Towards Decentralized and Adaptive Network Resource Management

D.Tuncer, M.Charalambides, G.Pavlou
Department of Electronic & Electrical Engineering
University College London
London, UK

N.Wang
Centre for Communication Systems Research
University of Surrey
Surrey, UK

Abstract—**Current practices for managing resources in fixed networks rely on off-line approaches, which can be sub-optimal in the face of changing or unpredicted traffic demand. To cope with the limitations of these off-line configurations new traffic engineering (TE) schemes that can adapt to network and traffic dynamics are required. In this paper, we propose an intra-domain dynamic TE system for IP networks. Our approach uses multi-topology routing as the underlying routing protocol to provide path diversity and supports adaptive resource management operations that dynamically adjust the volume of traffic sent across each topology. Re-configuration actions are performed in a coordinated fashion based on an in-network overlay of network entities without relying on a centralized management system. We analyze the performance of our approach using a realistic network topology, and our results show that the proposed scheme can achieve near-optimal network performance in terms of resource utilization in a responsive manner.**

Keywords- Adptative Resource Management, Online Traffic Engineering, Decentralized Network Configuration

I. INTRODUCTION

Today's TE practices mainly rely on off-line settings that use traffic demand estimates to derive network configurations. However, because of their static nature, these practices do not take network and traffic dynamics into account and can lead to sub-optimal overall performance. To cope with unexpected traffic variations and network dynamics, approaches that can dynamically adapt routing configurations and traffic distribution are required. Despite recent proposals to enable adaptive traffic engineering in plain IP networks [11][15][16], current approaches normally rely on a *centralized* TE manager to periodically compute new configurations according to dynamic traffic behaviors.

In this paper, we present DACoRM (Decentralized Adaptive Coordinated Resource Management), a new adaptive resource management approach for IP networks in which traffic distribution is dynamically adapted according to real-time network conditions. Our approach allows for the traffic between any pair of end points in the network to be balanced across several paths according to splitting ratios, which are (re-)computed by the network nodes themselves in real-time. The set of possible routes, enabled by multi-topology routing (MTR), is determined by an off-line configuration process, and is not modified by the adaptive scheme. The adjustment of the splitting ratios relies on run-time information about the network state and does not require any prior estimates of traffic demand i.e. traffic matrices. Most important, new configurations are not computed by a centralized management entity that has a global view of the network, but instead, the source nodes coordinate among themselves to decide on the course of action to follow. Each source node is responsible for adjusting the ratios of its locally originating traffic based on the result of the coordination.

Initial evaluations of our adaptive resource management scheme based on a realistic topology and with real traffic traces are encouraging. Our results show that near-optimal network performance can be achieved in terms of resource utilization in a responsive and stable manner.

The remainder of this paper is organized as follows. In section II, we introduce the necessary background on MTR and online TE. In section III, we present the main features of our scheme which we then detail in section IV. Section V presents the results of the evaluation of our approach and in section VI, we review related work. We finally present a summary and point to future work.

II. BACKGROUND

Current practices for intra-domain TE rely on off-line approaches, where a central management system is responsible for computing routing configurations, especially tuning link weights based on the estimation of the traffic demand. The goal of these approaches is to find a routing configuration that optimizes the network performance over long timescales, e.g. weekly or monthly. Off-line TE schemes have been extensively investigated both in the context of MPLS-based TE by using MPLS paths and in the context of IP-based TE by determining heuristics to tune the link weights that optimize some objective function given a set of traffic matrices [1][2][3][4]. As such, off-line approaches may be sub-optimal in the face of unexpected traffic demand.

In contrast to these off-line schemes, online TE approaches do not rely on the knowledge of any traffic matrix to configure the routing or the link weights. Instead, they dynamically adapt the settings in short timescales in order to rapidly respond to traffic dynamics [12]. These schemes do not rely on any knowledge of future demands to configure the settings but instead use monitored real-time information from the

network. In order to satisfy the future traffic demands, online TE approaches aim at adaptively distributing the traffic load as evenly as possible onto the network according to the changing traffic conditions.

There have been some proposals for both online MPLS-based TE such as [13][14] and online IP-based TE such as [11][15][16]. These approaches focus on dynamically adjusting the volume of traffic (represented by splitting ratio) sent across several available paths between each S-D pair in the network according to real-time traffic information.

