transported by paths in different layers. Consider an example of setting up a dynamic link of an Ethernet path through a network. The setup process is mainly handled by the IP router which is in Layer 3. When flow set up cost is 0, using dynamic links are cheaper than transporting the traffic on multiple hops of static simple links, because the switching cost in the intermediate nodes is saved, and there is no cost in setting up dynamic or virtual permanent links.

In Fig. 8, in MMA, VBR traffic streams are modelled by PPBP parameters and traffic-sharing is used in Layers 2 and 3, which implies that large volumes of traffic are directly passed down to Layer 2, and Layer 2 is actually handling the traffic. The results show that for the examples studied, when dynamic or virtual permanent links are applied, the total network cost can be lower by up to 10% when all the traffic streams are VBR as compared to the ILP + SM_EtE variant which optimizes the routing. We observe in the figure that the VBR proportion increases and the potential benefit of statistical multiplexing increases, so that MMA can further benefit from the reuse of the space capacity, the gap in network cost between ILP + SM and MMA reduces until MMA actually outperforms ILP + SM when all the traffic streams are VBR. One reason is that the traffic with large flow sizes could be transported through a path in the lower layer at lower transport cost. Another reason that when traffic-sharing is used together with dynamic links more opportunities for routing flows are available and therefore more efficient solutions can be found and MMA finds them apparently very well.

In Fig. 8, we also provide results for the MMA + SM variant, in this case, the VBR traffic streams are modelled by Gaussian processes. The reason that Gaussian traffic MMA + SM can save more total network cost compared to MMA can be explained by the fact that Gaussian is normally a far smoother process that PPBP. Notice that in our case $\gamma = 1.5$, so that the flow size (burst) variance is infinite.

In Fig. 9, we compare MMA, MMA with dynamic links and MMA with virtual permanent links, where parameters of traffic-sharing (in Layers 2 and 3) are set the same as in the previous cases of Fig. 8. The set-up cost for virtual permanent links and dynamic links are set to 0 at Layer 3. Both of the originating cost and terminating cost equal to 1 and a transit cost in Layer 3 is either zero or non-zero. The node-accounting cost model is explored.

In scenarios involving node-accounting in Layer 3, when traffic-sharing is used for MMA with dynamic or virtual permanent links, the total network cost can be lower by up to 25% when all the traffic streams are VBR. The total network costs are the same for both transit cost (i.e., cost of passing through a transit node) set to zero and non-zero cases when dynamic or virtual permanent links are enabled. The reason is that there will always be a fully meshed network of dynamic or virtual permanent links in the result network. Use of a full mesh in the top layer is normally optimal according to both ILP and MMA. Since the networks are fully-meshed, there will be no transit cost for traffic streams.

These results have demonstrated that a design based on MMA can outperform ILP in a range of variants. However, the most important attribute of MMA is its scalability. In our examples, the running time of MMA is within in 2000 milliseconds, compared to ILP with original topology only, which in average is 1200 seconds, and in the case of ILP with fully-meshed topology, the running time is over 10 hours. Therefore, MMA is not only applicable to small networks, but scalable to larger networks, i. e. practical-sized networks.

In Fig. 10, total network cost and the proportion of dynamic links in use are plotted as a function of the flow setup cost. All the traffic and the cost model are the same as the settings used in Fig. 8 and the proportion of VBR traffic streams is fixed to 0.6. The results demonstrate that when flow setup cost is low, the reduction in network cost due to the use of dynamic links is significant. As seen in Fig. 10, when the flow set-up cost is 0, all of the large flows use dynamic links, which results in a fully meshed topology in Layer 3. As flow set-up cost is increased, usage of dynamic links falls off quickly, and when cost is over 0.4 which is expressed as a percentage overhead of Layer 3, dynamic links are of no benefit at all.

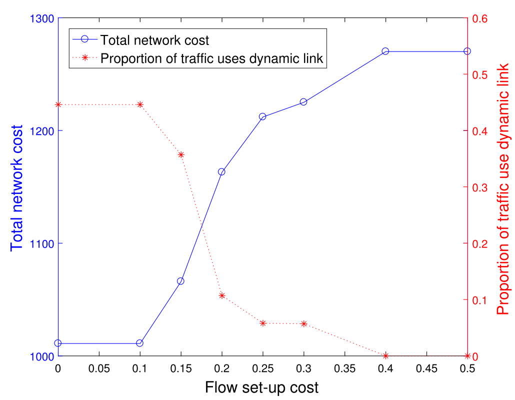

FIGURE 10. **Total cost and percentage of flow using dynamic link vs. flow set-up cost.**

B. RESULTS FOR LARGER NETWORKS

Since the PPBP traffic streams are split for the flow-size dependent routing based on a threshold value, it is necessary to explore whether flow-size dependent routing under different threshold is beneficial and can reduce the total network cost. The experiment is conducted by comparison two scenarios: traffic stream routes without flow-size dependent routing, the other is flow-size dependent routing with different flow set-up cost and different threshold. The traffic-sharing is applied to Layer 2 and 3.

