

TEAS Working Group
Internet-Draft
Intended status: Informational
Expires: April 27, 2020

A. Wang
China Telecom
X. Huang
C. Kou
BUPT
Z. Li
China Mobile
P. Mi
Huawei Technologies
October 25, 2019

Scenarios and Simulation Results of PCE in Native IP Network draft-ietf-teas-native-ip-scenarios-11

Abstract

Requirements for providing the End to End(E2E) performance assurance are emerging within the service provider networks. While there are various technology solutions, there is no single solution that can fulfill these requirements for a native IP network. In particular, there is a need for a universal (E2E) solution that is simultaneously applicable for both intra- and inter-domain scenarios.

One feasible E2E traffic engineering solution is the addition of central control in a native IP network. This document describes various complex scenarios and simulation results when applying the Path Computation Element (PCE) in a native IP network. This solution, referred to as Centralized Control Dynamic Routing (CCDR), integrates the advantage of using distributed protocols and the power of a centralized control technology, providing traffic engineering for native IP networks in a manner that applies equally to intra- and inter-domain scenarios.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 27, 2020.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
2.	Terminology	4
3.	CCDR Scenarios	4
3.1.	QoS Assurance for Hybrid Cloud-based Application	4
3.2.	Link Utilization Maximization	5
3.3.	Traffic Engineering for Multi-Domain	6
3.4.	Network Temporal Congestion Elimination	7
4.	CCDR Simulation	7
4.1.	Case Study for CCDR algorithm	8
4.2.	Topology Simulation	10
4.3.	Traffic Matrix Simulation	10
4.4.	CCDR End-to-End Path Optimization	11
4.5.	Network Temporal Congestion Elimination	12
5.	CCDR Deployment Consideration	14
6.	Security Considerations	14
7.	IANA Considerations	15
8.	Contributors	15
9.	Acknowledgement	15
10.	References	15
10.1.	Normative References	15
10.2.	Informative References	16
	Authors' Addresses	16

[1.](#) Introduction

A service provider network is composed of thousands of routers that run distributed protocols to exchange the reachability information. The path for the destination network is mainly calculated, and controlled, by the distributed protocols. These distributed

protocols are robust enough to support most applications, but have some difficulties supporting the complexities needed for traffic engineering applications, e.g. E2E performance assurance, or maximizing the link utilization within an IP network.

Multiprotocol Label Switching (MPLS) using Traffic Engineering (TE) technology (MPLS-TE)[[RFC3209](#)] is one solution for traffic engineering networks but it introduces an MPLS network and related technology which would be an overlay of the IP network. MPLS-TE technology is often used for Label Switched Path (LSP) protection and complex path set-up within a domain.

It has not been widely deployed for meeting E2E (especially in inter-domain) dynamic performance assurance requirements for an IP network.

Segment Routing [[RFC8402](#)] is another solution that integrates some advantages of using a distributed protocol and a centrally control technology, but it requires the underlying network, especially the provider edge router, to do a label push and pop action in-depth, and adds complexity when coexisting with the Non-Segment Routing network. Additionally, it can only maneuver the E2E paths for MPLS and IPv6 traffic via different mechanisms.

Deterministic Networking (DetNet)[[RFC8578](#)] is another possible solution. It is primarily focused on providing bounded latency for a flow and introduces additional requirements on the domain edge router. The current DetNet scope is within one domain. The use cases defined in this document do not require the additional complexity of deterministic properties and so differ from the DetNet use cases.

This draft describes several scenarios for a native IP network where a Centralized Control Dynamic Routing (CCDR) framework can produce qualitative improvement in efficiency without requiring a change of the data-plane behavior on the router. Using knowledge of BGP(Border Gateway Protocol) session-specific prefixes advertised by a router, the network topology and the near real time link utilization information from network management systems, a central PCE is able to compute an optimal path and give the underly routers the destination address to use to reach the BGP nexthop, such that the distributed routing protocol will use the computed path via traditional recursive lookup procedure. Some results from simulations of path optimization are also presented, to concretely illustrate a variety of scenarios where CCDR shows significant improvement over traditional distributed routing protocols.

This draft is the base document of the following two drafts: the universal solution draft, which is suitable for intra-domain and

inter-domain TE scenario, is described in [[I-D.ietf-teas-pce-native-ip](#)]; the related protocol extension contents is described in [[I-D.ietf-pce-pcep-extension-native-ip](#)]

2. Terminology

This document uses the following terms defined in [[RFC5440](#)]: PCE.

The following terms are defined in this document:

- o BRAS: Broadband Remote Access Server
- o CD: Congestion Degree
- o CR: Core Router
- o CCDR: Centralized Control Dynamic Routing
- o E2E: End to End
- o IDC: Internet Data Center
- o MAN: Metro Area Network
- o QoS: Quality of Service
- o SR: Service Router
- o TE: Traffic Engineering
- o UID: Utilization Increment Degree
- o WAN: Wide Area Network

3. CCDR Scenarios

The following sections describe various deployment scenarios where applying the CCDR framework is intuitively expected to produce improvements, based on the macro-scale properties of the framework and the scenario.

