Joint Optimization of Intra- and Inter-Autonomous System Traffic Engineering

Kin-Hon Ho, Michael Howarth, Ning Wang, George Pavlou and Stylianos Georgoulas

Centre for Communication Systems Research, University of Surrey, Guildford, Surrey GU2 7XH, UK

Email: {K.Ho, M.Howarth, N.Wang, G.Pavlou, S.Georgoulas}@surrey.ac.uk

***Abstract*—Traffic Engineering (TE) is used to optimize IP operational network performance. The existing literature generally considers intra- and inter-AS (Autonomous System) TE independently. However, the overall network performance may not be truly optimized when these aspects are considered separately. This is due to the interaction between intra- and inter-AS TE, where a solution of intra-AS TE may not be a good input to inter-AS TE and vice versa. To remedy this situation, we propose considering intra-AS aspects during inter-AS TE and vice versa. We propose a joint optimization of intra- and inter-AS TE to further improve the overall network performance by simultaneously finding the best egress points for the inter-AS traffic and the best routing scheme for the intra-AS traffic. Three strategies are presented to attack the problem, namely sequential, nested and integrated optimization. Our simulation study shows that, compared to sequential and nested optimization, integrated optimization can significantly improve the overall network performance by accommodating 30%-60% more traffic demands.**

I. INTRODUCTION

Traffic Engineering (TE) [1] is the set of techniques that optimize IP operational network performance by tactically routing traffic on a path other than the one would have been chosen if standard routing methods had been used. The task of TE is: given a network topology and a collected set of traffic demands (i.e. traffic matrix), determine the best routing for the traffic so that the overall network performance is optimized.

Today's Internet is a collection of more than 18,000 Autonomous Systems (ASes), each being an administrative region governed by its own network policies and routing protocols. Open Shortest Path First (OSPF) and Intermediate System-Intermediate System (ISIS) are the common intra-AS routing protocols, while Border Gateway Protocol (BGP) is the de facto inter-AS routing protocol. With the hierarchical characteristics of the Internet, traffic is routed *within* an AS or *between* ASes. Thus, TE can be broadly divided into two types: intra-AS and inter-AS.

In **intra-AS TE**, the operator of an AS controls traffic routing within the network by either optimizing the link weights of the corresponding routing protocol (mostly OSPF or IS-IS) or establishing Label Switched Paths (LSPs) through MultiProtocol Label Switching (MPLS). Typical intra-AS TE optimization objectives are to minimize network bandwidth consumption and to achieve load balancing within the network. **Inter-AS TE** [14,15], on the other hand, aims to control traffic entering and exiting an AS using optimization objectives such as load balancing over inter-AS links. It is commonly assumed that these inter-AS links are frequently congestion points [6,13]. For a particular AS, the network operator can control traffic exiting the AS by assigning the traffic to the 'best' egress points. This is called *Outbound* inter-AS TE. Likewise, the network operator can also control traffic entering the AS by selecting the 'best' ingress points, which is called *Inbound* inter-AS TE. In current practice, the commonly used method to enforce inter-AS TE is by adjusting BGP route attributes.

Internet Service Provider (ISP) networks typically carry both *intra-AS traffic* that is routed only within their networks and *inter-AS traffic* that is routed not only within their networks but also across other ASes. ISPs can employ both intra- and inter-AS TE to optimize the routing of these types of traffic. However, although some work exists on intra- and inter-AS TE, much of the existing literature deals with them *separately*: prior intra-AS TE work has assumed that ingress and egress points of inter-AS traffic do not change whereas prior inter-AS TE work has not considered route optimization within an AS. This is often inappropriate and results in suboptimal overall network performance due to the following two interaction effects between intra- and inter-AS TE:

- **Inter-on-Intra-AS TE**: The first interaction is the effect of inter-AS TE on the performance of intra-AS TE. Inter-AS TE can change the ingress and egress points of inter-AS traffic, thus causing the traffic to be routed on different ingress-to-egress paths within the network. This fundamentally changes the intra-AS traffic matrix, i.e. traffic load between each ingress and egress node pair. Such a change could therefore significantly affect the intra-AS TE solution.
- **Intra-on-Inter-AS TE**: The second interaction is the effect of intra-AS TE on the performance of inter-AS TE. It arises when outbound inter-AS TE and OSPF/ISIS-based intra-AS TE are used. One current practice is known as hot potato routing [2,3]. In this approach, a BGP router will choose the closest exit as measured by the lowest IGP (Interior Gateway Protocol) cost among multiple equally-good egress points towards a downstream routing prefix. If the IGP costs are changed by intra-AS TE, some inter-AS traffic flows will in general be shifted to a new set of closest egress points. This could lead to congestion on these new egress points.

If intra- and inter-AS TE are not jointly optimized, a sequential approach may therefore be regarded as the current practice in the collective use of intra- and inter-AS TE. In this sequential approach, the solution of inter-AS TE becomes the input for intra-AS TE or vice versa. However, since the objectives and constraints of the subsequent stage are not taken into account, the decisions made at one stage often do not provide a good input for the subsequent one, sometimes even leading to infeasible inputs. As a result, it is difficult to

1-4244-0143-7/06/$20.00

claim that a truly good overall TE solution has been found when each TE is considered separately. We therefore propose to consider inter-AS TE during intra-AS TE and vice versa. In this paper, we propose a joint optimization of intra- and inter-AS TE as an effective means to achieve better overall TE solutions than the one obtained by the sequential approach. Specifically, we investigate the following two challenges:

- *How should intra- and inter-AS TE be combined and how do we formulate their joint optimization?*
- *How can we solve the joint TE optimization problem to achieve a better overall network performance?*

Our contributions in this paper are as follows. First of all, we explain with examples the two interaction effects that can lead to suboptimal overall network performance. Then, for the first challenge above, we formulate a bi-criteria joint optimization problem of intra- and inter-AS TE with an aim of optimizing their objectives simultaneously. Since the two interaction effects can generally apply to any intra- and inter-AS TE approach, it is possible to formulate the joint optimization problem for all the combinations of intra- and inter-AS TE approaches. However, as the primary objective of this paper is to illustrate the point of interest on joint optimization of intra- and inter-AS TE in general, we only consider MPLS-based intra-AS TE and outbound inter-AS TE as an example in the joint optimization problem formulation. This problem formulation should be valid and practical as both intra-AS MPLS TE and outbound inter-AS TE are nowadays widely researched in academia and employed in industry. Moreover, a potential use of this problem formulation could be to optimize BGP/MPLS VPN provisioning, a subject which is currently attracting a great deal of industrial attention.

For the second challenge above, we consider three strategies to solve the joint TE optimization problem, namely sequential, nested and integrated optimization. These strategies aim to obtain non-dominated solutions with respect to the intra- and inter-AS TE objectives. We evaluate the performance of these strategies by simulation using Rocketfuel [4] topologies and synthetic traffic matrices. Our simulation results reveal that better overall network performance can be achieved by integrated optimization, solving intra- and inter-AS TE simultaneously. The performance improvement could allow the network to support a 30%-60% increase in the traffic demands. We believe that our work provides an insight into the interactions between intra- and inter-AS TE, enabling ISPs to further optimize the performances of their networks over the current practice sequential approaches.

The rest of this paper is structured as follows. In the next section, we provide background on intra- and inter-AS TE. Section III explains the two interaction effects with examples. We formulate the joint intra- and inter-AS TE optimization problem in Section IV and present strategies to solve it in Section V. In Sections VI and VII, we present our evaluation methodology and simulation results for these strategies. Sections VIII and IX present related work and conclusion.

II. TRAFFIC ENGINEERING

Traffic engineering typically takes input elements such as traffic matrix, network topology, and then executes algorithms to produce a set of traffic routing plans that optimize the overall network performance. In this section, we provide background on these elements that are of importance to understand our work. For ease of presentation, Table 1 shows the notation used throughout this paper.

