CTE: Cost-Effective Intra-domain Traffic Engineering

Baobao Zhang[1,2] Jun Bi[1,2] Jianping Wu[1,2] Fred Baker[3]
zbb@netarchlab.tsinghua.edu.cn junbi@tsinghua.edu.cn jianping@cernet.edu.cn fred@cisco.com
1, Institute for Network Sciences and Cyberspace, Tsinghua University
2, Tsinghua National Laboratory for Information Science and Technology(TNList) ; 3, Cisco Systems Inc., USA

Categories and Subject Descriptors

C.2.2 **[Computer-Communication Networks]:** Network Protocols

Keywords

Traffic engineering, Access control list, static routes

1. PROBLEM STATEMENT

Internet Traffic Engineering (TE) is a very important research topic for production networks, where TE is typically formulated into minimizing the maximum link utilization (MLU). For example, with increasing IPv6 traffic, the backbone network of CERNET2[1] is suffering big traffic pressure on some critical long-haul links under the traditional shortest-path routing. Taking CERNET2 as an example, we develop a practical and cost-effective intra-domain TE method in this paper. Existing intra-domain TE proposals, such as MPLS-based proposal [2], OpenFlow-based proposal [3] and reactive-TE proposal [4], need to make great modifications to routers, and even have to replace existing routers with brand new routers, which makes the deployment be very high-cost. Other TE proposals, such as the work [5], optimize OSPF weights for the goal of TE. Although these methods have low deployment cost, they need to frequently change OSPF weights based on dynamically changed traffic matrices. Each change to OSPF weights may lead to transient routing loops or traffic disruption. Frequent routing loops or traffic disruption is intolerable for network operators. Therefore, in this paper, we propose a new intra-domain TE method named CTE. CTE works in an OSPF/IS-IS network, where shortest-path routing is run. The core idea of CTE is to use loop-free next hops [7] to route some traffic, which is realized by configuring some static routes and some access control list (ACL) rules. CTE does not make any modifications to existing routers. We only need to develop a remote control program that can configure static routes and ACL rules on routers. In addition, CTE can be incrementally deployed.

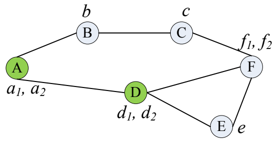

Figure 1. Network topology

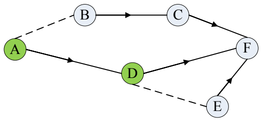

Figure 2. Shortest-path tree to F

Table 1. Prefix-granularity traffic matrix

$a_1 \rightarrow f_2$: 10;	$a_2 \rightarrow f_2$: 10
$a_2 \rightarrow f_1$: 30;	$d_1 \rightarrow f_1$: 10
$d_2 \rightarrow f_1$: 40;	$c \rightarrow f_1$: 20

Table 2. Node-granularity traffic matrix

$A \rightarrow F$: 50
$D \rightarrow F$: 50
$C \rightarrow F$: 20

We now use an example to illustrate how CTE works. Figure 1 is a network topology composed a set of nodes and a set of links. The lowercase letters beside a node denote the prefixes associated with the node. We suppose that the capacity of each link is 100 units and the weight of each link is 1. We give the shortest-path tree towards the destination *F* as shown in Figure 2. In Figure 2, the solid arrows direct the next hops towards the destination *F* and the dash lines will not be used by the traditional shortest-path routing towards the destination *F*. Table 1 shows a prefix-granularity traffic matrix. Table 2 shows the corresponding node-granularity traffic matrix. Under this traffic matrix, the MLU achieved by the traditional shortest-path routing is 1.0, which is high. We now show how CTE reduces the MLU. We suppose that Node *A* and Node *D* are two nodes that CTE can control. Node *B* is a loop-free next hop of Node *A* towards *F*. Node *E* is a loop-free next hop of Node *D* towards *F*. For CTE, the available next hops of Node *A* towards *F* are {*D, B*} and the available next hops of Node *D* towards *F* are {*F, E*}. Based on the node-granularity traffic matrix and the shortest-path routing, we can calculate the optimal splitting ratios across the available next hops of each CTE node towards each destination. We will give the specific optimization model of this step in Section 2. In this example, the optimal splitting ratio across the available next hops {*D, B*} of Node *A* towards *F* is 3:2 in order. The optimal splitting ratio across the available next hops {*F, E*} of Node *D* towards *F* is *1:1*. Using these splitting ratios, the minimum MLU, which is 0.4, can be achieved. The next step of CTE is to configure static routes and ACL rules to realize the calculated splitting ratios as close as possible. In this example, to realize the calculated splitting ratios, we only need to configure a static route '*reaching the destination* f_2 *via B*' on Node *A* and configure an ACL rule '*reaching the destination* f_1 *via E for the source* d_2' on Node *D*. Static routes and ACL rules have higher priority than OSPF routes. The MLU under the hybrid routers is 0.4, which the minimum. In addition, we design a mechanism of updating static routes and ACL rules to avoid routing loops and traffic disruption.