Two example networks are presented here: (i) NSFNET with 14 nodes and 21 links (Fig. 11) [61]; (ii) Internet2 with of 53 nodes and 122 links (Fig. 12) [62]. For these practical networks, we still consider three layers which are IP/MPLS, OTN and WDM for Layer 3, 2 and 1, respectively. The size of module and unit cost for each layer are set according to [9] in the network. The PPBP traffic streams are randomly generated between all possible SD pairs in the network.

In the experiments, different thresholds are set and are used to separate small from large flows. The networks used for these experiments are NSFNET and Internet2. The results of the total network cost with various thresholds for NSFNET and Internet2 are shown in Figs. 13 and 14, respectively. In both Figs. 13 and 14, variation of total network cost with the choice of the threshold, which distinguishes short from long flows, is shown and provides intuition for how

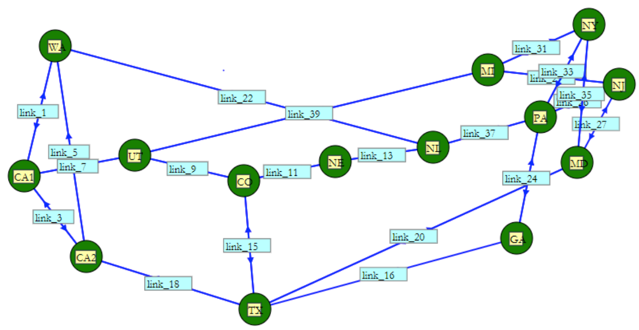

FIGURE 11. **Netml representation of NSFNET.**

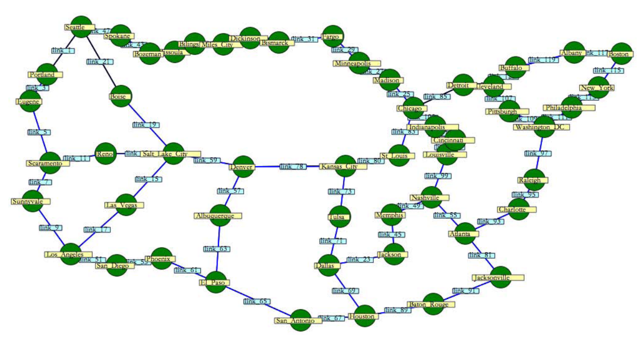

FIGURE 12. **Netml representation of Internet2.**

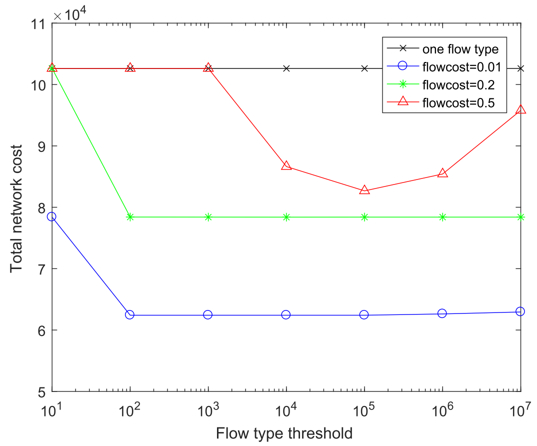

FIGURE 13. Total network cost and flow set-up Cost vs threshold for NSFNET.

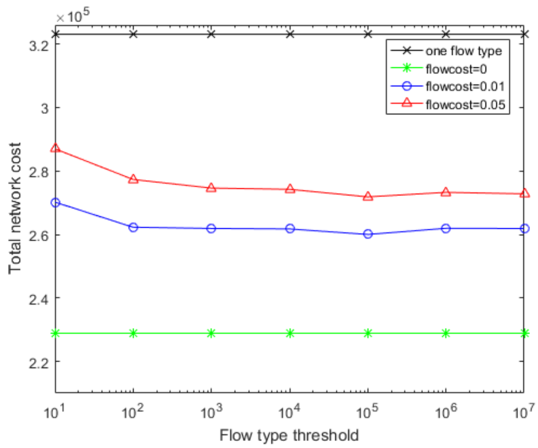

FIGURE 14. Total network cost and flow set-up Cost vs threshold for Internet2.

this threshold should be chosen. When the flow set-up cost becomes higher, the utilization of dynamic links is suppressed, and the choice of threshold between short and long flows becomes more critical to achieving lower total network cost by means of flow-size dependent routing. This is to be expected since by the nature of the Pareto distribution of flow sizes, most packets are in large flows, but most flows are small. Since switching cost in upper layers is saved by flow-size dependent routing, it is mainly larger flows which enjoy the benefits of dynamic links. The results also demonstrate when flow setup cost is small enough, flow-size dependent routing can save total network cost because of a large amount of traffic is carried in the layer below. when flow setup cost increases, the cost saving decreases dramatically. When the flow set-up cost is 0, because when traffic-sharing is in use, all the VBR traffic flows take advantage of the dynamic links by sharing the unused capacity in the links. In Figs. 13 and 14, total network cost saved by flow-size dependent routing is nearly 30%.