Table 1. Notation used in this paper

NOTATION	DESCRIPTION
K	A set of downstream routing prefixes
I	A set of ingress points
J	A set of egress points (inter-AS links)
E_{intra}	A set of intra-AS links
$t\_inter(i,k)$	Bandwidth demand of the inter-AS traffic flow at ingress point $i \in I$ destined to routing prefix $k \in K$
$t\_loc(i,j)$	Bandwidth demand of the local intra-AS traffic flow between ingress point $i \in I$ and egress point $j \in J$
$t\_intra(i,j)$	Bandwidth demand of the intra-AS traffic flow between ingress point $i \in I$ and egress point $j \in J$
$T\_Inter(I,K)$	Inter-AS (ingress-to-prefix) traffic matrix consisting of all $t\_inter(i,k)$
$T\_Loc(I,J)$	Local intra-AS (ingress-to-egress) traffic matrix consisting of all $t\_loc(i,j)$
$T\_Intra(I,J)$	Intra-AS traffic matrix consists of all $t\_intra(i,j)$
$Out(k)$	A set of egress points that has reachability to routing prefix k
C^j_{inter}	Capacity of egress point j (inter-AS link)
bw^j_{inter}	Residual bandwidth of C^j_{inter}
C^l_{intra}	Capacity of intra-AS link l
bw^l_{intra}	Residual bandwidth of C^l_{intra}
$x^j_{i,k}$	A binary variable indicating whether inter-AS traffic flow $t\_inter(i,k)$ is assigned to egress point j
$y^l_{i,j}$	A binary variable indicating whether intra-AS traffic flow $t\_intra(i,j)$ is assigned to intra-AS link l
$P_{i,j}$	A set of candidate paths realizing intra-AS traffic flow $t\_intra(i,j)$
$w_{i,j,p}$	A binary variable indicating whether path $p \in P_{i,j}$ is chosen to realize the traffic flow $t\_intra(i,j)$
$s(i,k)$	A variable storing the egress point that has been assigned to $t\_inter(i,k)$

A. Internet Traffic Types

According to [5], Internet traffic received by an AS can be classified into four types: *internal* traffic that travels from an ingress access link to an egress access links; *transit* traffic that travels from an ingress peering link to an egress peering link; *inbound* traffic that travels from an ingress peering link to an egress access link, and *outbound* traffic that travels from an ingress access link to an egress peering link. As mentioned in the introduction, this paper only considers outbound inter-AS TE which selects optimal egress peering points for inter-AS traffic (i.e transit and outbound traffic). We therefore do not manipulate the potential ingress peering point selection in a similar manner. Given the fact that ISPs can generally only suggest which ingress peering points to use and the final decisions are still made by their customers [6], we assume in this paper that the ingress point of *any* traffic is fixed and known in advance. Thus, we define the following types of Internet traffic and refer to them throughout this paper:

- *local intra-AS traffic*: traffic that is destined to egress access links. This corresponds to internal and inbound traffic.
- *inter-AS traffic*: traffic that is destined at downstream ASes and whose egress peering points can be varied by inter-AS TE. This corresponds to outbound and transit traffic.

- *intra-AS traffic*: all traffic that traverses the network, including both local intra-AS and inter-AS traffic.

B. Traffic Matrices

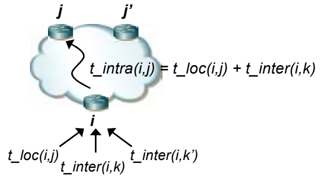

Figure 1. Example of intra-AS traffic matrix

A Traffic Matrix (TM) represents a matrix of traffic load from one network point to another one over a particular time interval. In the **inter-AS traffic matrix**, each element *t_inter(i,k)* represents the volume of inter-AS traffic that enters the network at ingress point *i* and is destined to routing prefix *k*. The egress point of each inter-AS traffic flow can be selected by inter-AS TE. We denote *s(i,k)* as the egress point assigned to *t_inter(i,k)*. On the other hand, in the **intra-AS traffic matrix**, each element *t_intra(i,j)* represents the volume of traffic that enters the network at ingress point *i* and exits at egress point *j*. It is the sum of the local intra-AS and inter-AS traffic volume between each ingress and egress node pair:

$$t\_intra(i,j) = t\_loc(i,j) + \sum_{k \in K: s(i,k)=j} t\_inter(i,k)$$

Figure 1 shows a network with ingress point *i* and egress points *j* and *j'*. We assume both egress points can reach routing prefixes *k* and *k'*. Given local intra-AS traffic flow *t_loc(i,j)* and inter-AS traffic flows *t_inter(i,k)* and *t_inter(i,k')* with *s(i,k) = j* and *s(i,k') = j'*, the elements of the intra-AS traffic matrix are *t_intra(i,j) = t_loc(i,j) + t_inter(i,k)* and *t_intra(i,j') = t_inter(i,k')*.

The intra- and inter-AS TM can be obtained through measurement or estimation. The intra-AS TM can be measured [5] through Cisco's NetFlow. It can also be estimated from measured link load statistics [7,8]. In contrast to the intra-AS TM, less attention has been paid to deriving the inter-AS TM. One method of deriving this is through measurement at each vantage point. The authors in [9] describe a methodology to compute inter-AS TM using Cisco's NetFlow and BGP routing data. In addition to measurement, the inter-AS TM may be estimated. The authors in [10] propose a methodology to estimate inter-AS publisher and web traffic matrices using server logs from content delivery networks and packet level traces from large user sets. These estimated traffic matrices are routing prefix basis, which can be used as the inter-AS TM in this paper.

C. Intra- and Inter-AS Traffic Engineering

A general intra-AS TE problem can be summarized as follows: given a network topology and an intra-AS traffic matrix, determine an appropriate set of OSPF link weights or MPLS LSPs so as to optimize the network performance such as bandwidth consumption and load balancing within the network. Many techniques exist to solve this problem. For example, Fortz and Thorup [11] proposed a tabu search technique to derive optimal link weight settings. Xiao et al [12] proposed a MPLS-based TE heuristic. More intra-AS TE work can be found in [26,28] and their cited references.

On the other hand, the general inter-AS TE problem [6] can be summarized as follows: given a network topology, an inter-AS traffic matrix and BGP routing prefixes, select ingress/egress points for the traffic so that the network performance is optimized. A common inter-AS TE objective is to balance the load over inter-AS links. A number of outbound inter-AS TE algorithms have been proposed recently [6,16,17]. In addition, a number of inbound inter-AS TE algorithms [18,19] have been proposed very recently based on AS-path prepending [21]. This is a commonly used technique that purposely makes a route less attractive to its upstream ASes by adding several instances of its own AS-Number to the AS-path attribute so as to increase the AS-path length of that route. In current practice, inter-AS TE is enforced by adjusting BGP route attributes such as *local-preference, AS path length*, etc. The reader is referred to [21] for the explanation of these techniques. In addition, ongoing work in inter-AS MPLS [22] provides an alternative method to enforce inter-AS TE.

III. Interactions between intra- and inter-AS TE

The existing literature considers intra- and inter-AS TE independently. However, the overall network performance may not be truly optimized when they are considered separately. In this section, we explain with examples two interaction effects between intra- and inter-AS TE which would lead to suboptimal overall network performance. Although, as mentioned in the introduction, this paper focuses on MPLS-based intra-AS TE and outbound inter-AS TE, we present the two interaction effects in a general way that is applicable to other intra- and inter-AS TE approaches.

A. Effect of Inter-AS TE on Intra-AS TE

The first interaction is the effect of inter-AS TE on the performance of intra-AS TE. We take outbound inter-AS TE as an example in our discussion although the effect is also applicable to inbound inter-AS TE as both can influence intra-AS TE performance by changing the intra-AS traffic matrix.

Recall from Section II.B that the intra-AS TM is derived from both local intra-AS and inter-AS traffic. Inter-AS TE assigns egress points to inter-AS traffic in order to balance the load over inter-AS links, for example. Consequently, the traffic will be routed on different ingress-to-egress paths according to the assigned egress points. This results in changing the intra-AS TM, with varying traffic load for each ingress-egress node pair. For example, referring to the scenario in Section II.B, if *s(i,k)* was changed to *j'*, the two elements of intra-AS traffic matrix would then become *t_intra(i,j) = t_loc(i,j)* and *t_intra(i,j') = t_inter(i,k) + t_inter(i,k')*. However, a change in intra-AS TM has a consequence. There are two cases where if:

- inter-AS TE is performed *before* intra-AS TE: by performing inter-AS TE, different egress point selections for the inter-AS traffic will result in different intra-AS TMs because the traffic load between ingress and egress node pairs is varied. However, different intra-AS TMs cause intra-AS TE to produce solutions with different performances, even though identical network topology and TE algorithm are used. This largely depends on the topological connectivity and characteristics, e.g. link capacity. In this case, a suboptimal network performance results if the intra-AS TM does not lead the intra-AS TE to produce a globally optimal performance.

- inter-AS TE is performed *after* intra-AS TE: when inter-AS TE is run, inter-AS traffic is shifted to use different ingress-to-egress paths, causing increased intra-AS link utilization and possibly congestion. This worsens the performance initially achieved by the intra-AS TE.