In a similar fashion to other intra-domain TE approaches [9][10][11], our approach uses multi-topology routing (MTR) [5][6] as the underlying network routing protocol to provide a set of available routes between each pair of edge nodes. MTR is a standardized extension to the common IGP routing protocols OSPF and IS-IS, that aims at determining several independent virtual IP topologies based on a single network topology, each having its own independent routing configurations, especially its own link weight settings. Based on these link weight settings, the Shortest Path First (SPF) algorithm can be applied independently between each source-destination (S-D) pair in each topology. Thus, it is possible to compute a set of paths between each pair of end points in the network with each path being related to one virtual topology. More precisely, the traffic demand between any pair S-D is virtually split into *n* independent sets at ingress nodes and each traffic set is assigned to one of the *n* topologies and routed according to that topology's configuration. Results in [9], [10] and [11] show that only a small number of topologies (typically between 3 and 5) is enough to offer pretty good path diversity.

III. PROPOSED APPROACH

A. Overview

Our online TE system performs adaptive resource management by dynamically adjusting the splitting ratios according to network conditions. The TE re-configuration actions performed are decided in a coordinated fashion between a set of source nodes forming an *in-network* overlay (INO).

Based on the path diversity provided by configuring the different virtual topologies, the proposed approach controls the distribution of traffic load in the network in an adaptive and decentralized manner through re-configuration actions. The objective of this adaptive control is to dynamically balance the traffic load such that traffic is moved from the most utilized[1] links towards less loaded parts of the network. As described in the section II, the traffic demand between each S-D pair is divided into *n* sets at source nodes, with each set being associated to one of the *n* topologies *T*. The proportion of traffic assigned to each set is determined by splitting ratios, which are used to distribute incoming flows at source nodes. Flows are routed to their destination according to the configuration of the topology they have been assigned to. Splitting ratios are not pre-computed by an off-line process as in other approaches, e.g. [9], but are instead adapted dynamically by the source nodes themselves, even without centralized control as is the case in [11]. New splitting ratios are computed by a re-configuration algorithm that executes only at source nodes, which allows them to react to traffic dynamics in an online fashion by adjusting the proportion of traffic assigned to each topology. If a link gets congested for instance, the nodes can automatically decide to re-configure the splitting ratios and hence move some of the load on that link to less utilized parts of the network. The adaptation is performed periodically in short time scales, every 5-10 minutes.

B. System Design

In order to realize the proposed adaptive resource management scheme a set of components need to be deployed at source nodes as depicted in Fig. 1.

Each source node *S* maintains locally both static and dynamic information related to each of the traffic flows originating at that node. Note that, in this paper, we refer to a traffic flow as the volume of traffic between a source node and a destination node. This information is stored in two tables that we call the Link Information Table (LIT) and the Demand Information Table (DIT) respectively – Fig. 2 shows the structure of each table. The LIT contains static information about the links traversed by the paths of all the locally originating flows. For each link, its stores the link capacity, references to the S-D pairs that use that link for routing their associated traffic flow in at least one topology, and for each of these S-D pairs, the involved topology(ies). Based on this information, source nodes can efficiently determine if a flow contributes to the load of a link in the network, and if so, on which topology(ies). The DIT contains dynamic information related to each flow entering the network at the source node. It maintains the current and previous splitting ratios assigned to each topology, as well as the associated traffic volume for the current time interval.

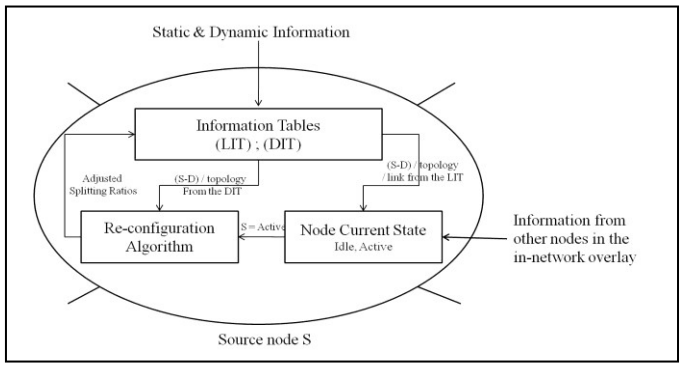

Figure 1. Components overview at the source node level

[1]The utilization of a link is defined as the ratio between the link load and the link capacity