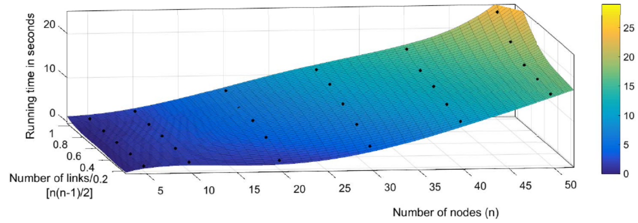

FIGURE 15. Running time in seconds of MMA on partially and fully meshed networks.

C. RUNNING TIME OF MMA

As stated in Section IV, as the number of nodes and links increases, an ILP optimal solution becomes more and more time-consuming, and eventually, the cost is prohibitive. The running time of MMA on a variety of sizes of networks has been determined to demonstrate that MMA is much faster than ILP. MMA has been applied to fully meshed networks with the number of nodes varying from 5 to 60 and with partially randomly deleted links; the time till convergence of MMA on a workstation with a 2.6 GHz processor has been recorded and the results are plotted in Fig. 15. In each network, the number of links is set to a specified fraction of $n(n-1)/2$ (which is the number of links for a fully meshed network). The surface shown in Fig. 15 is a least-squares best fit to the experiments, which are shown as the black dots in the plot. This numerical study results in a polynomial complexity of $O(|V|*|E|)+O(|V|^4)$ where $O(|V|^4)$ is the dominant term. The R-squared goodness-of-fit measure for this polynomial fit was found to be 0.9976, which indicates that the model fits the empirical results well. Given that the polynomial-time complexity is normally acceptable and scalable, MMA is a very significant step relative to ILP which is known to have high order complexity (and is probably NP-hard).

D. FURTHER VERIFICATION BY DOUBLE-ENTRY BOOKKEEPING

An important verification that we use is a method called double-entry bookkeeping as mentioned in Section I. It is a high-level check that the optimization algorithm targets the correct cost function. This method has been partly described in [10], that includes a cost report for an earlier version of MMA that does not include sharing schemes. It includes various settings of parameters, which verify that MMA achieves correct results. To implement double-entry bookkeeping, MMA computes a cost-per-packet for all traffic streams in the network. These costs enable MMA to calculate the total network cost by adding up these costs over all carried traffic, which should produce the same estimate of total network cost as obtained by adding up all the network resource costs associated with all links and nodes. An example of a full cost report of a four-node network experiment is provided in Appendix. In order to properly understand the cost model and running a network in MMA, we also provide examples with web-links as in Appendix.

E. FURTHER VALIDATION BY VISUALIZATION

In addition to the validation that we performed using ILP benchmarks, we can also validate the results, to a certain extent, by visualization. This highlights the importance of having visual tools, which enable users to see into the details of networks under study. Multi-layered networks are particularly difficult to visualize and it was felt, near the start of this research, that it would not be possible to develop satisfactory multi-layer design algorithms without powerful visualization tools for multi-layered networks.

The current version of Netml provides such means for visualization. This is illustrated in Fig. 16, that provides a Netml visualization of the optimal results achieved for a six-node network experiment, where we can observe that the links on average are indeed highly utilized. Observe that the red color on the links indicate highly utilized links and indicates efficient network operation. Also, observe that although most links are highly utilized, some links are not fully utilized because in some cases, a large module cannot be filled up. In addition, the details of the network design can also be observed.

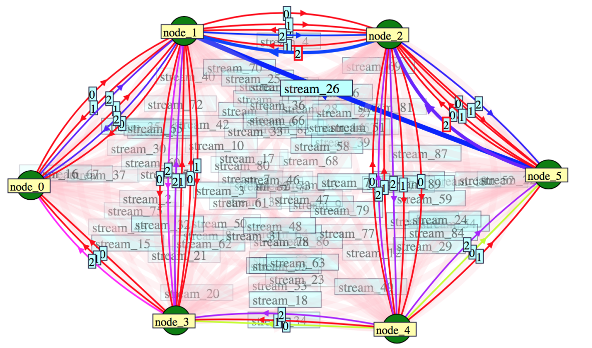

FIGURE 16. Netml visualization of a result network.

In Netml, we can observe the chosen path for each traffic stream. In Fig. 16, we present Netml output, where we choose most of the traffic streams to be blurred (light colored) except traffic stream 26, i.e. the traffic stream from Node 5 to Node 1, which we use as an example to describe its path. In particular, it is transported by links from Node 5 to Node 2 and Node 2 to Node 1 at Layer 2. Such method of validation by visualization is applicable to large networks that cannot be validated by ILP benchmarks.

VI. CONCLUSION

In this paper, we have presented a comprehensive model of a multi-layered network that involves realistic CBR and VBR (Gaussian and PPBP) traffic streams. The model has led to a new cost-based polynomial-time heuristic algorithm, called MMA, for optimizing resource provisioning, that achieves total network cost minimization by considering end-to-end routes. We have demonstrated the ability of MMA to consider dynamic links based on flow size dependent routing for PPBP traffic streams. Verification of MMA has been achieved by double-entry bookkeeping. Through extensive numerical experiments, we have presented results for MMA validation considering a range of ILP benchmarks for small networks. The results demonstrate the cost-benefit achieved by MMA through classification of flow sizes.