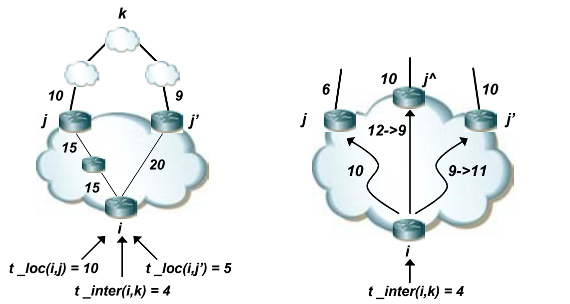

Figure 2(a) and (b). Interaction effects between intra- and inter-AS TE

We illustrate the first interaction effect through the following example of Figure 2(a). We denote inter-AS costs[1] at egress point *j* and *j'* by *Cost(j)* and *Cost(j')*, and intra-AS costs[1] of route *r(i,j)* (the best route that could be found by intra-AS TE between ingress point *i* and egress point *j*) and *r(i,j')* by *Cost(i,j)* and *Cost(i,j')* respectively. These routes are assumed to be the ultimate intra-AS TE solution, regardless of whether the inter-AS TE performs before or after the intra-AS TE; this corresponds to the abovementioned two consequences. Assume there are two local intra-AS traffic flows *t_loc(i,j)* and *t_loc(i,j')*, and one inter-AS traffic flow *t_inter(i,k)*. Their traffic loads are 10, 5 and 4 (unitless) respectively as shown in the figure. The number besides each link is the link capacity. We define the cost[1] of each link to be the link's total traffic divided by its capacity.

For inter-AS TE, there are two possible egress points where *t_inter(i,k)* can be assigned. If it is assigned to *j*, *Cost(j)* becomes 4 / 10 = 0.4 and the intra-AS TM becomes: *t_intra(i,j)* = *t_loc(i,j)* + *t_inter(i,k)* = 14 and *t_intra(i,j')* = *t_loc(i,j')* = 5. In this case, *Cost(i,j)* and *Cost(i,j')* are 1.866 (0.933 x 2 hops) and 0.25 respectively. Hence, the overall TE cost is 0.4 + 1.866 + 0.25 = 2.516. On the other hand, if the traffic flow is assigned to *j'*, *Cost(j')* becomes 0.44 (4 / 9) and the intra-AS TM becomes: *t_intra(i,j)* = *t_loc(i,j)* = 10 and *t_intra(i,j')* = *t_loc(i,j')* + *t_inter(i,k)* = 9. In this case, *Cost(i,j)* and *Cost(i,j')* are 1.33 and 0.45 respectively. Thus, the overall TE cost is 0.44 + 1.33 + 0.45 = 2.22.

For inter-AS TE, egress point *j* is chosen primarily because of the lowest resulting inter-AS cost. However, as can be seen, the lowest overall TE cost solution is to assign the traffic flow to egress point *j'* as the intra-AS cost can now be significantly reduced with only a slightly increase in the inter-AS cost. Even if the resulting inter-AS cost of both assignments are the same, the first one may still be a result since intra-AS route optimization is not considered during inter-AS TE. In fact, there are many issues influencing the performance of intra-AS TE. An important issue which has been shown in the example is the amount of traffic routed on each ingress-to-egress path. Varying the traffic load between ingress and egress node pairs caused by inter-AS TE results in different traffic demands in the network. If a large traffic demand can only be routed on a low capacity link or a longer path, a high intra-AS TE cost usually results, for example *t_intra(i,j)* and *Cost(i,j)* in the scenario where *t_inter(i,k)* is assigned to egress point *j*.

Regardless of whether the routes have been optimized or are to be optimized by intra-AS TE, shifting the traffic egress points by inter-AS TE yields different intra-AS TE solutions. We therefore propose to operate on local intra-AS TM and inter-AS TM collectively in order to find a 'good' intra-AS TM so as to improve the overall network performance, rather than merely assuming that an intra-AS TM is given.

B. Effect of Intra-AS TE on Inter-AS TE

The second interaction effect between intra- and inter-AS TE is caused by hot-potato routing. We explain this effect using the example in [2] illustrated by Figure 2(b). The definition of each element in the figure is the same as those in Figure 2(a), except that the numbers on the three arrows towards the three egress points are the IGP costs of *r(i,j)*, $r(i,\hat{j})$ and *r(i,j')*. Assume that ingress point *i* has three equally good routes to routing prefix *k* through egress points *j*, $\hat{j}$ and *j'*. Under hot-potato routing, *i* directs traffic to the closest egress point - the router with the smallest IGP cost (i.e., router *j'*). As a result, *Cost(j)*, $Cost(\hat{j})$ and *Cost(j')* are equal to (0.0, 0.0, 0.4) respectively. However, suppose the IGP cost to *j'* changes from 9 to 11, in response to a change in link weights for a TE purpose. Although the route through *j'* is still available, the IGP cost change would cause *i* to select the route through egress point *j* since this has the lowest IGP cost. In this case, the inter-AS costs are changed to (0.66, 0.0, 0.0). Consequently, although the intra-AS TE cost could be improved, the change of IGP costs affects the utilization on some inter-AS links and possibly causes congestion. In this case, intra-AS TE can not only affect the performance of intra-AS but also the inter-AS performance. Note that this effect would not have an influence on MPLS-based intra-AS TE since the traffic forwarding decision does not depend on IGP costs but LSPs and labels. Nevertheless, irrespective of which intra-AS TE approach is used, the performance of intra-AS TE may not be optimized without reassigning some inter-AS traffic flows to other ingress/egress points that could eventually lead the traffic to be routed on better-performed paths in the network.

In order to reduce the high inter-AS cost in the above example, the status of both intra- and inter-AS link utilization need to be simultaneously considered. For example, one could change the IGP cost of $r(i,\hat{j})$ from 12 to 9 if this change does not affect the overall intra-AS cost while still meeting the TE objectives. As a result, the traffic flow will be assigned to egress point $\hat{j}$ and the inter-AS costs now become (0.0, 0.4, 0.0). This produces the same inter-AS cost as the original case where IGP costs have not changed.

1 The cost here is simply a performance metric that approximates the link utilization. It is *not* the IGP cost or weight that is used by OSPF routing.

C. The Need for Joint Optimization

The two interaction effects have shown that inter-AS TE indeed can affect the performance of intra-AS TE and vice versa. A recent study has also shown that the Internet bottleneck is approximately equally distributed on intra- and inter-AS links [13]. Therefore, a good TE solution should perform satisfactorily with respect to both intra- and inter-AS TE objectives. It is important that inter-AS TE aspects should be considered during intra-AS TE and vice versa. For example, if we look for the best overall TE result, we can see that egress point *j*' and the suggested solution of the preceding two examples would be chosen if intra- and inter-AS TE can be jointly optimized. In the next section, we therefore propose a bi-criteria *integer* programming formulation for the Joint intra- and inter-AS TE optimization (Joint-TE) problem.

IV. Joint Intra- and Inter-AS TE Optimization

A bi-criteria optimization formulation is that one can express two notions that are of concern in defining what represents an optimal solution. Their objectives are typically expressed in a form of cost functions. In this paper, we formulate a bi-criteria Joint-TE problem by taking into account both intra- and inter-AS TE cost functions.

A. Cost functions

We employ the commonly used cost function proposed by Fortz and Thorup [11]. This is a piecewise linear function of link utilization, which imitates the response time of *M/M/1* queues to represent the cost of network links. By using the piecewise linear cost function, two objectives of minimizing bandwidth consumption and achieving load balancing are taken into account simultaneously. In this paper, we use the piecewise linear cost function for both intra- and inter-AS TE for consistency and generality, as well as for its ability to express the common key objectives of each TE. Nevertheless, these cost functions may be different by domains according to their operational objectives.

B. Bi-Criteria Problem Formulation

The objective of the Joint-TE problem is to minimize both overall intra- and inter-AS costs:

$$Minimize \sum_{l \in E_{intra}} \Upsilon_l + \sum_{j \in J} \Psi_j \quad (1)$$

Subject to

$$u^l_{intra} = \sum_{i \in I}\sum_{j \in J} y^l_{i,j} \cdot t\_intra(i,j) \Big/ C^l_{intra} \ , \ \forall l \in E_{intra} \quad (2)$$

$$u^j_{inter} = \sum_{i \in I}\sum_{k \in K} x^j_{i,k} \cdot t\_inter(i,k) \Big/ C^j_{inter} \ , \forall j \in J \quad (3)$$

$$\sum_{p \in P_{i,j}} w_{i,j,p} = 1 \ , \ \forall (i \in I, j \in J) \quad (4)$$

$$\Upsilon_l = f(u^l_{intra}) \quad (5)$$

$$\Psi_j = f(u^j_{inter}) \quad (6)$$

$$f(\theta) = \theta \ , \ 0 \le \theta \le 1/3 \quad (7)$$

$$f(\theta) = 3\theta - \frac{2}{3} \ , \ 1/3 \le \theta \le 2/3 \quad (8)$$

$$f(\theta) = 10\theta - \frac{16}{3} \ , \ 2/3 \le \theta \le 9/10 \quad (9)$$

$$f(\theta) = 70\theta - \frac{178}{3} \ , \ 9/10 \le \theta \le 1 \quad (10)$$

$$f(\theta) = 500\theta - \frac{1468}{3} \ , \ 1 \le \theta \le 11/10 \quad (11)$$

$$f(\theta) = 5000\theta - \frac{16318}{3} \ , \ 11/10 \le \theta \le \infty \quad (12)$$

Equations (2) and (3) define the utilization of intra- and inter-AS links. Constraint (4) ensures that each intra-AS traffic flow *t_intra(i,j)* is routed along a single LSP within the network. This is an optional constraint but we consider preserving scalability and minimizing complexity on network management by avoiding excessive LSPs to be managed and arbitrary traffic splitting, though such splitting could permit better network performance [23]. The remaining constraints define the cost of each intra- and inter-AS link as a function of its utilization based on the piecewise linear cost function [11]. Given the lossless property of the links, a general constraint is a flow conservation constraint which ensures that the total traffic demand incoming to an intermediate node is equal to the total traffic for this demand outgoing from the node.