Link l_1	Capacity $C(l_1)$	S-D pair id_1	Topology T_1
			….
			Topology T_N
		…..	….
		S-D pair id_N	Topology T_1
			…
			Topology T_N

a) Link Information Table (LIT)

S-D pair id_1	Traffic Volume v(S-D)	$TopoT_1$	Previous Ratio $x_{/T1}(S\text{-}D)_{pre}$	CurrentRatio $x_{/T1}(S\text{-}D)_{cur}$
		…..		
		$TopoT_N$	Previous Ratio $x_{/TN}(S\text{-}D)_{pre}$	CurrentRatio $x_{/TN}(S\text{-}D)_{cur}$

b) Demand Information Table (DIT)

Figure 2. Information tables maintained by each source node

Based on information stored in the tables and on information received through the INO, a source node can determine its current state, which can be either idle or active depending on whether or not it needs to perform re-configuration. When in the latter state, the node executes the re-configuration algorithm over its locally originating flows to determine the new splitting ratios that can decrease the utilization of the most utilized link in the network. If with these new splitting ratios no other link in the network gets overloaded, the adjusted splitting ratios are updated in the corresponding DIT and enforced at the next time interval.

IV. ADAPTIVE RESOURCE MANAGEMENT

A. Coordinated Re-configuration

Performing a re-configuration involves adjusting the traffic splitting ratios for some of the S-D pairs for which traffic is routed across the link with the maximum utilization in the network (noted *lmax*). This means that more traffic is assigned to topologies not using *lmax* to route traffic thus decreasing the traffic volume assigned to topologies that do use *lmax*.

The splitting ratios for each traffic flow are configured only by the corresponding source node. In realistic scenarios, links in the network are used by multiple flows and therefore, several source nodes may be eligible to adapt the ratios of flows traversing *lmax*. Due to the limited network view of individual source nodes, actions taken by more than one node at a time may lead to inconsistent decisions, which may jeopardize the stability and the convergence of the overall network behavior. For instance, in the process of shifting traffic away from *lmax*, the different reacting nodes can re-direct traffic flows towards the same links, as depicted in Fig. 3 thus potentially causing congestion. In Fig. 3, source nodes N1 and N2 send traffic over link $l_{5\text{-}6}$. In (a), $l_{5\text{-}6}$ is identified as being the most utilized link in the network. N1 and N2 both react to this information (b) and decide to perform some re-configurations locally, which results in more traffic from both N1 and N2 routed towards link $l_{3\text{-}4}$, which then becomes overloaded (c).

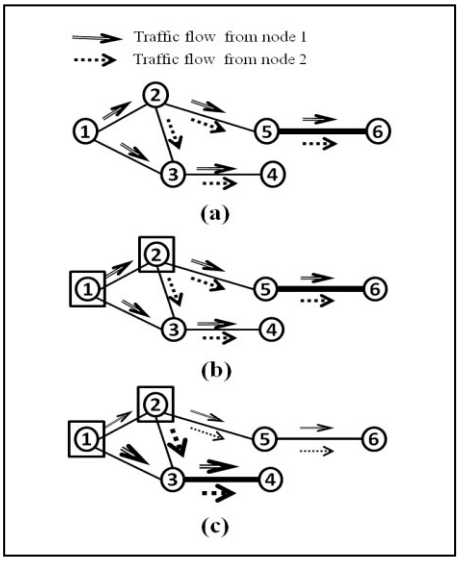

Figure 3. Example of conflicting decisions between node N1 and node N2

To avoid such inconsistent decisions, DACoRM is designed so that only one source node is permitted to change the splitting ratios of one of its local traffic flows at a time. When an adaptation is required, the source nodes coordinate through the INO to select one of them that will compute and enforce the new ratios. The selected node is responsible for executing the re-configuration algorithm over its locally originating traffic flows with the objective to re-balance the network load.