2. CTE APPROACH

We first develop an optimization model of calculating splitting ratios under a whole traffic matrix as shown in Table 3. This model is named TEW model for short. We now explain the TEW model. $V = \{1,2,\dots,|V|\}$ is the set of nodes. $E = \{1,2,\dots,|E|\}$ is the set of links. We use c_e to denote the capacity of a link $e \in E$. $\{d_{ij}\}$ is a traffic matrix, where d_{ij} denotes the size of traffic from Node *i* to Node *j*. $\{g_{jk} \geq 0 \mid j \in V, k \in E\}$ is the set of routing variables, where g_{jk} denotes the size of traffic destined to Node *j* over Link *k*. The splitting ratio over Link *e* on Node *i* is $\frac{g_{je}}{\sum_{k \in OUT(i)} g_{jk}}$. $Ava(i,j)$ denotes the available next hops of Node *i* towards Node *j*. If Node *i* is a legacy node, the available next hops of Node *i* are only the shortest-path next hops of Node *i* towards Node *j*. If Node *i* is a CTE node, the available next hops of Node *i* are the shortest-path next hops and the loop-free next

SIGCOMM'14, August 17–22, 2014, Chicago, Illinois, USA.
ACM 978-1-4503-2836-4/14/08.
http://dx.doi.org/10.1145/2619239.2631438

Table 3. TE model under a whole traffic matrix (TEW model)

$$\textbf{min}\ \alpha + \delta \textstyle\sum_{j \in V, k \in E} \left(\frac{g_{jk}}{|E| * c_k} \right),\ \textbf{subject}\ \text{to:}$$

$$\textstyle\sum_{k \in OUT(i)} g_{jk} - \sum_{k \in IN(i)} g_{jk} = d_{ij} \quad \forall i \in V,\ j \in V, and\ i \neq j$$

$$\textstyle\sum_{k \in OUT(i) - Ava(i,j)} g_{jk} = 0 \quad \forall i \in V,\ j \in V, and\ i \neq j$$

$$\textstyle\sum_{j \in V} \left(g_{je} \right) \leq \alpha c_e \ \ \forall e \in E$$

hops of Node i towards Node j. $OUT(i)$ denotes the set of outgoing links of Node i. $IN(i)$ denotes the set of incoming links of Node i. α is the MLU variable. Note that only $\{g_{jk}\}$ and α are variables. All the other signs are constants. δ is a small enough constant. The traffic matrix can be collected by deploying NetFlow on each node. Inspired by the work [3], we develop a model of calculating optimal splitting ratios under the situation that NetFlow is only deployed on the routers that CTE can control. This model is named TEP model for short. We refer readers to our technical report [6] for the TEP model. We now give the performance gap between the TEW model and the TEP model.

Theorem 1. We suppose that the MLU achieved by splitting ratios calculated by the TEW model is r_1 and the MLU achieved by splitting ratios calculated by the TEP model is r_2. $r_1 \leq r_2 \leq \frac{1}{1-p} r_1$ must stand, where p is the maximum packet loss rate among all the ingress-egress pairs.

Proof. See our technical report [6].

In the previous paragraphs, we have described how to compute proper splitting ratios across multiple next hops. We can configure static routes and ACL to implement calculated splitting ratios. In this paper, we mainly focus on how to update static routes and ACL rules without producing any transient routing loops and traffic disruption. For description simplicity, we use $f(S, L, D)$ to denote the fraction of traffic that is routed by the node S via its neighbor L towards the destination D. Some static routes and ACL rules will be configured to achieve the routing fraction $f(S, L, D)$. Our TEW and TEP models can guarantee that if $f(S, L, D) > 0, f(L, S, D) = 0$ must stand. If some static routes and ACL rules associated with $f(S, L, D) > 0$ are installed before the static routes and ACL rules associated with $f(L, S, D) > 0$ are completely removed, transient routing loops may occur. Therefore, we design a reliable update mechanism as follows.