Fundamentally, the Joint-TE problem is the combination of intra- and inter-AS TE problems. The problem formulation of intra-AS TE consists of the reduced objective function $Minimize \sum_{l \in E_{intra}} \Upsilon_l$ as well as constraints (2), (4), (5) and (7)-(12). On the other hand, the problem formulation of inter-AS TE consists of the reduced objective function $Minimize \sum_{j \in J} \Psi_j$ as well as constraints (3) and (6)-(12).

C. Optimization Criteria

An optimal solution of the bi-criteria optimization problem is that each of the two TE objectives simultaneously attains an optimal value. However, in general, either it is not possible to find such optimal solutions or they do not exist. In other words, the two objective functions are conflicting. For example, the cost of a particular egress point may be low but the cost of the intra-AS path towards the egress point is high. Moreover, it is not possible to compare the two objective values mathematically and sensibly. For example, we cannot distinguish mathematically which is a better TE solution, (10,30) or (30,10), where (x,y) represents intra- and inter-AS costs respectively. However, we can observe that (10,30) is better than (20,30) or (20,40). On the other hand, the value of cost function varies with the number of links and their utilizations. It does not make sense to compare the two objective values when the number of intra- and inter-AS links and their capacities are different, as is typically the case. Consequently, this leads us to finding non-dominated solutions, which is a primary goal when solving a multi-criteria optimization problem. A solution is called *non-dominated* if there is no any other solution that is strictly better in one of the objective functions, and has the same or better values in the others [24]. Thus, solution (10,30) in the above example is a non-dominated solution of (20,30) and (20,40).

There are multiple ways to identify non-dominated solutions. A commonly used method is to design a metric or cost function that combines both intra- and inter-AS TE objectives. However, it is often unclear how to determine the relative weights between the two objectives. A more intuitive approach, which we consider in this paper, is to search non-dominated solutions in such a way that the inter-AS cost remains at least near-optimal while substantially improving the intra-AS cost. In other words, inter-AS TE is assumed to be more important than intra-AS TE. The rationale for this assumption is the following:

- intra-AS overprovisioning has been employed by ISPs as an effective means to provide high-quality service to all traffic on their IP backbone networks [20].
- inter-AS links are common points of congestion in the Internet [6,13]. For example, peer-to-peer traffic consumes the major part of inter-AS link bandwidth [25]. Moreover, an inter-AS link is relatively more difficult to extend than an intra-AS link due to time-consuming and complicated negotiation between two ASes. Therefore, the ASes need to control traffic especially on inter-AS links.

The method of predefining a lexicographic importance order is commonly used in solving multi-criteria optimization problems. It allows us to generate Joint-TE solutions that can be mathematically distinguished and sensibly compared.

D. TE Algorithm Selection – A Black-Box Approach

Intra- and inter-AS TE algorithms will be used to solve the Joint-TE problem. In this paper, we deliberately treat both intra- and inter-AS TE as black-boxes that we combine in a plug-and-play manner. Both sides use TE algorithms that are based on previously established techniques and can achieve near-optimal solutions. The intention of using a TE algorithm that produces near-optimal solutions is to minimize the possibility that any performance improvement is solely caused by a large performance gap between the optimal solution and the solution achieved by the algorithm.

We use the optimal aware heuristic proposed by Sridharan et al in [26] as the intra-AS TE algorithm. The algorithm solves a MPLS-based intra-AS TE problem with the piece-wise linear cost function, which is the intra-AS TE problem considered in this paper. The basic idea of the algorithm is that Linear Programming (LP) formulation of the intra-AS TE problem is solved to obtain an optimal routing solution. This optimal solution permits arbitrary traffic splitting but it is not the solution required by the intra-AS TE that does not allow such splitting. Therefore, a greedy heuristic based on traffic demand sorting is then performed to transform the optimal routing solution to the traffic non-splittable solution while attempting to maintain its optimality. We call this intra-AS TE algorithm ***intra-optimal-aware-alg*** throughout this paper. The reader is referred to [26] for more details of the algorithm.

The idea of the optimal aware heuristic algorithm has also been used by prior TE work [27,28]. Their simulation results showed that the algorithm can achieve near-optimal solutions. Hence, we also propose using it to solve the inter-AS TE problem. We call it ***inter-optimal-aware-alg*** throughout this paper. The algorithm works as follows:

1. A LP of the inter-AS TE problem is solved. The outcomes are an optimal inter-AS cost and the desired utilization of each inter-AS link that achieves this optimal cost. This solution permits arbitrary splitting of inter-AS traffic over multiple egress points.
2. Set the desired inter-AS link utilization as capacity constraints. This constraint ensures that the total traffic on each inter-AS link does not exceed the desired utilization.
3. Sort inter-AS traffic flows in descending order according to traffic demand. Assign each traffic flow in that order to the egress point (i.e. inter-AS link) that has the lowest utilization while not violating its capacity constraint[2].
4. If there exists unassigned inter-AS traffic flows, remove the capacity constraint on all the inter-AS links and re-run step 3 until all the traffic flows have been assigned.

The LP formulations (i.e. the first step) of *intra-optimal-aware-alg* and *inter-optimal-aware-alg* were modeled in AMPL and solved with CPLEX optimization engine[3]. To enable us to show that these two algorithms produce near-optimal solutions, we include their optimal solutions (i.e. by only solving the LP of the corresponding TE problem) in our experiments for comparison.

V. Strategies for Joint-TE

Based on the problem formulation, assumptions and algorithms described in the preceding section, we present three strategies to solve the Joint-TE problem, namely sequential, nested and integrated optimization.

A. Sequential Optimization

Sequential optimization solves the Joint-TE problem *sequentially* where the optimal solution of one TE problem becomes the input for the other one. It can be regarded as the current practice in the collective use of intra- and inter-AS TE. As with the two interaction effects in Section III, sequential optimization can be divided into two approaches with different execution sequences in the TE problems:

1) SeqOpt-Inter-Intra: egress points are first assigned to inter-AS traffic using *inter-optimal-aware-alg* so as to optimize the inter-AS cost. Then the intra-AS traffic matrix is computed by taking into account the local intra-AS TM and the inter-AS TM with the assigned egress points. Finally, intra-AS TE is performed using *intra-optimal-aware-alg* to optimize the intra-AS cost. This strategy is logical in the sense that intra-AS TE is not performed until both ingress and egress points of the inter-AS traffic have been determined.

2) SeqOpt-Intra-Inter: initially, intra-AS TE is first performed using *intra-optimal-aware-alg* to optimize the intra-AS cost based on the intra-AS TM that is computed by combining the local intra-AS TM with the inter-AS TM that has the egress points randomly assigned. This random egress point assignment mimics the situation where inter-AS TE has not been performed, thus the inter-AS cost is not optimized. At this stage, the outcome of intra-AS TE is a set of ingress-to-egress LSPs. The next step is to perform inter-AS TE using *inter-optimal-aware-alg* to optimize the inter-AS cost. Since inter-AS TE does not consider route optimization within the network, each inter-AS traffic flow is eventually routed on the LSP between the ingress and the assigned egress point, which has been determined by the intra-AS TE. *SeqOpt-Intra-Inter*

---

[2] If there are several such egress points, the selection tie-break is in order of maximum residual capacity and visited sequence.

[3] Information about AMPL and CPLEX can be found at www.ampl.com and www.ilog.com respectively.

takes the TE approach that inter-AS TE is performed on top of an intra-AS traffic engineered network.

The advantage of sequential optimization is that different analysis techniques can be applied to each of the TE problems. However, neither of the above sequential optimization approaches consider intra-AS route optimization during inter-AS TE nor inter-AS egress point selection during intra-AS TE.

B. Nested Optimization

Sequential optimization generates only one solution from each TE algorithm, and this is used as input to the other algorithm. However, the initial solution may not be a good input to the subsequent TE problem. In fact, for many optimization problems, there exists more than one optimal solution. Hence, there may exist inter-AS TE solutions that are nearly optimal and are also good with respect to intra-AS TE. Thus, we seek for those optimal solutions from a TE problem and then input them to the subsequent TE problem one at a time until the solution with the best overall cost is found.