B. In-Network Overlay of Coordinated Entities

The INO of source nodes is built during the initial configuration of the network in an off-line manner. Its formation is based on the identification of ingress nodes in the physical network, i.e. the nodes which are potential sources of traffic. In the case of a PoP (Point of Presence) level network, for instance, each node is a potential source of traffic and would therefore be part of the INO. Each node *N* in the INO is associated with a set of neighbors – nodes that are directly connected to the INO – with direct communication only possible between neighboring nodes.

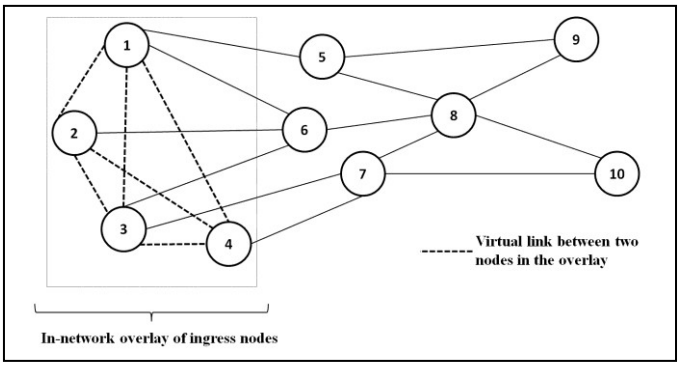

Figure 4. Example of a network and its associated full-mesh in-network overlay of ingress nodes

Although different types of INO topologies can be used, e.g. ring, star, full-mesh, in this paper we concentrate on a full-mesh topology, where there exists a direct virtual link between all source nodes. Such a topology is depicted in Fig. 4 where the four ingress nodes on the physical network are logically connected in a full-mesh. This topology offers a greater flexibility in the choice of neighbors with which to communicate since all source nodes belong to the set of neighbors. However, the choice of the INO topology may be driven by different parameters related to the physical network, such as its topology, the number of source nodes, but also by the constraints of the coordination mechanism and the associated communication protocol. The number and frequency of messages exchanged, for example, are factors that influence the choice of topology.

We plan to analyze the performance and flexibility of different topologies in future extensions of this work.

C. Adaptation Process

The overall objective of DACoRM is to balance the load in the network by moving some traffic away from highly utilized links towards less utilized ones. To achieve this objective, the proposed adaptive resource management scheme successively adjusts the splitting ratios of traffic flows through a sequence of re-configurations. At each iteration, DACoRM identifies the link with the maximum utilization, *lmax*, and a set of other heavily utilized links, S_{HU}, in the network. S_{HU} is defined as the set of links in the network with a utilization within α % of the utilization of *lmax*.

This information is shared among the source nodes in the INO and is used to select one of them that will compute new ratios. The selected source node is responsible for modifying the splitting ratios of one of its traffic flows contributing to the load on *lmax* such that: a) some traffic is moved away from *lmax*, and, b) the diverted traffic is not directed towards links in the set S_{HU}. The adaptation process terminates if a successful configuration cannot be determined or if it reaches the maximum number of permitted iterations (a parameter of the algorithm).

The adaptation process consists in adjusting the splitting ratios of some traffic flows such that the load-balancing objective is satisfied. At each iteration of this process, the selected source node is responsible for executing a re-configuration algorithm. The objective of the re-configuration algorithm is to determine if re-configuration can be performed on one of the local traffic flows. The outcome of the algorithm is either positive, which means that part of a local flow can be diverted from *lmax*, or negative if this is not possible. More precisely, the algorithm considers each local traffic flow at a time and tries to adjust its splitting ratios. These are adjusted such that the ratios related to the topologies that use *lmax* to route the traffic flow are decreased while the ratios related to the alternative topologies not using *lmax* are increased. The resulting configuration is then analyzed to decide whether it is acceptable or not. If so, the new splitting ratios are accepted.

One of the challenges addressed by our re-configuration algorithm is determining the volume of traffic that can be diverted from *lmax* in each iteration, while at the same time preserving the network stability. If too much traffic is shifted, other links may become overloaded. This may cause oscillations as in the next iteration traffic will need to be removed from these links. To guarantee stability, the volume of traffic that can be diverted at each iteration is computed to satisfy an upper bound constraint.