3.1. QoS Assurance for Hybrid Cloud-based Application

With the emergence of cloud computing technologies, enterprises are putting more and more services on a public oriented cloud environment, but keeping core business within their private cloud. The communication between the private and public cloud sites will span the Wide Area Network (WAN) network. The bandwidth requirements

between them are variable and the background traffic between these two sites varies over time. Enterprise applications require assurance of the E2E Quality of Service(QoS) performance on demand for variable bandwidth services.

CCDR, which integrates the merits of distributed protocols and the power of centralized control, is suitable for this scenario. The possible solution framework is illustrated below:

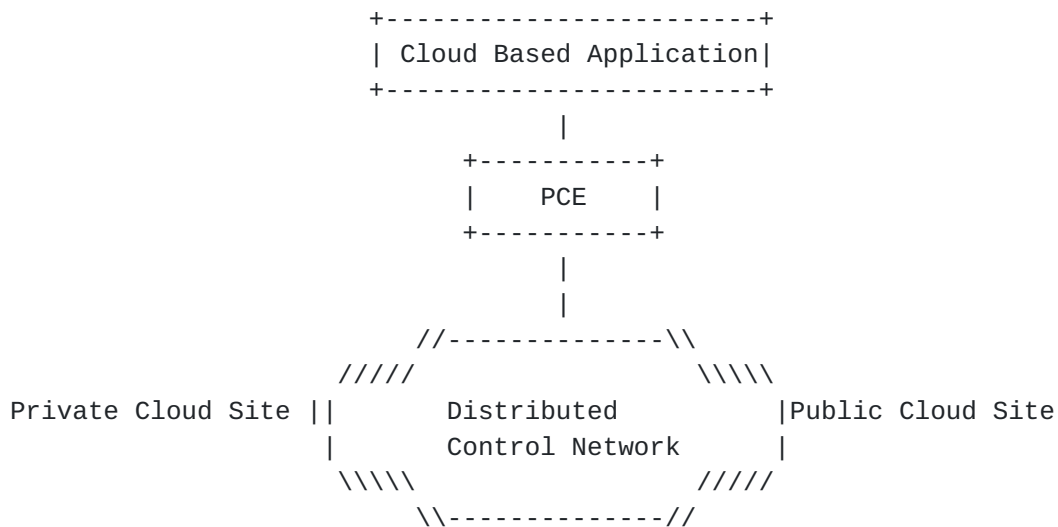


Figure 1: Hybrid Cloud Communication Scenario

As illustrated in Figure 1, the source and destination of the "Cloud Based Application" traffic are located at "Private Cloud Site" and "Public Cloud Site" respectively.

By default, the traffic path between the private and public cloud site is determined by the distributed control network. When application requires the E2E QoS assurance, it can send these requirements to the PCE, and let the PCE compute one E2E path which is based on the underlying network topology and the real traffic information, to accommodate the application's QoS requirements. [Section 4.4](#) of this document describes the simulation results for this use case.

3.2. Link Utilization Maximization

Network topology within a Metro Area Network (MAN) is generally in a star mode as illustrated in Figure 2, with different devices connected to different customer types. The traffic from these customers is often in a tidal pattern, with the links between the Core Router(CR)/Broadband Remote Access Server(BRAS) and CR/Service Router(SR) experiencing congestion in different periods, because the

subscribers under BRAS often use the network at night, and the leased line users under SR often use the network during the daytime. The link between BRAS/SR and CR must satisfy the maximum traffic volume between them, respectively, and this causes these links often to be under-utilized.

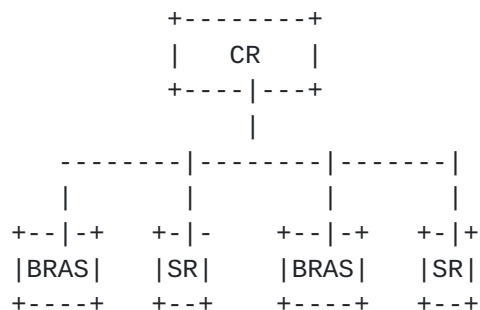


Figure 2: Star-mode Network Topology within MAN

If we consider connecting the BRAS/SR with a local link loop (which is usually lower cost), and control the overall MAN topology with the CCDR framework, we can exploit the tidal phenomena between the BRAS/CR and SR/CR links, maximizing the utilization of these central trunk links (which are usually higher cost than the local loops).

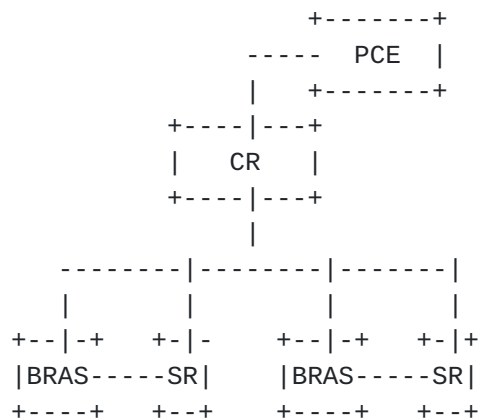


Figure 3: Link Utilization Maximization via CCDR

3.3. Traffic Engineering for Multi-Domain

Service provider networks are often comprised of different domains, interconnected with each other, forming a very complex topology as illustrated in Figure 4. Due to the traffic pattern to/from the MAN and IDC, the utilization of the links between them are often asymmetric. It is almost impossible to balance the utilization of these links via a distributed protocol, but this unbalance can be overcome utilizing the CCDR framework.