We propose a nested optimization to implement the abovementioned idea. It can be regarded as an enhanced and *iterative* version of sequential optimization. The algorithm proposed for the nested optimization in a similar sequence to *SeqOpt-Inter-Intra*:

1. A Genetic Algorithm (GA) is used to identify a set of lowest-cost inter-AS TE solutions. When the GA converges, all chromosomes are the solutions that have nearly identical lowest cost but they may have different results. In fact, in order to explore a larger solution searching space for intra-AS TE, these solutions should not be restricted to an identical cost. In this paper, the solution with inter-AS cost not exceeding 0.001% of the visited lowest cost is considered.
2. For each of this set of inter-AS TE solutions, we perform intra-AS TM computation and then intra-AS TE using *intra-optimal-aware-alg*. During intra-AS TE optimization, the best and the worst visited solutions are recorded. We call them *NestedBest* and *NestedWorst*. The *NestedBest* and *NestedWorst* solutions reflect respectively the extent to which the sequential optimization solution can be further optimized and how worse could be the solution.

We modified our GA [17] previously proposed to solve the inter-AS TE problem. The GA has included a heuristic similar to step 3 of *inter-optimal-aware-alg* to enhance the quality of the solution. In fact, the number of candidate inter-AS TE solutions may be quite large and we did identify a large number of such solutions as alternative inputs for the intra-AS TE. We observed that many of those inter-AS TE solutions have significantly different egress point selection results, which produced intra-AS TE solutions with very different performances. The nested optimization maintains a simply interdisciplinary partitioning used in sequential optimization while attempting to obtain better overall TE solutions.

C. Integrated Optimization

Integrated optimization aims to solve the Joint-TE problem by *simultaneously* optimizing the intra- and inter-AS TE objectives. We propose an integrated approach that requires as starting solutions an inter-AS and an intra-AS routing configurations with known egress points and ingress-to-egress paths. The starting solutions can be any quality, regardless of whether they are optimized by TE or not. The integrated approach then proceeds to enhance the quality of the starting solution using neighborhood search algorithm. The integrated approach guarantees that the produced solutions are no worse than the input solutions and in practice are much better.

Algorithm NSA

1. Obtain the starting solutions
2. *iter* ← 0
3. **While** *iter* < *MAX_ITER* /* stopping criteria */
4. *iter* ← *iter* + 1
5. **If** no significant cost improvement for a certain number of iterations **then**
6. perform intra-AS TE on the current solution /* diversification */
7. **For** each inter-AS traffic flow *t_inter(i,k)*
8. *f(i,k)* ← *s(i,k)*
9. $\Phi \leftarrow \alpha\Psi_{s(i,k)}(bw_{inter}^{s(i,k)}) + \sum_{l \in p \in P_{i,j}: w_{i,j,p}=1} \Upsilon_l(bw_{intra}^{l})$
   $\Phi' \leftarrow \alpha\Psi_{s(i,k)}(bw_{inter}^{s(i,k)} + t\_inter(i,k)) + \sum_{l \in p \in P_{i,j}: w_{i,j,p}=1} \Upsilon_l(bw_{intra}^{l} + t\_inter(i,k))$
10. $\Delta \leftarrow \Phi - \Phi'$ /* delta is the saved cost */
11. **For** each $j \in Out(k)$ which does not constitute a move in the memory list
12. $\Omega \leftarrow \alpha\Psi_j(bw_{inter}^{j}) + \sum_{l \in p \in P_{i,j}: w_{i,j,p}=1} \Upsilon_l(bw_{intra}^{l})$
13. virtually add *t_inter(i,k)* to *t_intra(i,j)*
14. virtually release the resource used by *t_inter(i,k)* on $p \in P_{i,j} : w_{i,j,p} = 1$
15. re-compute intra-AS path *z* between ingress point *i* and egress point *j*
16. $\Omega' \leftarrow \alpha\Psi_j(bw_{inter}^{j} - t\_inter(i,k)) + \sum_{l \in z} \Upsilon_l(bw_{intra}^{l})$
17. **If** $\Delta \geq \Omega' - \Omega$ **then** /* if saved cost >= added cost */
18. $\Delta \leftarrow \Omega' - \Omega$ /* lines 17-20 aim to find the minimum added
19. *f(i,k)* ← *j* cost, i.e. to find the maximum profit cost */
20. *sel_path* ← *z*
21. **If** *f(i,k)*≠*s(i,k)*
22. break the most outer **For** loop /* Implementation of FM */
23. **If** *f(i,k)*≠*s(i,k)*
24. update resource utilization on intra- and inter-AS link with respect to the new assignment
25. replace $p \in P_{i,s(i,k)} : w_{i,s(i,k),p} = 1$ by *sel_path*
26. add *t_inter(i,k)* to *t_intra(i,f(i,k))*
27. *s(i,k)* ← *f(i,k)*

Remarks:
$\Psi_j(\Phi)$Represents the cost of inter-AS link *j* with residual bandwidth Φ
$\gamma_l(\Phi)$Represents the cost of intra-AS link *l* with residual bandwidth Φ

Figure 3. Neighborhood search algorithm

1) Neighborhood Search Algorithm

a) Overview

Neighborhood Search Algorithm (NSA) is widely regarded as an important tool to solve hard combinatorial optimization problems efficiently. The primary reasons for the widespread use of neighborhood search techniques in practice are their intuitive appeal, flexibility and ease of implementation, and their excellent empirical results [29].

The basic steps of NSA are as follows. Consider a current starting solution *x*. NSA explores the solution space by identifying the neighborhood of *x*, *N(x)*. The neighbors of *x* are solutions that can be obtained by applying a single local transformation (also called a *move*) on *x*. The best solution in the neighborhood is selected as the new current solution. This neighborhood searching iterates until the stopping criteria is satisfied. Finally the algorithm returns the best visited solution.

During neighborhood search, NSA can move the current solution to the best neighbor that either improves or worsens the quality of the solution. To avoid cycling, a specially designed memory list is used to store previously visited solutions or certain attributes of them for a certain number of iterations. A neighbor solution is rejected if it is already in the list. In order to make the neighborhood searching more effective, an intensification or diversification technique is used to force the algorithm to explore parts of the solution space that has not been searched yet. Our NSA is outlined in Figure 3 and its fundamental components are described as follows.

b) Non-TE Starting Solutions

Starting solutions for inter- and intra-AS routing can be respectively obtained by randomly assigning egress points to the inter-AS traffic and then routing each intra-AS traffic flow on the shortest hop paths. They are regarded as non-TE solutions. Nevertheless, we will also evaluate the impact of using TE optimized starting solutions on network performance.

c) Neighborhood Structure and Search Strategy

We consider a neighborhood structure that is based on shifting inter-AS traffic to different egress points while at the same time rerouting the corresponding ingress-to-egress paths. Details of this local transformation are as follows:

We define path cost to be sum of the cost of the inter-AS link and the cost of each link on the intra-AS route to which an inter-AS traffic flow has been assigned. In order to place more importance on optimizing inter-AS cost relative to the intra-AS cost, we introduce α as a factor with large value to scale the inter-AS cost. Note that α itself has no particular meaning to the Joint-TE problem. It is an intermediary to identify all the non-dominated combinations of intra- and inter-AS costs.

For each inter-AS traffic flow, calculate the profit cost, i.e. the saved cost minus the added cost. The *saved cost* (line 10 in Figure 3) is the path cost of the traffic flow minus the path cost of the traffic flow that would have been removed. This saved cost reflects how much cost would have been reduced if the traffic flow was removed on the path. The *added cost* (line 17) is calculated for each potential egress point except for the one that is currently assigned to the inter-AS traffic flow. It is the cost of a new path towards the potential egress point that the traffic flow would have been assigned minus the path cost of the original path towards this egress point. The new path is the result of rerouting the original path taking into account the traffic flow (lines 13-15) and it is a minimum cost path that can be found by Dijkstra's algorithm using the instantaneous intra-AS link cost as the routing metric. Consequently, the added cost reflects how much cost would have been increased when the traffic flow is assigned onto a new egress point.

The neighborhood search strategy specifies which solution in the neighborhood is chosen at each iteration. The following two methods are commonly used:

- *Best Method (BM)*: Compute the profit cost for each inter-AS traffic flow. Choose the one yielding the solution with the *highest* profit cost as the next move.
- *First Method (FM)*: Compute the profit cost for each inter-AS traffic flow. Choose the *first* one yielding a solution with positive profit cost.