V. EXPERIMENTAL RESULTS

We have evaluated the performance of our approach using the real PoP-level topology of the Abilene network and the traffic matrices available from [17] that provide traffic traces for 5 minute intervals during a 7 day period. The Abilene network topology consists of 12 PoP nodes and 30 unidirectional links.

To analyze the performance of the proposed adaptive scheme in terms of maximum utilization (max-u) in the network, we compare the results achieved by DACoRM with the results obtained by three other schemes:

Original scheme: the original link weight settings are used in the original topology and no adaptation is performed.

MTR scheme: the computed virtual topologies are used to provide path diversity and initial random splitting ratios are applied, but no further adaptation of these ratios is performed.

Optimal scheme: the TOTEM toolbox is used to compute the optimal maximum link utilization for each traffic matrix.

The objective of the comparison with the MTR scheme, where no adaptation is performed, is to evaluate the performance of our adaptive resource management scheme, which performs periodic re-configurations, in terms of resource utilization gain.

The settings of the different parameters of the algorithm used to perform the experiments presented in this section are summarized in TABLE I.

Fig. 5 presents the evolution of max-u in the Abilene network at 5 minute intervals over a period of one week for the following schemes: (a) original scheme, (b) the MTR scheme, and, (c) DACoRM. We can observe that the maximum utilization achieved with DACoRM is less than the ones obtained by the original and the MTR schemes.

TABLE I. ALGORITHM PARAMETER SETTINGS

Parameter	Value
Number of topologies	4
α	10 %
Max number of iterations	15
Frequency of adaptation	Every 5 min

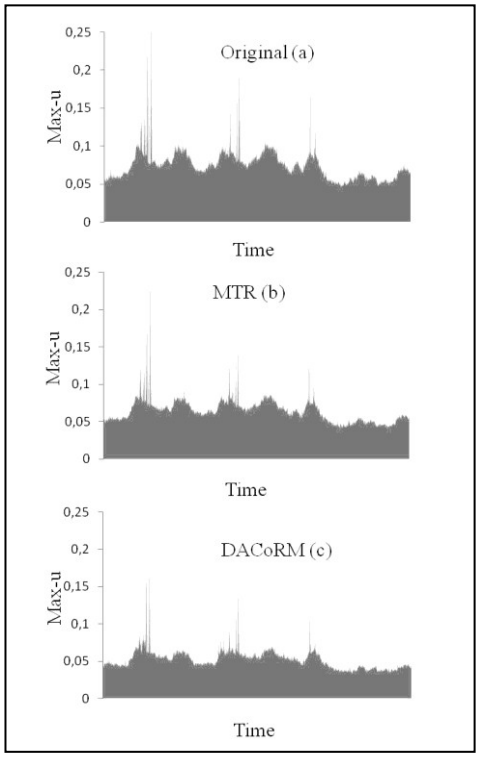

Figure 5. Evolution of the maximum utilization in the Abilene network over a period of 7-days for the different schemes

To quantify the actual gain, TABLE II. presents the average deviation from the optimum over a period of one week. This corresponds to more than 2000 traffic matrices that thus represents a wide variety of traffic conditions. Our approach achieves a near optimal result with an average deviation of less than 10% from the optimum, while the other two schemes do not perform as well. The reason for MTR performing better than using only original link weight settings is that traffic is more evenly balanced over the different links in the network. The substantial difference in performance is more obvious when fewer samples are plotted, as in the case of Fig. 6 where only 200 traffic matrices (corresponding to roughly 1 day) are used.

To investigate the influence of MTR on our adaptation algorithm, we also analyze how DACoRM performs depending on the number of topologies used. Using the 2000 traffic traces we varied the number of topologies and observed the average deviation of the maximum utilization. The results are presented in Fig. 7. The figure shows that the average deviation decreases as the number of topologies increases. This is consistent with results in [9], [10] and [11] that show that between 3 and 5 topologies are typically required to offer good path diversity.

A performance indicator of any online adaptive scheme is time-complexity. Different factors may affect the convergence time of our adaptive resource management scheme, such as the size of the network, the structure of the INO, the number of selected neighbors, as well as the design of the communication protocol between nodes in the INO.