MMA also enables comparison between competing technologies in the same layer. This can be done by multiple MMA runs with different traffic sharing schemes (technologies) and comparison of the total network costs. With the evolution of technologies, different sharing schemes help service providers to explore different combinations between multiple alternative technologies in the same network with different layers. MMA can provide resource provisioning by consideration of efficient traffic management, and can guide telecommunications providers in evolution prediction and choice of technologies for future multi-layered networks.

APPENDIX

This appendix provides a cost verification report of a four-node network experiment which aims to verify that MMA achieves correct results. The cost is calculated by two means: cost by modules of each node and link and also computed by each cost-per-packet. The results show that the total network cost obtained by adding up cost associated with all links and nodes produces the same total network cost by adding up the cost over all carried traffic by using the cost per packet for the path used. This appendix also provides a table that includes examples with web links in which parameters are set in the Netml system.

12/28/2018 Netml V 4.65 Network view

Cost Report for Network fournode_costreport

Cost Per Layer and Module Type

Layer 0

phy: number of modules (of type 0, sharing type unsharing): 6 ($3 per module) , total cost for this module type: $18

Discount for unutilised capacity: $ 0.000000

Total cost of Layer 0 and below: $18

Layer 1

eth: number of modules (of type 0, sharing type unsharing): 9 ($5 per module) , total cost for this module type: $45

Discount for unutilised capacity: $ 0.000000

Total cost of Layer 1: $45

Total cost of Layer 1 and below: $63

Layer 2

ip: number of modules (of type 0, sharing type unsharing): 18 ($10 per module) , total cost for this module type: $180

Discount for unutilised capacity: $ 0.000000

Total cost of Layer 2: $180

Total cost of Layer 2 and below: $243

Cost Per Layer and Traffic Flow from cost-per-packet

This network has 3 layers and its average packet size is 1000

Layer 0.

trafficname: traffic (pkts/s) x price = cost-below + transport-cost + flowsetup-cost + pathsetup-cost = cost-for-this-traffic (added up from components)

T_link_1_Lay1_simple_physexists_1398: 375000 pkts/s x 0.000016$/packet/s/y = 0 + 6 + 0 + 0 = $6 [Only correct when all fs have the same costperpkt](6)

T_link_3_Lay1_simple_physexists_1399: 0 pkts/s x 0.000012$/packet/s/y = 0 + 0 + 0 + 0 = $0 [Only correct when all fs have the same costperpkt](0)

T_link_13_Lay1_simple_physexists_1400: 375000 pkts/s x 0.000016$/packet/s/y = 0 + 6 + 0 + 0 = $6 [Only correct when all fs have the same costperpkt](6)

T_link_15_Lay1_simple_physexists_1401: 0 pkts/s x 0.000012$/packet/s/y = 0 + 0 + 0 + 0 = $0 [Only correct when all fs have the same costperpkt](0)

T_link_17_Lay1_simple_physexists_1402: 375000 pkts/s x 0.000016$/packet/s/y = 0 + 6 + 0 + 0 = $6 [Only correct when all fs have the same costperpkt](6)

T_link_19_Lay1_simple_physexists_1403: 0 pkts/s x 0.000012$/packet/s/y = 0 + 0 + 0 + 0 = $0 [Only correct when all fs have the same costperpkt](0)

Total cost of Layer 0 and below: $18.000000 (18)

Layer 1.

trafficname: traffic (pkts/s) x price = cost-below + transport-cost + flowsetup-cost + pathsetup-cost = cost-for-this-traffic (added up from components)

T_link_1_Lay1_simple_physexists__Lay2_simple_physexists_1392: 375000 pkts/s x 0.000056$/packet/s/y = 6 + 15 + 0 + 0 = $21 [Only correct when all fs have the same costperpkt](21)

T_link_3_Lay1_simple_physexists__Lay2_simple_physexists_1393: 0 pkts/s x 0.00004$/packet/s/y = 0 + 0 + 0 + 0 = $0 [Only correct when all fs have the same costperpkt](0)

http://localhost:8080/netml4_65/costreport.jsp?netname=fournode_costreport&location=Results 1/2

12/28/2018 Netml V 4.65 Network view

T_link_13_Lay1_simple_physexists__Lay2_simple_physexists_1394: 375000 pkts/s x 0.000056$/packet/s/y = 6 + 15 + 0 + 0 = $21 [Only correct when all fs have the same costperpkt](21)

T_link_15_Lay1_simple_physexists__Lay2_simple_physexists_1395: 0 pkts/s x 0.00004$/packet/s/y = 0 + 0 + 0 + 0 = $0 [Only correct when all fs have the same costperpkt](0)

T_link_17_Lay1_simple_physexists__Lay2_simple_physexists_1396: 375000 pkts/s x 0.000056$/packet/s/y = 6 + 15 + 0 + 0 = $21 [Only correct when all fs have the same costperpkt](21)