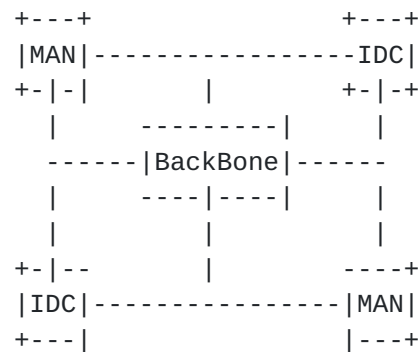


Figure 4: Traffic Engineering for Complex Multi-Domain Topology

A solution for this scenario requires the gathering of NetFlow information, analysis of the source/destination AS, and determining what is the main cause of the congested link(s). After this, the operator can use the external Border Gateway Protocol(eBGP) sessions to schedule the traffic among the different domains according to the solution described in CCDR framework.

3.4. Network Temporal Congestion Elimination

In more general situations, there are often temporal congestion within the service provider's network, for example due to daily or weekly periodic bursts, or large events that are scheduled well in advance. Such congestion phenomena often appear regularly, and if the service provider has methods to mitigate it, it will certainly improve their network operations capabilities and increase satisfaction for their customers. CCDR is also suitable for such scenarios, as the controller can schedule traffic out of the congested links, lowering the utilization of them during these times. [Section 4.5](#) describes the simulation results of this scenario.

4. CCDR Simulation

The following sections describe a specific case study to illustrate the workings of the CCDR algorithm with concrete paths/metrics, as well as a procedure for generating topology and traffic matrices and the results from simulations applying CCDR for E2E QoS (assured path and congestion elimination) over the generated topologies and traffic matrices. In all cases examined, the CCDR algorithm produces qualitatively significant improvement over the reference (OSPF) algorithm, suggesting that CCDR will have broad applicability.

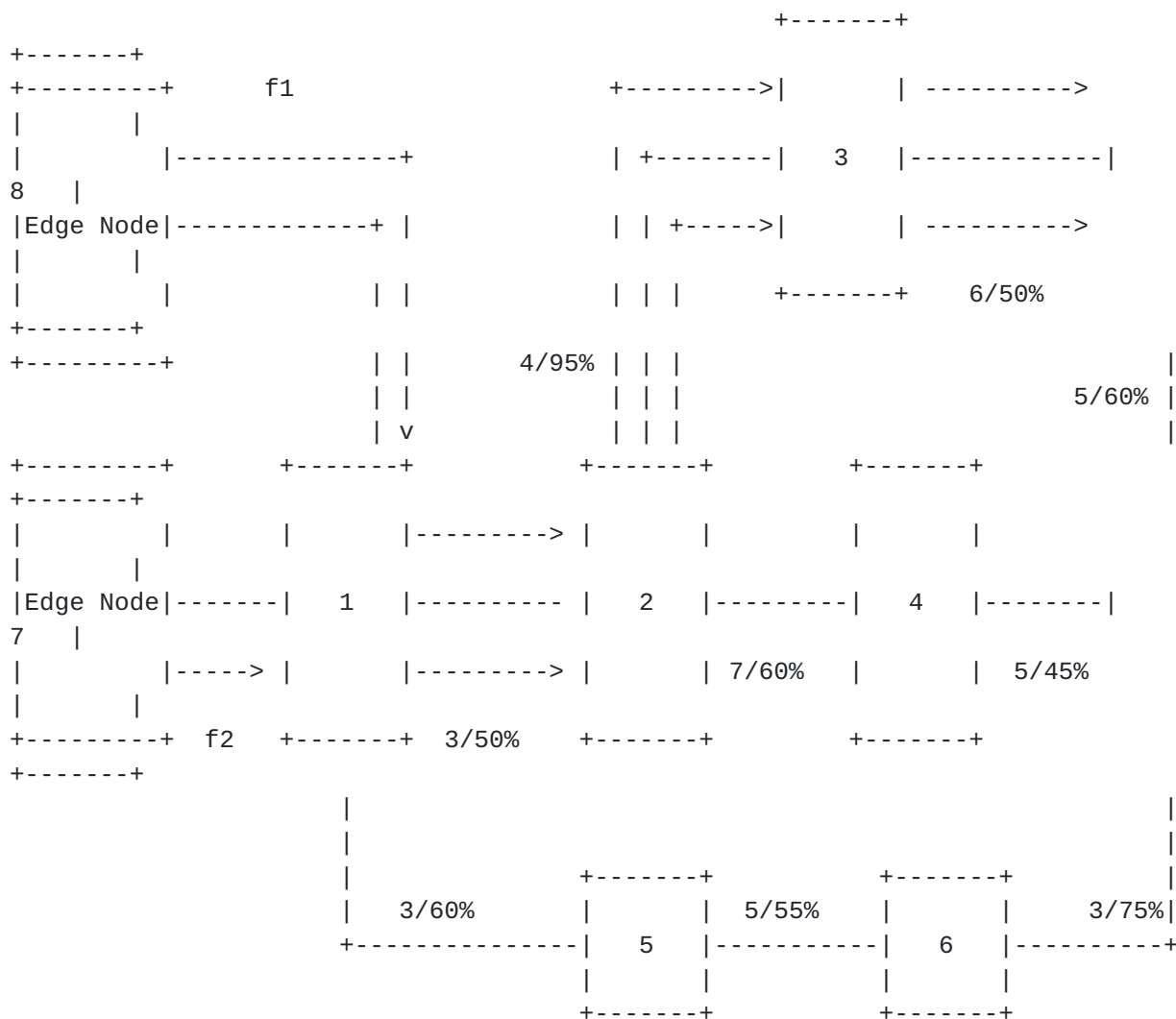
The structure and scale of the simulated topology is similar to that of the real networks. Multiple different traffic matrices were generated to simulate different congestion conditions in the network,