It is of great importance for the solution quality and the search efficiency. We have found in our experiments that BM can achieve approximately 5%-10% performance improvement over FM, but the computational complexity of BM is several orders of magnitudes higher than the FM, which makes it impractical to use. We therefore decided to use FM in our NSA. Our finding is consistent with the prior work that has evaluated the tradeoff between quality and efficiency of BM and FM [30].

d) Use of Memory List

The memory list is operated as a first-in-first-out queue. The first element in the list is removed and then a new solution is pushed into the tail of the list. As suggested in [29], the size of the list depends on the size and the characteristics of the problem. We define the size of the list to be a large value (100) in order to avoid looping. This number does not significantly affect the performance that can be achieved by the NSA because the number of potential traffic-to-egress-point assignments that are not in the memory list is still very large.

e) Diversification

We notice that if the NSA works on the initial solution for many iterations, this may lead the algorithm get stuck in a local optima. To overcome this, a diversification is needed. If there is no obvious improvement in the solution for a certain number of iterations, we modify the current solution by re-running intra-AS TE on it. We define the threshold of obvious improvement to be 10% and the number of iterations to be 500.

f) Stopping Criterion

Many stopping criteria can be developed depending on the nature of the problem being studied. The most common criterion, which is employed in this paper, is a maximum number of iterations (*MAX_ITER*). However, we do not arbitrarily select this number since the performance of the NSA is mainly dependent on how many times inter-AS traffic flows can be reassigned. Therefore, the maximum iteration number should be related to the number of inter-AS traffic flow. In our experiments, we found that setting the maximum iteration number to be 4 times the number of inter-AS traffic flows gives us a sufficiently good result.

g) Neighborhood search algorithm complexity

The worst-case time complexity of the NSA is analyzed. We denote n and e the number of inter-AS traffic flows and the number of egress points. The NSA calculates the profit cost for each inter-AS traffic flow by evaluating each potential egress point (lines 11-20). The most time consuming step in this block is to find a new minimum cost route by using the Dijkstra's algorithm (line 15). The time complexity of Fibonacci-heap implementation of the algorithm is $O(|V|log|V|+|E|)$ where $|V|$ and $|E|$ are the number of nodes and edges in the network [31]. Since the worst-case is to examine all the inter-AS traffic flows until the first positive profit cost solution is found, the whole step (lines 7-20) could take $O(n \cdot e \cdot (|V|log|V|+|E|))$ time. The NSA then iterates until the maximum iteration number (*MAX_ITER*) has been reached. Therefore, the overall worst-case time complexity of the NSA can be summarized as $O(MAX\_ITER \cdot n \cdot e \cdot (|V|log|V|+|E|))$.

VI. EVALUATION METHODOLOGY

In this section, we present our methodology to evaluate the performances of different proposed Joint-TE strategies.

A. Network Topology

We use the Rocketfuel [4] Point-of-Presence (POP) level maps published by the authors, shown in Table 2, as our topologies. For each topology, POPs correspond to cities. Some POPs have inter-AS links connected to other ASes and we call them border POPs. Without loss of generality and for simplicity, we assume that each border POP is associated with a virtual inter-AS link that is the abstraction of one or multiple physical inter-AS links. Since the Rocketfuel data do not contain link bandwidth, we set the intra- and inter-AS links to be OC-48/STM-16 and OC-12/STM-4 respectively.

Table 2. Rocketfuel topologies used in evaluation

AS	Name	# POP nodes	# Intra-AS links	# Border POP nodes
1239	Sprint (US)	52	84	40
7018	AT&T (US)	114	148	41
6461	Abovenet (US)	19	34	14
5400	BT Backbone (Europe)	29	45	16

B. Internet Routing prefixes

For scalability and stability, inter-AS TE can focus on only a small fraction of routing prefixes which are responsible for a large fraction of the traffic [14]. In this paper, we consider 200 such popular routing prefixes. Nevertheless, each of them may not merely represent an individual prefix but also a group of distinct routing prefixes that have the same set of candidate egress points [5] in order to improve network and TE algorithm scalability. Hence, the number of routing prefixes we consider could actually represent even a larger value.

Each border PoP can be an ingress or egress point. In order to evaluate the effect of inter-AS TE on the performance of intra-AS TE, we consider the situation where if a border POP receives a route advertisement towards routing prefix k from adjacent AS Y, then AS Y cannot inject traffic for k into it. This corresponds to multi-hop traffic [5] in which the traffic traverses across the network instead of being directed to another egress link of the same border POP. As a result, we cannot assign all the routing prefixes on each border POP as route advertisements. Instead, we consider half of the routing prefixes are randomly selected as route advertisements and the other half as inter-AS traffic in each border POP. We note that this routing prefix generation process is just a best effort attempt to model prefix distribution, as no synthetic model for the actual behavior of real networks was found in the literature.

C. Traffic Matrices

We generate synthetic traffic matrices for our evaluation. We generate inter-AS traffic from each POP towards each of the considered routing prefixes. Note that if the POP is a border POP, it can only inject traffic headed towards the routing prefixes that have not been selected as route advertisements. Previous work has shown that intra- and inter-AS traffic are not uniformly distributed [32,35]. According to [33], AS traffic volumes are top-heavy and can be approximated by Weibull distribution with shape parameter 0.2-0.3. We therefore generate the inter-AS TM with traffic demand using this distribution with the shape parameter 0.2.

We use the Gravity Model (GM) outlined in [34] to generate the local intra-AS TM. The GM approach was proposed based on the findings in [8]. Following the suggestions in [35], we randomly classify 40% of POPs as “small”, 40% as “medium” and 20% as “big”. The amount of incoming traffic at a POP is proportional to its size.

D. Algorithm Parameters

For the GA in the nested optimization, we use the suggested values from previous GA research to achieve satisfactory effectiveness and convergence rate of the algorithm [36]. The population size is 200 and the probability of mutation is 0.01. We set the GA to produce maximum 200 distinct inter-AS TE solutions to compute *NestedBest* and *NestedWorst*.

E. Performance Metrics

The following performance metrics are used to evaluate the Joint-TE strategies. For these metrics, lower values are better than high values.

- **Overall Intra- and Inter-AS cost**: these metrics capture the overall network cost of the objective function (1).
- **Total bandwidth consumption**: the amount of bandwidth needed to accommodate all traffic flows within the network, being the sum of the traffic loads over all the intra-AS links.
- **Maximum intra- and inter-AS link utilization**: the maximum intra-AS (inter-AS) link utilization is the maximum utilization on all the intra-AS (inter-AS) links in a network. Minimizing this value ensures that traffic is moved away from congested to less utilized links and is balanced over the links.

VII. SIMULATION RESULTS

All the results presented in this paper are an average of 30 trials. For the largest network topology, the simulation took approximately two hours in average.

A. Evaluation of Overall Inter-AS TE Cost

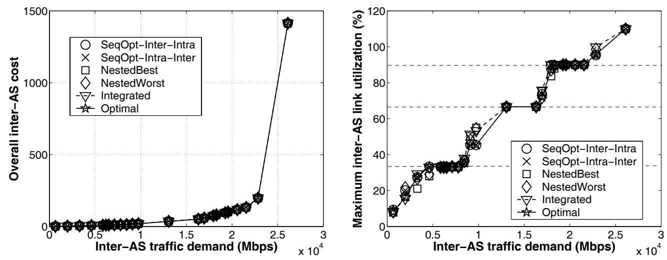

Figure 4(a) and (b). Overall Inter-AS cost and maximum inter-AS link utilization (Sprint topology)

We have evaluated the overall inter-AS TE costs achieved by all the strategies for all the network topologies. We found that their results exhibit a common characteristic – their overall inter-AS TE costs are nearly identical and very close to the optimal solution that is obtained by solving the LP of the inter-AS TE problem. We therefore arbitrarily present and analyze the results of one network topology for brevity.

Figure 4(a) shows the overall inter-AS cost (y-axis) achieved by all the strategies as a function of the inter-AS traffic demand (x-axis) for the Sprint topology. The curve denoted by *Optimal* is the optimal solution of the inter-AS TE problem. The shapes of the result curves follow the piecewise linear cost function. The figure shows that all strategies have almost identical overall inter-AS costs. The inter-AS costs of sequential and nested optimization are similar because the *inter-optimal-aware-alg* and the GA use a similar heuristic. These algorithms effectively produce near-optimal inter-AS costs (within 1% from the optimal solution). The NSA of integrated optimization has also reached a similar inter-AS cost because a significant importance has been given to optimize inter-AS cost over intra-AS cost, as mentioned in Section IV.C. As a consequence, it is expected that those Joint-TE strategies will strive to obtain a near-optimal inter-AS TE cost and, hence, their costs are almost identical.

Figure 4(b) shows the corresponding maximum inter-AS link utilization. On the whole, all the strategies obtain a similar utilization level although there are some small differences among them at some traffic demands. This mainly result from the fact that the piecewise linear cost function gives the same cost to links with utilizations in the same piecewise linear segment, such as from 1/3 to 2/3 (as shown between any two dashed horizontal lines in the figure). In other words, such links are considered as at same utilization level. In the figure, we see that all the strategies have utilization level within the same block at any traffic demand.