T_link_19_Lay1_simple_physexists__Lay2_simple_physexists_1397: 0 pkts/s x 0.00004$/packet/s/y = 0 + 0 + 0 + 0 = $0 [Only correct when all fs have the same costperpkt](0)

Total cost of Layer 1: $45 (cost below: 18, thislayer: 45.000000 (added up from components: 45))

Total cost of Layer 1 and below: $63.000000 (63)

Layer 2.

trafficname: traffic (pkts/s) x price = cost-below + transport-cost + flowsetup-cost + pathsetup-cost = cost-for-this-traffic (added up from components)

stream_8: 112500 pkts/s x 0.00072$/packet/s/y = 21 + 60 + 0 + 0 = $81 [Only correct when all fs have the same costperpkt](81)

stream_38: 112500 pkts/s x 0.00072$/packet/s/y = 21 + 60 + 0 + 0 = $81 [Only correct when all fs have the same costperpkt](81)

stream_68: 112500 pkts/s x 0.00072$/packet/s/y = 21 + 60 + 0 + 0 = $81 [Only correct when all fs have the same costperpkt](81)

Total cost of Layer 2: $180 (cost below: 63, thislayer: 180.000000 (added up from components: 180))

Total cost of Layer 2 and below: $243.000000 (243)

http://localhost:8080/netml4_65/costreport.jsp?netname=fournode_costreport&location=Results 2/2

TABLE 3. URLs of example networks used in Netml.

Experiment	Figures	URL
Non-traffic sharing scheme with port-accounting	Fig. 3	http://netml.usq.edu.au/netml4_65/index.jsp?netname=fournode_unsharing_costreport&location=Demo&userid=cxing5
Traffic-sharing scheme with port-accounting	Fig. 3	http://netml.usq.edu.au/netml4_65/index.jsp?netname=fournode_linksharing_costreport&location=Demo&userid=cxing5
Traffic sharing scheme with node-accounting	Fig. 3	http://netml.usq.edu.au/netml4_65/index.jsp?netname=fournode_nodesharing_costreport&location=Demo&userid=cxing5
Six-node network example	Figs. 7-9	http://netml.usq.edu.au/netml4_65/index.jsp?netname=sixnode_9_10&location=Demo&userid=cxing5

REFERENCES

[1] R. Bush and D. Meyer, *Some Internet Architectural Guidelines and Philosophy*, document RFC 3439 (Informational), Internet Engineering Task Force, Dec. 2002. [Online]. Available: https://www.ietf.org/rfc/rfc3439.txt

[2] F. Rambach *et al.*, "A multilayer cost model for Metro/Core networks," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 5, no. 3, pp. 210–225, Mar. 2013.

[3] S. Roy *et al.*, "Evaluating efficiency of multi-layer switching in future optical transport networks," in *Proc. OFC/NFOEC*, Anaheim, CA, USA, Mar. 2013, pp. 1–3.

[4] *Architecture of Optical Transport Networks (OTN)*, document Rec. ITU-T G.872, 2012. [Online]. Available: http://www.itu.int/rec/T-REC-G.872

[5] J. W. Guck, A. Van Bemten, M. Reisslein, and W. Kellerer, "Unicast QoS routing algorithms for SDN: A comprehensive survey and performance evaluation," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 388–415, 1st Quart., 2018.

[6] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "A survey on software-defined networking,' *Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 27–51, 1st Quart., 2014.

[7] R. G. Addie, Y. Peng, and M. Zukerman, "NetML: Networking networks," in *Proc. ACGS*, Sydney, NSW, Australia, Dec. 2011, pp. 1055–1060.

[8] M. Pióro and D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks*. San Mateo, CA, USA: Morgan Kaufmann, Jul. 2004.

[9] I. Katib and D. Medhi, "IP/MPLS-over-OTN-over-DWDM multilayer networks: An integrated three-layer capacity optimization model, a heuristic, and a study," *IEEE Trans. Netw. Service Manage.*, vol. 9, no. 3, pp. 240–253, Sep. 2012.

[10] R. G. Addie, Y. Peng, M. Albdair, C. Xing, D. Fatseas, and M. Zukerman, "Cost modelling and validation in network optimization," in *Proc. ITNAC*, Sydney, NSW, Australia, Nov. 2015, pp. 11–16.

[11] A. Sangster, "The genesis of double entry bookkeeping," *Accounting Rev.*, vol. 91, no. 1, pp. 299–315, 2016.

[12] W. Vereecken *et al.*, "Power consumption in telecommunication networks: Overview and reduction strategies," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 62–69, Jun. 2011.

[13] J. R. Current, C. S. ReVelle, and J. L. Cohon, "The hierarchical network design problem," *Eur. J. Oper. Res.*, vol. 27, no. 1, pp. 57–66, 1986.

[14] E. Oki, K. Shiomoto, D. Shimazaki, N. Yamanaka, W. Imajuku, and Y. Takigawa, "Dynamic multilayer routing schemes in GMPLS-based IP+optical networks," *IEEE Commun. Mag.*, vol. 43, no. 1, pp. 108–114, Jan. 2005.