but only one of them is illustrated since the others produce similar results.

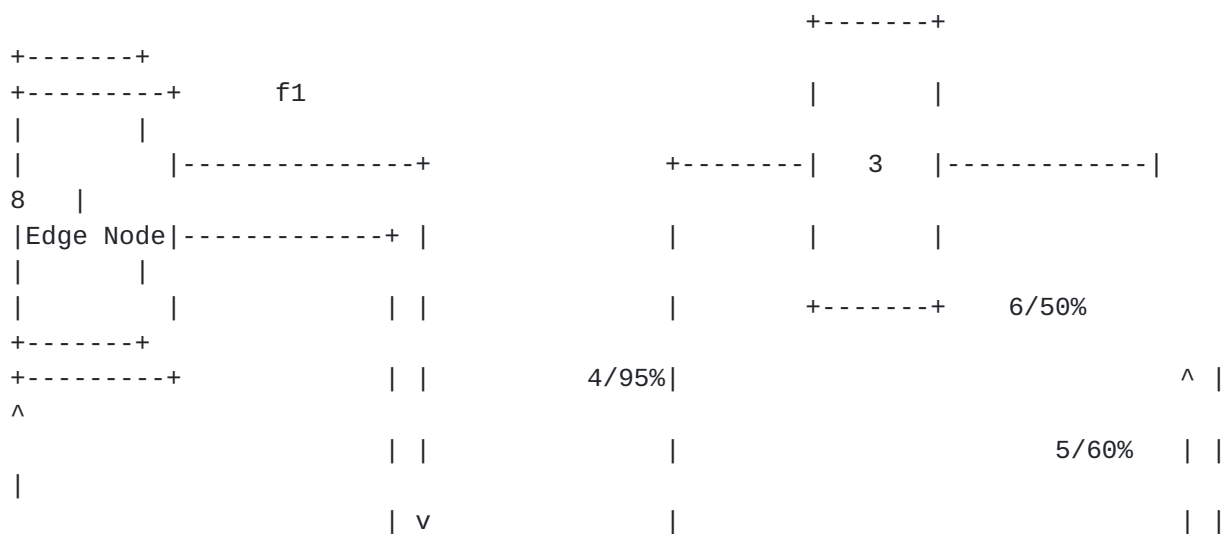
4.1. Case Study for CCDR algorithm

In this section we consider a specific network topology for a worked case study, examining the path selected by OSPF and CCDR and evaluating how and why the paths differ. Figure 5 depicts the topology of the network in question. There are 8 forwarding devices in the network. The original cost and utilization are marked on it, as shown in the figure. For example, the original cost and utilization for the link (1,2) are 3 and 50% respectively. There are two flows: f1 and f2. Both of these two flows are from node 1 to node 8. For simplicity, it is assumed that the bandwidth of the link in the network is 10Mb/s. The flow rate of f1 is 1Mb/s, and the flow rate of f2 is 2Mb/s. The threshold of the link in congestion is 90%.

If OSPF protocol (ISIS is similar, because it also use the Dijkstra's algorithm) is applied in the network, which adopts Dijkstra's algorithm, the two flows from node 1 to node 8 can only use the OSPF path (p1: 1->2->3->8). It is because Dijkstra's algorithm mainly considers original cost of the link. Since CCDR considers cost and utilization simultaneously, the same path as OSPF will not be selected due to the severe congestion of the link (2,3). In this case, f1 will select the path (p2: 1->5->6->7->8) since the new cost of this path is better than that of OSPF path. Moreover, the path p2 is also better than the path (p3: 1->2->4->7->8) for for flow f1. However, f2 will not select the same path since it will cause the new congestion in the link (6,7). As a result, f2 will select the path (p3: 1->2->4->7->8).