We conclude from Figure 4(a) and (b) that the inter-AS TE solutions achieved by all the strategies are identical and near-optimal. Recall in Section IV.C that, in order to achieve an unambiguous comparison, our aim is to derive TE solutions that remain overall inter-AS cost near-optimal while substantially improving intra-AS cost. At this point, the objective of inter-AS TE has been achieved. In order to determine which strategies produce the best and the worst overall network performance, we proceed to evaluate their overall intra-AS costs.

B. *Evaluation of Overall Intra-AS TE Cost*

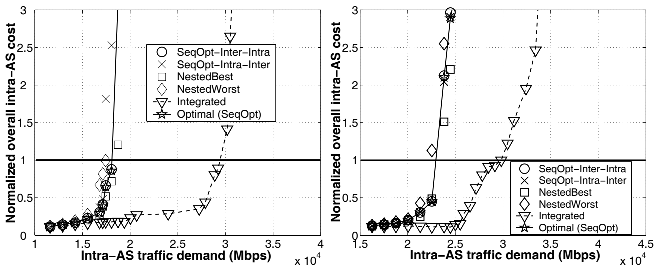

Figure 5. Normalized overall intra-AS cost (left: Sprint, right: AT&T)

In our intra-AS TE experiments, the intra-AS TM is the entire local intra-AS TM with an increasing amount of inter-AS traffic. For brevity, we only illustrate the results of Sprint and AT&T topologies using figures because they have larger topologies and have similar results to the BT backbone and the Abovenet topologies. Following the method of deriving a universal measure of congestion in [11], we normalize the resulting overall intra-AS cost by the cost derived from the hop-count based shortest path routing with infinite capacity on each intra-AS link under *SeqOpt-Inter-Intra* which is the current practice sequential approach. If the normalized cost is larger than one, it implies that the algorithm is performing as badly as if all flows were along shortest hop paths with loads matching the capacities.

Considering the general picture for the normalized overall intra-AS costs in Figure 5, we first see that *NestedWorst* has the worst overall intra-AS cost. Then come the two sequential optimization methods closely together and *NestedBest* is the best among all of them. The curve denoted by *Optimal (SeqOpt)* is the optimal intra-AS TE solution of *SeqOpt-Inter-Intra*, which is obtained by solving the LP problem rather than *intra-optimal-aware-alg*. We see that the intra-AS cost of *SeqOpt-Inter-Intra* is within 3% of the optimal solution, thus showing that the *intra-optimal-aware-alg* can achieve near-optimal solutions.

The above results are expected and can be explained as follows. First of all, the nested optimization finds the best and the worst intra-AS TE solution by evaluating many equal-cost acceptable inter-AS TE solutions, with each solution performing a sequential optimization. Hence, *NestedBest* and *NestedWorst* can be regarded as the lower and the upper bound of the sequential optimization. As could be seen, indeed, the performances of the two sequential optimization methods (*SeqOpt-Inter-Intra* & *SeqOpt-Intra-Inter*) are between *NestedBest* and *NestedWorst*. The difference in performance between the sequential and the nested optimization sufficiently demonstrates that there indeed exists optimal inter-AS TE solutions that are far better and worse with respect to intra-AS TE. Since the fundamental characteristics of both sequential and nested optimization follow a sequential model that do not optimize intra- and inter-AS TE costs simultaneously, their performances are generally poor. This contrasts to the superior performance of integrated optimization.

We follow the methodology in [11] to quantify and compare the performances of different Joint-TE strategies. The comparison metric is the amount of traffic demand the network can cope with before it gets congested (i.e. the normalized intra-AS cost gets to one). Our experiments show that the integrated optimization allows the network to cope with 30%-60%[4] more traffic demand than the other non-integrated optimization approaches. This significant improvement implies that the intra-AS traffic matrices for the sequential and the nested optimization are not yet optimized and leading to better overall network performance.

The intra-AS cost can reflect the performance of maximum intra-AS link utilization and total bandwidth consumption. Turning our attention first to the maximum intra-AS link utilization, Figure 6 shows the performance achieved by different strategies. The integrated optimization is the best strategy. It attempts to keep the maximum utilization below 100% to avoid the high cost penalty as the network load increases. As a consequence, the number of additional traffic flows the network can support before suffering congestion (which we define here as a maximum utilization above 100%) is approximately at least 30% more than the non-integrated optimization strategies.

[4] Approximately 30% for AT&T, 35% for Abovenet, 46% for BT Backbone and 60% for Sprint

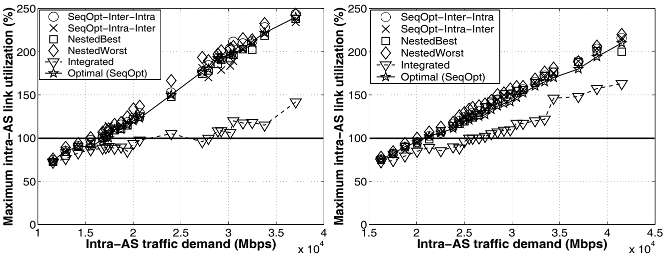

Figure 6. Maximum intra-AS link utilization (left: Sprint, right: AT&T)

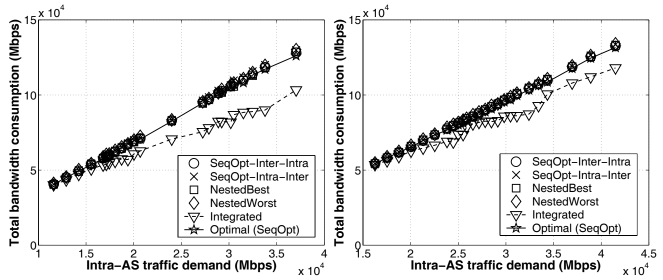

Figure 7. Total bandwidth consumption (left: Sprint, right: AT&T)

Figure 7 shows that the total intra-AS bandwidth consumption achieved by the integrated optimization is lower than the other strategies. Together with the results of the maximum intra-AS link utilization, we see that the integrated optimization has employed more intra-AS routes that are short and well load balanced in the network. This explains why integrated optimization can achieve better overall intra-AS cost than the other strategies.

C. *Summary of the Evaluations*

From our evaluations of inter- and intra-AS costs, we see that the integrated optimization has successfully produced from non-TE starting solutions to the final solutions that have the same inter-AS cost as those obtained from inter-AS TE (*inter-optimal-aware-alg* and the GA) with an improved intra-AS cost. It is worth noticing that, in comparison to those intelligent intra- and inter-AS TE algorithms that produce near-optimal solutions in a decoupled mode, the overall performance improvement of the integrated optimization is remarkable. An implication of this finding is to encourage ISPs to move towards a fully integrated TE approach that is aware of both intra- and inter-AS TE simultaneously.

D. *Optimized vs. Non-Optimized Starting Solutions for Integrated Optimization*

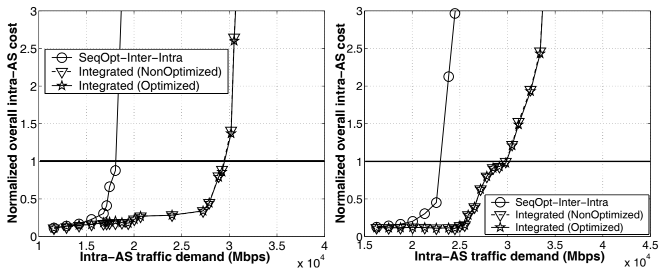

Figure 8. Optimized vs. Non-optimized starting solutions (left: Sprint, right: AT&T)

In the preceding experiments, we have used the non-TE (non-optimized) starting solutions for the integrated optimization. In this section, we evaluate the performance of the integrated optimization using a good quality starting solution. Our hypothesis is that much better performance improvement can be obtained when an optimized starting solution is used. We use the solution obtained from *SeqOpt-Inter-Intra* as optimized starting solution, which is achieved by the existing intra- and inter-AS TE algorithms. As previously shown, the solutions are near-optimal with respect to inter- and intra-AS TE costs when these TE are accomplished separately.

In line with the phenomenon observed in Section VII.A, the integrated optimization using non-optimized and optimized starting solutions produce very similar overall inter-AS costs, which can be regarded as identical. For intra-AS performance, Figure 8 shows that the integrated optimization using optimized starting solution can achieve much better intra-AS cost than the sequential optimization. However, it is only slightly better than that using non-optimized starting solution. This refutes our hypothesis.