[15] H. Höller and S. Voß, "A heuristic approach for combined equipment-planning and routing in multi-layer SDH/WDM networks," *Eur. J. Oper. Res.*, vol. 171, no. 3, pp. 787–796, 2006.

[16] O. Gerstel *et al.*, "Multi-layer capacity planning for IP-optical networks," *IEEE Commun. Mag.*, vol. 52, no. 1, pp. 44–51, Jan. 2014.

[17] V. Lopez, D. Konidis, D. Siracusa, C. Rožić, I. Tomkos, and J. P. Fernandez-Palacios, "On the benefits of multilayer optimization and application awareness," *J. Lightw. Technol.*, vol. 35, no. 6, pp. 1274–1279, Mar. 15, 2017.

[18] R. Huelsermann, M. Gunkel, C. Meusburger, and D. A. Schupke, "Cost modeling and evaluation of capital expenditures in optical multilayer networks," *J. Opt. Netw.*, vol. 7, no. 9, pp. 814–833, 2008.

[19] D. T. Hai, "An optimal design framework for 1 + 1 routing and network coding assignment problem in WDM optical networks," *IEEE Access*, vol. 5, pp. 22291–22298, 2017.

[20] C. Rožić, D. Klonidis, and I. Tomkos, "A survey of multi-layer network optimization," in *Proc. ONDM*, Cartagena, Spain, May 2016, pp. 1–6.

[21] T. Engel, A. Autenrieth, and J.-C. Bischoff, "Packet layer topologies of cost optimized transport networks multi-layer network optimization," in *Proc. ONDM*, Braunschweig, Germany, Feb. 2009, pp. 1–7.

[22] L. Gouveia, P. Patricio, A. F. de Sousa, and R. Valadas, "MPLS over WDM network design with packet level QoS constraints based on ILP models," in *Proc. INFOCOM*, San Francisco, CA, USA, Mar./Apr. 2003, pp. 576–586.

[23] S. Fu, H. Wen, J. Wu, and B. Wu, "Cross-networks energy efficiency tradeoff: From wired networks to wireless networks," *IEEE Access*, vol. 5, pp. 15–26, 2016.

[24] C. Obreque and V. Marianov, "An optimal procedure for solving the hierarchical network design problem," *IIE Trans.*, vol. 39, no. 5, pp. 513–524, 2007.

[25] P. Papanikolaou, K. Christodoulopoulos, and E. Varvarigos, "Incremental planning of multi-layer elastic optical networks," in *Proc. ONDM*, Budapest, Hungary, May 2017, pp. 1–6.

[26] G. Thouénon, C. Betoule, E. Pincemin, P. S. Khodashenas, J. M. Rivas-Moscoso, and I. Tomkos, "All-optical vs. electrical aggregations CAPEX comparisons in a fully-flexible multi-layer transport network," in *Proc. ECOC*, Valencia, Spain, Sep./Oct. 2015, pp. 1–3.

[27] A. M. C. A. Koster, S. Orlowski, C. Raack, G. Baier, and T. Engel, "Single-layer cuts for multi-layer network design problems," in *Telecommunications Modeling, Policy, and Technology* (Operations Research/Computer Science Interfaces), vol. 44, S. Raghavan, B. Golden, and E. Wasil, Eds. Boston, MA, USA: Springer, 2008.

[28] B. Fortz and M. Poss, "An improved Benders decomposition applied to a multi-layer network design problem," *Oper. Res. Lett.*, vol. 37, no. 5, pp. 359–364, 2009.

[29] R. Aparicio-Pardo, B. Garcia-Manrubia, P. Pavon-Marino, N. Skorin-Kapov, and M. Furdek, "Balancing CapEx reduction and network stability with stable routing–virtual topology capacity adjustment (SR–VTCA)," *Opt. Switching Netw.*, vol. 10, no. 4, pp. 343–353, 2013.