(a) Dijkstra's Algorithm(OSPF/ISIS)



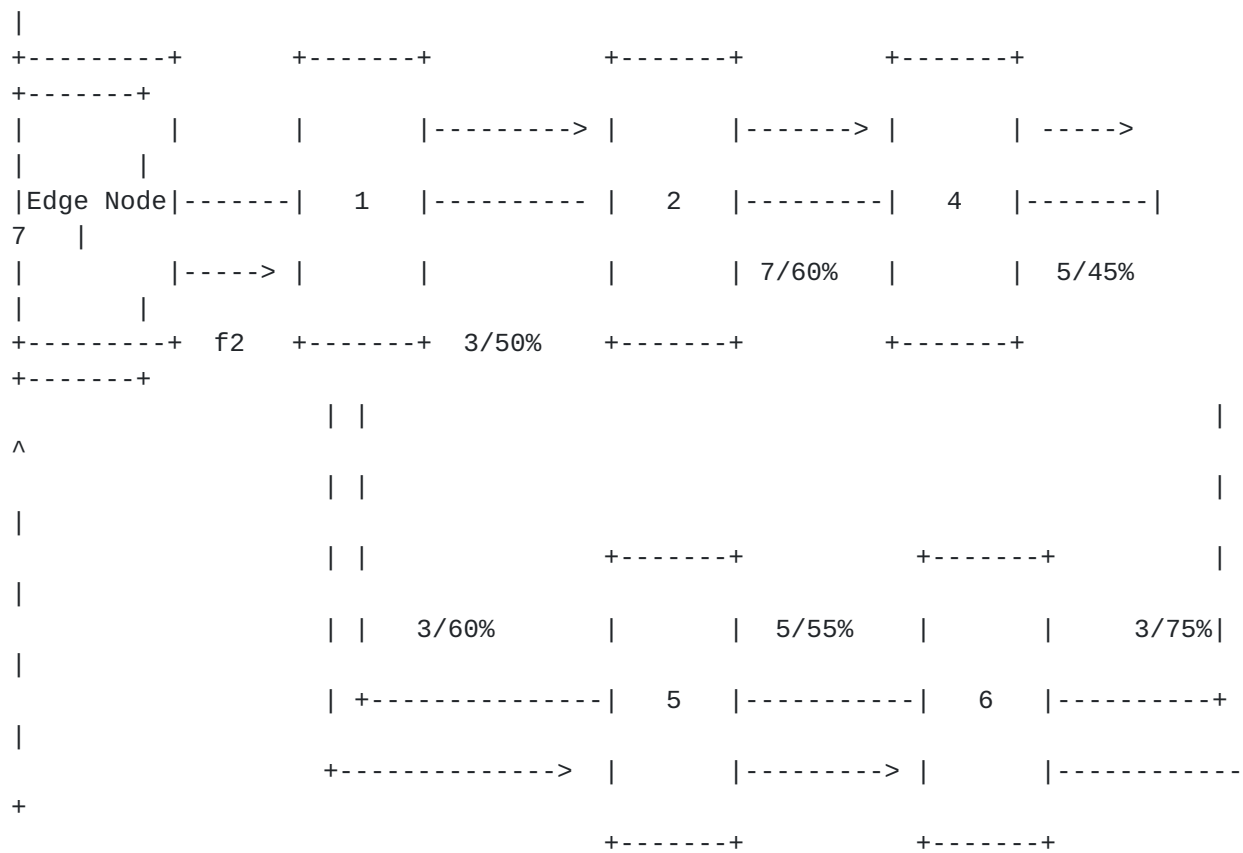


Figure 5: Case Study for CCDR's Algorithm

4.2. Topology Simulation

Moving on from the specific case study, we now consider a class of networks more representative of real deployments, with a fully-linked core network that serves to connect edge nodes, which themselves connect to only a subset of the core. An example of such a topology is shown in Figure 6, for the case of 4 core nodes and 5 edge nodes. The CCDR simulations presented in this work use topologies involving 100 core nodes and 400 edge nodes. While the resulting graph does not fit on this page, this scale of network is similar to what is deployed in production environments.

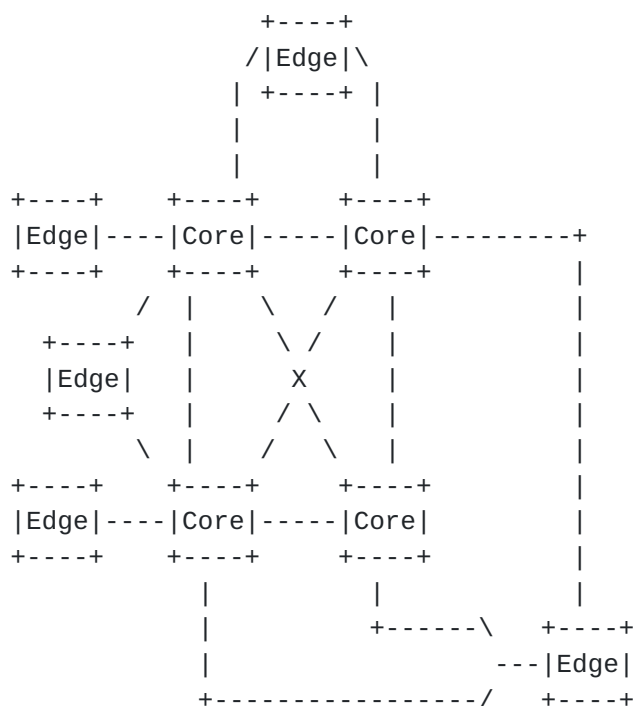


Figure 6: Topology of Simulation

For the simulations, the number of links connecting one edge node to the set of core nodes is randomly chosen between 2 to 30, and the total number of links is more than 20000. Each link has a congestion threshold, which can be arbitrarily set to (e.g.) 90% of the nominal link capacity without affecting the simulation results.