TABLE II. DEVIATION OF THE MAXIMUM UTILIZATION FROM THE OPTIMAL

	Deviation from the optimal (%)
Original scheme	54 %
MTR scheme	34 %
DACoRM	7,2 %

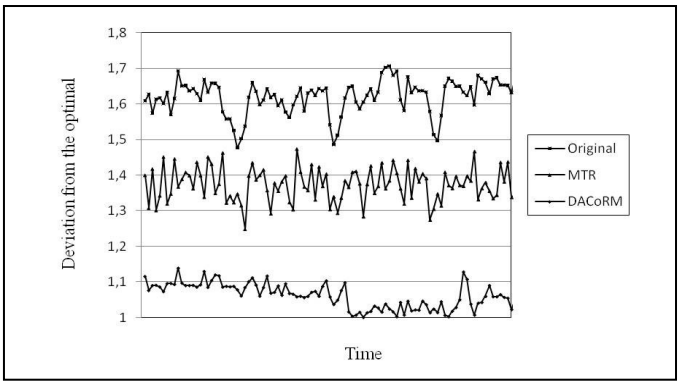

Figure 6. Deviation of the maximum utilization from the optimum in the Abilene network

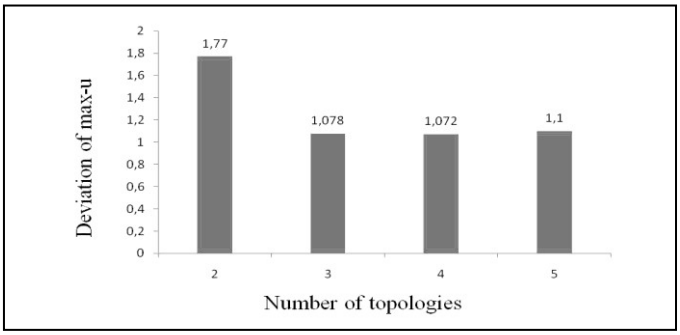

Figure 7. Influence of the number of topologies

In addition, the execution time of our algorithm may also be influenced by the re-configuration parameter settings, such as the value of α and the maximum number of iterations. Although we have not implemented the communication protocol between the source nodes in the INO, the coordination process is emulated by executing the algorithm for each source node in parallel and by randomly selecting one among those that can determine a new configuration for one of its local flow. On average, it takes 100ms for DACoRM to determine a new configuration, i.e. for computing and enforcing new splitting ratios, and 7ms for a source node to determine new splitting ratios. This amount can be considered negligible compared to the frequency at which the adaptive resource management scheme is invoked, i.e. every 5 minutes.

VI. RELATED WORK

As advocated in [10], our approach uses MTR to support path diversity between each source-destination pair. While MT extensions have been initially developed for the purpose of routing different types of services and traffic across different paths in the network, there has been interest to use MT principles for other purposes such as network resilience [7][8],

intra-domain off-line TE [9], and online TE [10][11]. In [10], the authors do not propose a detailed online TE solution, but they instead expose the incentives for extending and using MTR for TE purposes.

Dynamic adaptation of the splitting ratios was initially proposed by two MPLS-based TE solutions, MATE [13] and TeXCP [14], where ingress routers use periodical information from the network to adjust the ratios. Unlike MATE and TeXCP where re-configurations are performed at ingress nodes only, in AMP [15] and REPLEX [16], all the nodes in the network are responsible for dynamically splitting the traffic between the different available next hops, based on information received from upstream routers. In DACoRM, core nodes do not participate in the re-configuration process since adaptations are performed at network edges only. However, unlike TeXCP, where the adjustment actions are supported by a control mechanism implemented by the core nodes, re-configurations in our approach are the result of a coordination process between the source nodes in the INO.

Unlike the above distributed approaches, the authors in [11] use a central controller that has a global knowledge of the network state to perform the re-configurations. The advantage of such a centralized decision-making process is that the consistency between different re-configuration actions is guaranteed. However, using a centralized approach is less scalable than a solution where decisions are taken by nodes themselves inside the network, since at each re-configuration period the central controller needs to gather information from all the links and nodes in the network, which incurs a significant communication overhead.