[30] T. M. J. Ng and D. B. Hoang, "Joint optimization of capacity and flow assignment in a packet-switched communications network," *IEEE Trans. Commun.*, vol. COM-35, no. 2, pp. 202–209, Feb. 1987.
[31] E. V. Denardo, *Dynamic Programming: Models and Applications*. New York, NY, USA: Dover, Apr. 2003.
[32] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, Feb. 1994.
[33] M. E. Crovella, M. S. Taqqu, and A. Bestavros, "Heavy-tailed probability distributions in the World Wide Web," in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, R. J. Adler, R. E. Feldman, and M. S. Taqqu, Eds. Cambridge, MA, USA: Birkhäuser, 1998, pp. 3–25.
[34] M. F. Arlitt and C. L. Williamson, "Internet Web servers: Workload characterization and performance implications," *IEEE/ACM Trans. Netw.*, vol. 5, no. 5, pp. 631–645, Oct. 1997.
[35] A. Williams, M. Arlitt, C. Williamson, and K. Barker, "Web workload characterization: Ten years later," in *Web Content Delivery*, vol. 2, X. Tang, J. Xu, and S. T. Chanson, Eds. New York, NY, USA: Springer, 2005, pp. 3–21.
[36] A. B. Downey, "Lognormal and Pareto distributions in the Internet," *Comput. Commun.*, vol. 28, no. 7, pp. 790–801, May 2005.
[37] V. Eramo and F. G. Lavacca, "Processing and bandwidth resource allocation in multi-provider NFV cloud infrastructures interconnected by elastic optical networks," in *Proc. ICTON*, Bucharest, Romania, Jul. 2018, pp. 1–6.
[38] M. Parulekar and A. M. Makowski, "Tail probabilities for M/G/∞ input processes (I): Preliminary asymptotics," *Queueing Syst.*, vol. 27, nos. 3–4, pp. 271–296, Apr. 1997.
[39] N. G. Duffield, "Queueing at large resources driven by long-tailed $M/G/\infty$-modulated processes," *Queueing Syst.*, vol. 28, pp. 245–266, May 1998.
[40] R. G. Addie, T. D. Neame, and M. Zukerman, "Performance evaluation of a queue fed by a Poisson Pareto burst process," *Comput. Netw.*, vol. 40, no. 3, pp. 377–397, Oct. 2002.
[41] M. Zukerman, T. D. Neame, and R. G. Addie, "Internet traffic modeling and future technology implications," in *Proc. INFOCOM*, San Francisco, CA, USA, Mar./Apr. 2003, pp. 587–596.
[42] R. G. Addie, T. D. Neame, and M. Zukerman, "Performance analysis of a Poisson–Pareto queue over the full range of system parameters," *Comput. Netw.*, vol. 53, no. 7, pp. 1099–1113, May 2009.
[43] W. Whitt, "The queueing network analyzer," *Bell Syst. Tech. J.*, vol. 62, no. 9, pp. 2779–2815, Nov. 1983.
[44] P. J. Kuehn, "Approximate analysis of general queuing networks by decomposition," *IEEE Trans. Commun.*, vol. COM-27, no. 1, pp. 113–126, Jan. 1979.
[45] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*, 4th ed. Hoboken, NJ, USA: Wiley, 2008.
[46] J. Roberts, U. Mocci, and J. Virtamo, Eds., *Broadband Network Teletraffic: Performance Evaluation and Design of Broadband Multiservice Networks, Final Report of Action COST 242* (Lecture Notes in Computer Science). Berlin, Germany: Springer, Nov. 1996.
[47] R. van de Meent, M. Mandjes, and A. Pras, "Gaussian traffic everywhere?" in *Proc. ICC*, Istanbul, Turkey, Jun. 2006, pp. 573–578.
[48] A. Simonian and J. Virtamo, "Transient and stationary distributions for fluid queues and input processes with a density," *SIAM J. Appl. Math.*, vol. 51, no. 6, pp. 1732–1739, Dec. 1991.
[49] C. Knessl and J. A. Morrison, "Heavy-traffic analysis of a data-handling system with many sources," *SIAM J. Appl. Math.*, vol. 51, no. 1, pp. 187–213, Feb. 1991.
[50] R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 7, pp. 968–981, Sep. 1991.
[51] R. G. Addie, D. Fatseas, and M. Zukerman, "Optimizing multi-layered networks towards a transparently optical Internet," in *Proc. ICTON*, Munich, Germany, Jun./Sep. 2010, pp. 1–4.
[52] V. Abramov, R. G. Addie, F. Li, Y. Peng, and M. Zukerman, "A new teletraffic approach for network planning and evolution prediction," in *Proc. OFC/NFOEC*, Los Angeles, CA, USA, Mar. 2012, pp. 1–3.
[53] R. G. Addie, D. Fatseas, Y. Peng, F. Li, and M. Zukerman, "How good (or bad) is shortest path routing in layered networks," in *Proc. ATNAC*, Brisbane, QLD, Australia, Nov. 2012, pp. 1–6.
[54] Y. Peng *et al.*, "Validation of multi-layer network optimization," in *Proc. ICTON*, Trento, Italy, Jul. 2016, pp. 1–4.
[55] Y. Peng, "Modeling, design and optimization of multi-layer telecommunications networks," Ph.D. dissertation, Dept. Electron. Eng., City Univ. Hong Kong, Hong Kong, 2016.
[56] L. Guo and I. Matta, "The war between mice and elephants," in *Proc. ICNP*, Riverside, CA, USA, Nov. 2001, pp. 180–188.
[57] M. Ajmone Marsan, M. Garetto, P. Giaccone, E. Leonardi, E. Schiattarella, and A. Tarello, "Using partial differential equations to model TCP mice and elephants in large IP networks," *IEEE/ACM Trans. Netw.*, vol. 13, no. 6, pp. 1289–1301, Dec. 2005.
[58] R. G. Addie, M. Zukerman, and T. D. Neame, "Broadband traffic modeling: Simple solutions to hard problems," *IEEE Commun. Mag.*, vol. 36, no. 8, pp. 88–95, Aug. 1998.
[59] ILOG CPLEX, ILOG, Inc., Mountain View, CA, USA. *IBM ILOG CPLEX Optimization Studio*. Accessed: Jan. 24, 2019. [Online]. Available: https://www.ibm.com/products/ilog-cplex-optimization-studio?mhq=ilog&mhsrc=ibmsearch_a
[60] F. Pukelsheim, "The three sigma rule," *Amer. Statist.*, vol. 48, no. 2, pp. 88–91, 1994.
[61] S. Arakawa, M. Murata, and H. Miyahara, "Functional partitioning for multi-layer survivability in IP over WDM networks," *IEICE Trans. Commun.*, vol. E83-B, no. 10, pp. 2224–2233, 2000.
[62] *Internet2 Advanced Networking*. Accessed: Jan. 24, 2019. [Online]. Available: http://www.internet2.edu/network/