4.3. Traffic Matrix Simulation

For each topology, a traffic matrix is generated based on the link capacity of topology. It can result in many kinds of situations, such as congestion, mild congestion and non-congestion.

In the CCDR simulation, the dimension of the traffic matrix is 500*500 (100 core nodes plus 400 edge nodes). About 20% of links are overloaded when the Open Shortest Path First (OSPF) protocol is used in the network.

4.4. CCDR End-to-End Path Optimization

The CCDR E2E path optimization is to find the best path which is the lowest in metric value and for each link of the path is far below link's congestion threshold. Based on the current state of the network, the PCE within CCDR framework combines the shortest path algorithm with a penalty theory of classical optimization and graph theory.

Given a background traffic matrix, which is unscheduled, when a set of new flows comes into the network, the E2E path optimization finds the optimal paths for them. The selected paths bring the least congestion degree to the network.

The link Utilization Increment Degree (UID), when the new flows are added into the network, is shown in Figure 7. The first graph in Figure 7 is the UID with OSPF and the second graph is the UID with CCDR E2E path optimization. The average UID of the first graph is more than 30%. After path optimization, the average UID is less than 5%. The results show that the CCDR E2E path optimization has an eye-catching decrease in UID relative to the path chosen based on OSPF.

While real-world results invariably differ from simulations (for example, real-world topologies are likely to exhibit correlation in the attachment patterns for edge nodes to the core, which are not reflected in these results), the dramatic nature of the improvement in UID and the choice of simulated topology to resemble real-world conditions suggests that real-world deployments will also experience significant improvement in UID results.