VII. SUMMARY AND FUTURE WORK

In this paper, we have presented DACoRM, an adaptive resource management scheme for intra-domain TE, where source nodes coordinate among themselves through an in-network overlay to decide on the course of action to follow to re-balance the traffic load across several paths according to network conditions. Unlike off-line TE approaches which rely on static configurations, DACoRM can efficiently deal with traffic and network dynamics by enabling adaptation of routing configuration in short timescales. The results of our experiments, based on the Abilene network and real traffic traces, show that our approach can efficiently achieve substantial gain in terms of network resource utilization.

Future work will focus on the implementation of the communication protocol between source nodes in the in-network overlay. We plan to further analyze the impact of the different factors and parameter settings on the performance of our approach, both in terms of time-complexity as well as resource utilization gain. We also plan to evaluate our approach in different network topologies and to compare its performance to other adaptive TE schemes.

ACKNOWLEDGMENT

This work was partially supported by the European Union UniverSELF and COMET projects of the 7th Framework Program.

REFERENCES

[1] B. Fortz, M. Thorup, “Internet traffic engineering by optimizing OSPF weights”, in: Proceedings of IEEE INFOCOM 2000, Tel-Aviv, Israel, 2000.

[2] B. Fortz et al., “Traffic Engineering with Traditional IP Routing Protocols,” IEEE Commun. Mag., vol. 40, no. 10, Oct. 2002, pp. 118–24.

[3] A. Sridharan, R. Guerin, C. Diot, “Achieving Near-Optimal Traffic Engineering Solutions for Current OSPF/IS-IS Networks”, in: IEEE INFOCOM 2003, San Francisco, CA, 2003.

[4] Dahai Xu, Mung Chiang, and Jennifer Rexford, "Link-state routing with hop-by-hop forwarding can achieve optimal traffic engineering," to appear in IEEE/ACM Transactions on Networking.

[5] P. Psenak et al., “Multi-Topology (MT) Routing in OSPF,” RFC 4915, June 2007.

[6] T. Przygienda, N. Shen, N. Sheth, “M-ISIS: multi topology (MT) routing in Intermediate System to Intermediate Systems (IS-ISs)” , IETF RFC 5120, February 2008.

[7] Menth, M.; Martin, R.; , "Network resilience through multi-topology routing," Design of Reliable Communication Networks, 2005. (DRCN 2005). Proceedings.5th International Workshop on , vol., no., pp. 271-277, 16-19 Oct. 2005

[8] A. Kvalbein et al., “Multiple Routing Configurations for Fast IP Network Recovery” IEEE/ACM Transactions on Networking, 17(2), 2009, pp. 473.486

[9] J. Wang, Y. Yang, L. Xiao, K. Nahrstedt, “Edge-based traffic engineering for OSPF networks”, *Comput. Netw.* 48, 4 (July 2005), 605-625

[10] A. Kvalbein, O. Lysne, “How can multi-topology routing be used for intradomain traffic engineering,” in Proceeding of IEEE SIGCOMM Workshop, 2007.

[11] N. Wang, K.-H. Ho, and G. Pavlou, “Adaptive Multi-Topology IGP Based Traffic Engineering with Near-Optimal Network Performance,” in IFIP-TC6 Networking Conference (Networking), Singapore, May 2008.

[12] Wang et al. , “An Overview of Routing Optimization for Internet Traffic Engineering”, in IEEE Communications Surveys volume 10, N°1, 2008.

[13] A. Elwalid, C. Jin, S. Low, and I. Widjaja, “MATE: MPLS adaptive traffic engineering”, In Proceedings of IEEE INFOCOM Conference, 2001.

[14] S. Kandula, D. Katabi, B. Davie, and A. Charny, “Walking the tightrope: responsive yet stable traffic engineering”, In Proceedings of ACM SIGCOMM, 2005.

[15] Gojmerac I., Reichl P., Jansen L., “Towards low-complexity Internet traffic engineering: The Adaptive Multi-Path algorithm”, (2008) Computer Networks, 52 (15), pp. 2894-2907.

[16] S. Fischer, N. Kammenhuber, and A. Feldmann, “Replex: dynamic traffic engineering based on wardrop routing policies,” in Proceedings of the 2006 ACM CoNEXT conference (CoNEXT ’06), Lisboa, Portugal, December 2006, pp. 1–12.

[17] The Abilene topology and traffic matrices dataset available at h t t p ://w w w . cs . utexas . edu / ~ yzhang / research /AbileneTM