The significant performance improvement over the sequential optimization is expected due to the simultaneous TE optimization and that the integrated optimization guarantees the performance of its solution is at least no worse than the starting solutions. On the other hand, the small performance improvement over that using non-optimized starting solutions reflects that the quality of starting solution has not much influence on the quality of the final solution. That is simply because the optimized starting solution itself is not really optimized from the viewpoint of Joint-TE and its quality could be even far inferior from the optimal Joint-TE solution than a non-optimized starting solution. Thus, an optimal inter-AS TE solution can be a mediocre starting point with respect to intra-AS TE, but there are inter-AS TE solutions that are nearly optimal and are far better with respect to the intra-AS TE objectives. An implication of this finding is that the effort on devising the existing intra- and inter-AS TE algorithms, which are assumed used in a decoupled mode, may not be sufficient to achieve a truly optimized network performance due to the TE interactions and the fact that the overall network performance can be significantly improved through their joint optimization. The existing inter-AS TE approaches may not improve intra-AS TE performance even though an opportunity for such improvement exists. On the other hand, the existing intra-AS TE approaches may not be able to achieve a truly optimized performance without making it aware of inter-AS TE. In line with our proposal in Section VII.C, the integrated TE approach is an appropriate solution.

VIII. RELATED WORK

Some recent work has investigated the interactions between intra- and inter-AS routing such as the dynamic and the disruption effects between BGP and hot-potato routing [2, 3]. The authors in [37] have investigated the impact of BGP route changes on intra-AS traffic, which is related to our work. In addition, some work [38, 39] has considered the interactions between BGP and IGP in TE tools. However, none of these papers has either investigated the interactions between intra- and inter-AS routing from a TE perspective, or proposed any strategy or algorithm for their joint optimization.

Few attempts have been recently carried out towards this TE interaction. Our previous work first proposes preliminary thoughts and approaches for the TE interaction [40]. Agarwal et al [41] evaluate the behavior of hot-potato routing during intra-AS link weight optimization. The key difference between our work and those previous works is that ours investigates several very different approaches for the TE interaction. These approaches covered rather complete and representative operations that can be considered by real-world network management. In addition, we propose bi-criteria algorithms for solving the TE interaction problem, which has not been investigated by previous works.

IX. CONCLUSION

This paper has considered a joint intra- and inter-AS TE optimization scheme. We showed two interaction effects between intra- and inter-AS TE that can lead to suboptimal overall network performance. These interactions motivate the need for joint optimization of intra- and inter-AS TE in order to further optimize the overall network performance. We first formulated this joint TE optimization as a bi-criteria optimization problem. Then we presented three strategies, namely sequential, nested and integrated optimization, to solve it. Our experimental evaluations revealed that the integrated optimization, which solves intra- and inter-AS TE simultaneously, allows the network to accommodate 30% - 60% more future traffic demand in comparison to the other strategies that deal with intra- and inter-AS TE separately. The integrated optimization therefore provides a marked improvement on current industry practice towards the collective use of intra- and inter-AS TE.

ACKNOWLEDGEMENT

This work was undertaken in the context of FP6 Information Society Technologies AGAVE (IST-027609) and EMANICS Network of Excellence (IST-026854) projects, which are partially funded by the Commission of the European Union.

REFERENCES

[1] D. Awduche et al., "Overview and Principles of Internet Traffic Engineering," *IETF RFC* 3272, May 2002.

[2] R. Teixeira et al., "Dynamics of Hot-Potato Routing in IP Networks," Proc. *ACM SIGMETRICS* 2004.

[3] R. Teixeira et al., "Network Sensitivity to Hot Potato Disruptions," Proc. *ACM SIGCOMM* 2004.

[4] N. Spring et al., "Measuring ISP Topologies with Rocketfuel," Proc. *ACM SIGCOMM* 2002.

[5] A. Feldmann et al., "Deriving Traffic Demands for Operational IP Networks: Methodology and Experience", *IEEE/ACM Transactions on Networking*, 2001, pp. 265-279.

[6] T.C. Bressoud et al., "Optimal Configuration for BGP Route Selection," Proc. *IEEE INFOCOM* 2003.

[7] Y. Zhang et al., "An Information-Theoretic Approach to Traffic Matrix Estimation," Proc. *ACM SIGCOMM* 2003.

[8] A. Medina et al., "Traffic Matrix Estimation: Existing Techniques and New Directions," Proc. *ACM SIGCOMM 2002*.

[9] R. Teixeira et al., "Traffic Matrix Reloaded: Impact of Routing Changes," Proc. *Passive and Active Measurement Workshop*, 2005.

[10] A. Feldmann et al., "A Methodology for estimating Interdomain Web Traffic Demand," Proc. *ACM IMC* 2004.

[11] B. Fortz and M. Thorup, "Internet Traffic Engineering by Optimizing OSPF Weights." Proc. *IEEE INFOCOM* 2000.

[12] X. Xiao et al., "Traffic Engineering with MPLS in the Internet," *IEEE Network Magazine*, 14(2), 2000.

[13] N. Hu et al., "Locating Internet Bottlenecks: Algorithms, Measurements, and Implications," Proc. *ACM SIGCOMM* 2004.

[14] N. Feamster et al., "Guidelines for Interdomain Traffic Engineering," *ACM SIGCOMM CCR*, October 2003.

[15] B. Quoitin et al., "Interdomain Traffic Engineering with BGP," *IEEE Communication Magazine*, May 2003.

[16] S. Uhlig and B. Quoitin, "Tweak-it: BGP-based Interdomain Traffic Engineering for transit ASes," Proc. *EuroNGI Conference* 2005.

[17] K.H. Ho et al., "An incentive-based Quality of Service Aware Algorithm for Offline Inter-AS Traffic Engineering," Proc. *IEEE IPOM* 2004.

[18] R. Gao et al., "Interdomain Ingress Traffic Engineering through Optimized AS-Path Prepending," Proc. *IFIP Networking* 2005.

[19] R.K.C. Chang and M. Lo, "Inbound Traffic Engineering for Multihomed ASes using AS Path Prepending," *IEEE Networks,* 19(2), 2005.

[20] C.N. Chuah, "A Tier-1 ISP perspective & observations of routing behavior", A talk, 2002.

[21] B. Halabi, *Internet Routing Architectures*, 2/E, Cisco Press.

[22] J.P. Vasseur and A. Ayyangar, "Inter-area and Inter-AS MPLS Traffic Engineering", *IETF Internet Draft*, 2004.

[23] K. Kar, M. Kodialam and T.V. Lakshman, "Minimum Interference Routing of Bandwidth Guaranteed Tunnels with MPLS Traffic Engineering Applications", *IEEE JSAC*, 18(12), December 2000.

[24] R.E. Steuer, *Multi Criteria Optimization – Theory, Computation, and Application*, New York: Wiley, 1986.

[25] S. Saroiu et al., "An Analysis of Internet Content Delivery Systems," Proc. *USENIX OSDI* 2002.

[26] A. Sridharan and R. Guerin., "The Impact of Traffic Granularity on robustness of Traffic Aware Routing," Technical Report, Multimedia and Networking Lab, University of Pennsylvanians, U.S, April 2004.

[27] C.T. Chou, "Traffic Engineering for MPLS-based Virtual Private Networks," *Computer Networks*, 44(3), 2004.

[28] A. Sridharan et al., "Achieving Near-Optimal Traffic Engineering Solutions for Current OSPF/IS-IS Networks," *IEEE ToN*, 13(2), 2005.

[29] F. Glover and M. Laguna, *Tabu Search*, Kluwer Academic Publishers, Norwell, MA, 1997.

[30] J. Shen et al., "A tabu search algorithm for the routing and capacity assignment problem in computer networks," *Computers & Operations Research*, 32, 2005.

[31] T.H. Cormen et al., *Introduction to Algorithms*, MIT Press.

[32] W. Fang and L. Peterson, "Inter-AS Traffic Patterns and Their Implications", Proc. *IEEE Global Internet Symposium* 1998.

[33] A. Broido et al., "Their shares: diversity and disparity in IP traffic," Proc. *Passive and Active Measurement Workshop*, 2004.

[34] A. Nucci et al., "IGP link weight assignment for transient link failures," *Proc. 18th ITC*, 2003.

[35] S. Bhattacharyya et al., "POP-Level and Access-Link-Level Traffic Dynamics in a Tier-1 POP," Proc. *ACM IMW* 2001.

[36] X.H. Lin et al., "A Genetic Algorithm Based Approach to Route Selection and Capacity Flow Assignment," *Computer Comm..*, 26, 2003.

[37] S. Agarwal et al., "The Impact of BGP Dynamics on Intra-Domain Traffic," Proc. *ACM SIGMETRICS* 2004.

[38] A. Feldmann et al., "NetScope: Traffic Engineering for IP Networks," *IEEE Network Magazine*, 2000, pp. 11-19.

[39] N. Feamster et al., "A Model of BGP Routing for Network Engineering," Proc. *ACM SIGMETRICS* 2004.

[40] M. Howarth et al, MESCAL deliverable D1.2, January 2004. Available at www.mescal.org.

[41] S. agarwal et al., "Measuring the Shared Fate of IGP Engineering and Interdomain Traffic," Proc. *IEEE ICNP* 2005, Boston, USA.