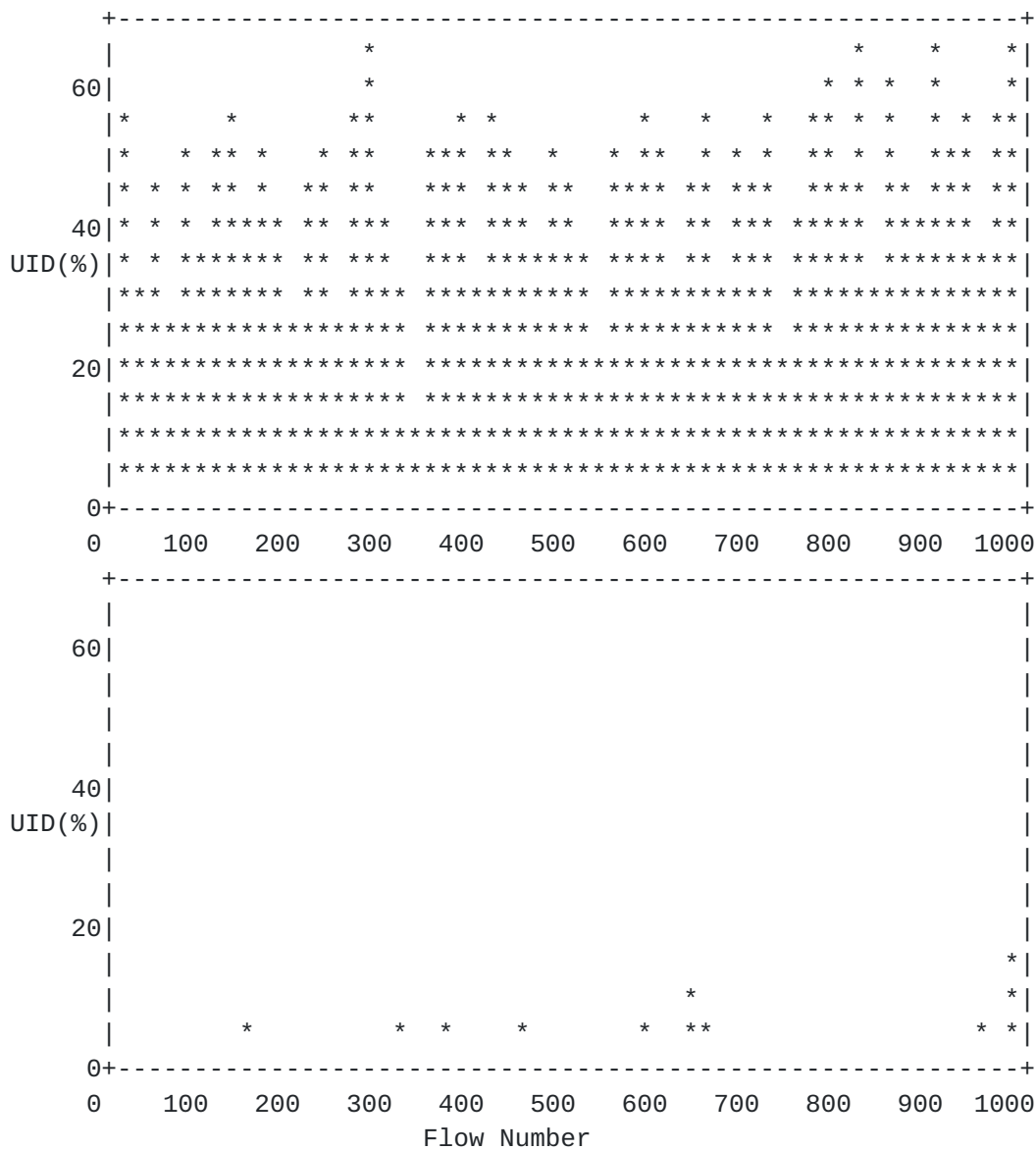


Figure 7: Simulation Result with Congestion Elimination

4.5. Network Temporal Congestion Elimination

During the simulations, different degrees of network congestion were considered. To examine the effect of CCDR on link congestion, we consider the Congestion Degree (CD) of a link, defined as the link utilization beyond its threshold.

The CCDR congestion elimination performance is shown in Figure 8. The first graph is the CD distribution before the process of congestion elimination. The average CD of all congested links is about 20%. The second graph shown in Figure 8 is the CD distribution after using the congestion elimination process. It shows that only 12 links among the total of 20000 links exceed the threshold, and all

the CD values are less than 3%. Thus, after scheduling of the traffic away from the congested paths, the degree of network congestion is greatly eliminated and the network utilization is in balance.