CHANG XING received the B.Eng. degree in electronic information engineering from the Hebei University of Science and Technology, Hebei, China, and the M.Sc. degree from the City University of Hong Kong, in 2014, where she is currently pursuing the Ph.D. degree with the Department of Electronic Engineering. Her research interests include telecommunication network optimization and multi-layered network design.

RONALD G. ADDIE received the B.Sc. degree and the Ph.D. degree in the area of semi-Markov queues from Monash University, in 1972 and 1986, respectively. From 1972 to 1992, he worked in Telecom Australia Research Laboratories, where he was involved in research in digital communications, teletraffic, and network analysis and design. He was also a contributor to the concept of virtual path which has since been adopted as a part of the B-ISDN standard. In 1992, he moved to the University of Southern Queensland, where he is a member of the Faculty of Science, Department of Mathematics and Computing, where he has served as the Head of the department. His current research interests include quality of service management in TCP/IP networks, rare event simulation, and the architecture of multi-layered networks.

YU PENG received the B.Eng. degree in information engineering, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong, in 2010 and 2016, respectively. His research interests include teletraffic theory, optical network dimensioning, and performance evaluation of telecommunication networks. Since 2015, he has been working in the finance industry in Hong Kong.

RONGPING LIN received the Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2013. He is currently an Associate Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China. From 2013 to 2014, he was a Senior Research Assistant with the Electronic Engineering Department, City University of Hong Kong. His research interests include optical networks, SDN networks, and network function virtualization.

FAN LI received the B.Eng. degree in electronic information engineering from the Nanjing University of Science and Technology, Nanjing, China, and the M.Sc. degree from the City University of Hong Kong, where he is currently pursuing the Ph.D. degree with the Department of Electronic Engineering. His research interests include optical network dimensioning, network congestion control, and net neutrality.

WENJIE HU received the B.Eng. degree in information engineering from the City University of Hong Kong, and the M.Sc. degree in computer science from Yale University, in 2013 and 2014, respectively. She is currently a Software Engineer at Google LLC, Menlo Park, CA, USA.

VYACHESLAV M. ABRAMOV received the degree from the Tajik State University, Dushanbe, Tajikistan, in 1977, and the Ph.D. degree from Tel Aviv University, in 2005. From 1977 to 1992, he was with the Research Institute of Economy, under the Tajikistan State Planning Committee (GosPlan). In 1992, he repatriated to Israel, and from 1994 to 2001, he worked in software companies of Israel as a software engineer and algorithms developer. From 2002 to 2005, he was an Assistant and a Lecturer with the Judea and Samaria College, Tel Aviv University, and the Holon Institute of Technology. Since 2005, he has been with the School of Mathematical Sciences, Monash University, Australia. He was also with Swinburne University and RMIT University, Australia, and the City University of Hong Kong, China. His scientific interests are mainly focused on the theory and application of queueing systems. He has authored various publications in leading scientific journals on operations research, applied mathematics, and probability, such as *Journal of Applied Probability*, *Journal of Theoretical Probability*, *Queueing Systems*, *SIAM Journal of Applied Mathematics*, *Annals of Operations Research*, and other journals.

MOSHE ZUKERMAN (M'87–SM'91–F'07) received the B.Sc. degree in industrial engineering and management, and the M.Sc. degree in operations research from the Technion – Israel Institute of Technology, Haifa, Israel, and the Ph.D. degree in engineering from the University of California, Los Angeles, CA, USA, in 1985. He was an independent Consultant with the IRI Corporation, and a Postdoctoral Fellow with the University of California, from 1985 to 1986. He was with Telstra Research Laboratories, first as a Research Engineer (1986–1997), and as a Project Leader (1988–1997). He has also taught and supervised graduate students at Monash University, from 1990 to 2001. From 1997 to 2008, he was with The University of Melbourne, Melbourne, VIC, Australia. In 2008, he joined the City University of Hong Kong, as a Chair Professor of information engineering, and was the Team Leader. He has over 300 publications in scientific journals and conference proceedings. He has served on various editorial boards such as *Computer Networks*, *IEEE Communications Magazine*, IEEE Journal of Selected Areas in Communications, the IEEE/ACM Transactions on Networking, and the *International Journal of Communication Systems*.