Reliable update mechanism: (a) we should remove all the static routes and ACL rules associated with $f(S, L, D) > 0$ before we install some static routes and ACL rules associated with $f(L, S, D) > 0$. (b) if $f(L, S, D) = 0$, we can directly install or remove any static routes and ACL rules associated with $f(S, L, D)$.

Theorem 2. The reliable update mechanism will not produce any routing loops and traffic disruption. Using the reliable update mechanism, we can finish the updating in at most two steps.

Proof. See our technical report [6].

3. EVALUATIONS

All the routers in the CERNET2 backbone have deployed NetFlow. We collected a range of traffic matrices from 2013-02-19 22:10:00 to 2013-03-26 15:20:00 under the traditional OSPF routing. In addition, we collected the capacity of each link and the OSPF weight of each link. The traffic matrices are packaged per five minutes. In the simulations of this paper, we assume that our desired splitting ratios can be achieved by configuring static routes and ACL, which we will evaluate in the future. We applied splitting ratios computed by CTE based on the previous five-minute traffic matrix to the next five-minute traffic matrix and calculated the MLU reduction ratio, which is measured by the ratio of the MLU achieved by the optimized routing (such as CTE) to the original MLU achieved by OSPF. Figure 3 shows results of the average MLU reduction ratio during all the five-minute time intervals with the number of nodes that CTE can control. We find that if we only upgrade two nodes to support CTE, the average MLU reduction ratio will be 0.75. In addition, we give the CDF curves of the MLU reduction ratios as shown in Figure 4. 'CTE 2' denotes the results of upgrading two routers. 'CTE all' denotes the results of upgrading all the routers. 'Limit' denotes the TE limit results, i.e. assuming that MPLS is universally deployed, desired splitting ratios can be realized and traffic matrices are known in advance. From Figure 4, we find that if CTE is universally deployed, the near optimal TE can be achieved.

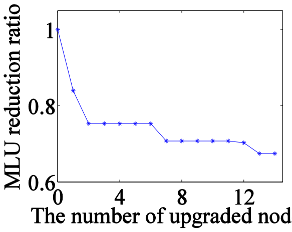

Figure 3. Partially deployed results

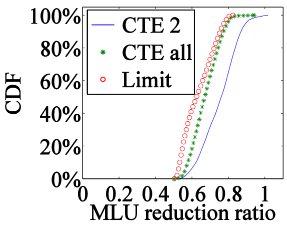

Figure 4. CDF curves

4. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a promising TE method named CTE. Without making any modifications to existing routers, CTE only needs a remote control program to configure static routes and ACL rules. Moreover, CTE is incrementally deployable. Our evaluation results show that CTE achieves near optimal TE performance when fully deployed, and achieves considerable TE performance when partially deployed. Some future work include: how to use minimum number of static routes and ACL rules to achieve the desired splitting ratios and how to incrementally update static routes and ACL rules.

5. ACKNOWLEDGMENTS

This research is supported by the National High-tech R&D Program ("863" Program) of China (No.2013AA013505), the National Science Foundation of China (No.61161140454), and Tsinghua-cisco joint lab research project. We also thank Kevin Yin, Zhiyan Zheng, Jiahai Yang, Lujing Sun and Chen Sun for their valuable comments, data collecting and processing. Jun Bi is the corresponding author.

6. REFERENCES

[1] Jianping Wu, et. al,, CNGI-CERNET2: an IPv6 deployment in China, ACM SIGCOMM Computer Communication Review, Vol.41, No.2, pp.48-52, 2011

[2] Hao Wang, et al., COPE: traffic engineering in dynamic networks, ACM SIGCOMM CCR, Vol. 36. No. 4., pp.99-100, 2006

[3] Sugam Agarwal, et al., Traffic engineering in software defined networks, INFOCOM13, pp. 2211-2219, 2013

[4] Amund Kvalbein, et al., Multipath load-adaptive routing: Putting the emphasis on robustness and simplicity, ICNP2009, pp.203-212, 2009

[5] Bernard Fortz, et al., Internet traffic engineering by optimizing OSPF weights, INFOCOM 00, Vol.2, pp.519-528, 2000

[6] Technical Report: http://netarchlab.tsinghua.edu.cn/~zbb/THU-NetArchLab-RTG-TR-CTE-20140515.pdf

[7] Alia K. Atlas, et al., Basic specification for IP fast-reroute: loop-free alternates, RFC 5286, 2008