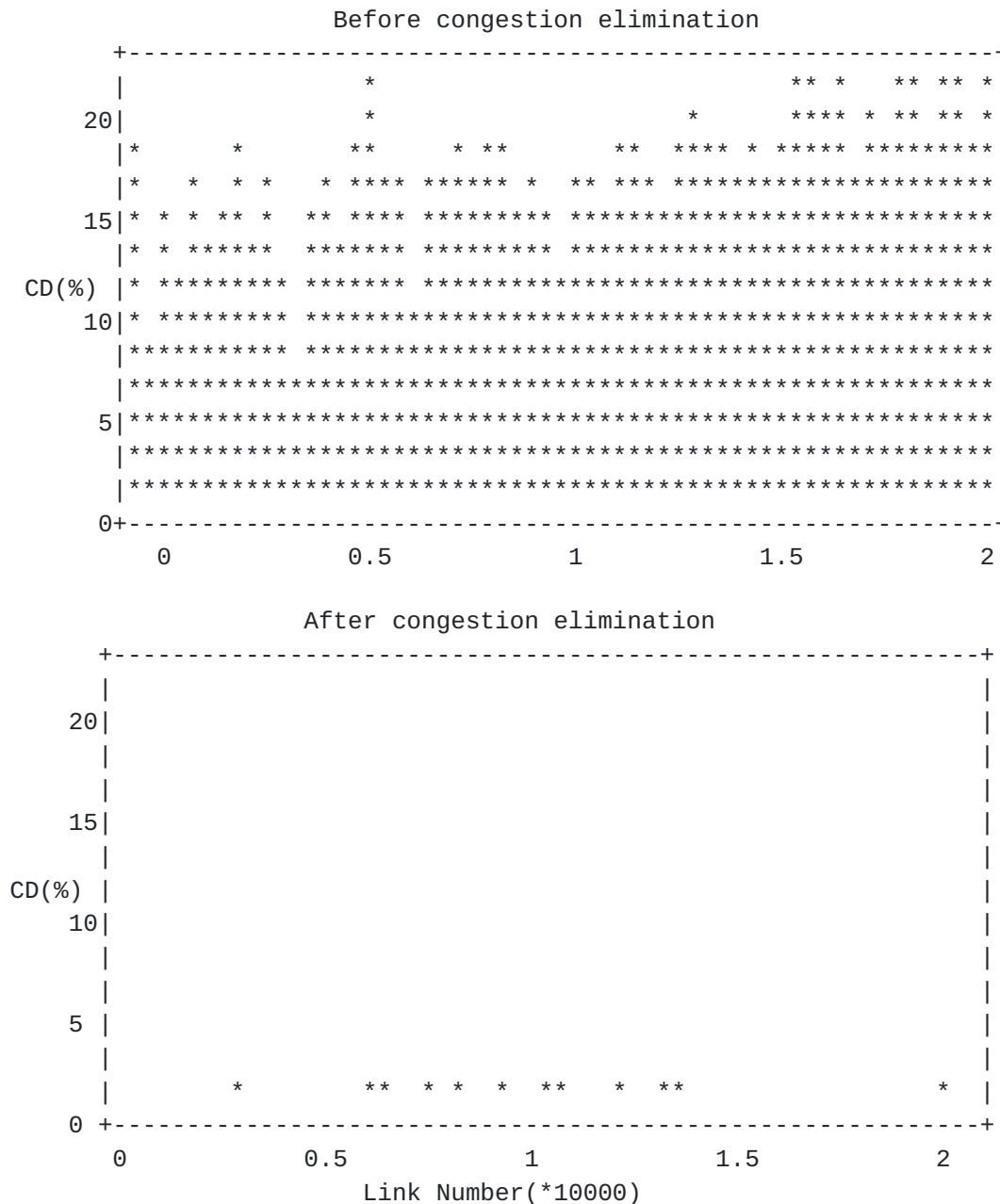


Figure 8: Simulation Result with Congestion Elimination

It is clear that using an active path-computation mechanism that is able to take into account observed link traffic/congestion, the occurrence of congestion events can be greatly reduced. Only when a preponderance of links in the network are near their congestion threshold will the central controller be unable to find a clear path, as opposed to when a static metric-based procedure is used, which

will produce congested paths once a single bottleneck approaches its capacity. More detailed information about the algorithm can be found in[PTCS] .

5. CCDR Deployment Consideration

The above CCDR scenarios and simulation results demonstrate that a single general solution can be found that copes with multiple complex situations. The specific situations considered are not known to have any special properties, so it is expected that the benefits demonstrated will have general applicability. Accordingly, the integrated use of a centralized controller for the more complex optimal path computations in a native IP network should result in significant improvements without impacting the underlay network infrastructure.

For intra-domain or inter-domain native IP TE scenarios, the deployment of a CCDR solution is similar, with the centralized controller being able to compute paths and no changes required to the underlying network infrastructure. This universal deployment characteristic can facilitate a generic traffic engineering solution, where operators do not need to differentiate between intra-domain and inter-domain TE cases.

To deploy the CCDR solution, the PCE should collect the underlay network topology dynamically, for example via BGP-LS[RFC7752]. It also needs to gather the network traffic information periodically from the network management platform. The simulation results show that the PCE can compute the E2E optimal path within seconds, thus it can cope with the change of underlay network on the scale of minutes. More agile requirements would need to increase the sample rate of underlay network and decrease the detection and notification interval of the underlay network. The methods to gather and decrease the latency of these information are out of the scope of this draft..

6. Security Considerations

This document considers mainly the integration of distributed protocols and the central control capability of a PCE. While it certainly can ease the management of network in various traffic engineering scenarios as described in this document, the centralized control also bring a new point that may be easily attacked. Solutions for CCDR scenarios need to consider protection of the PCE and communication with the underlay devices.

[RFC5440] and [[RFC8253](#)] provide additional information.

The control priority and interaction process should also be carefully designed for the combination of distributed protocol and central control. Generally, the central control instruction should have higher priority than the forwarding actions determined by the distributed protocol. When the communication between PCE and the underlay devices is not in function, the distributed protocol should take over the control right of the underlay network.

[[I-D.ietf-teas-pce-native-ip](#)] provide more considerations corresponding to the solution.

7. IANA Considerations

This document does not require any IANA actions.

8. Contributors

Lu Huang contributed to the content of this draft.

9. Acknowledgement

The author would like to thank Deborah Brungard, Adrian Farrel, Huaimo Chen, Vishnu Beeram and Lou Berger for their support and comments on this draft.

Thanks Benjamin Kaduk for his careful review and valuable suggestions to this draft. Also thanks Roman Danyliw, Alvaro Retana and Eric Vyncke for their views and comments.

10. References

10.1. Normative References

- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", [RFC 5440](#), DOI 10.17487/RFC5440, March 2009, <<https://www.rfc-editor.org/info/rfc5440>>.
- [RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", [RFC 7752](#), DOI 10.17487/RFC7752, March 2016, <<https://www.rfc-editor.org/info/rfc7752>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", [RFC 8253](#), DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.

10.2. Informative References

- [I-D.ietf-pce-pcep-extension-native-ip]
Wang, A., Khasanov, B., Cheruathur, S., Zhu, C., and S. Fang, "PCEP Extension for Native IP Network", [draft-ietf-pce-pcep-extension-native-ip-04](#) (work in progress), August 2019.
- [I-D.ietf-teas-pce-native-ip]
Wang, A., Zhao, Q., Khasanov, B., Chen, H., and R. Mallya, "PCE in Native IP Network", [draft-ietf-teas-pce-native-ip-04](#) (work in progress), August 2019.
- [PTCS]
Zhang, P., Xie, K., Kou, C., Huang, X., Wang, A., and Q. Sun, "A Practical Traffic Control Scheme With Load Balancing Based on PCE Architecture", IEEE Access 18526773, DOI 10.1109/ACCESS.2019.2902610, March 2019, <<http://ieeexplore.ieee.org/document/8657733>>.
- [RFC3209]
Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", [RFC 3209](#), DOI 10.17487/RFC3209, December 2001, <<https://www.rfc-editor.org/info/rfc3209>>.
- [RFC8402]
Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", [RFC 8402](#), DOI 10.17487/RFC8402, July 2018, <<https://www.rfc-editor.org/info/rfc8402>>.
- [RFC8578]
Grossman, E., Ed., "Deterministic Networking Use Cases", [RFC 8578](#), DOI 10.17487/RFC8578, May 2019, <<https://www.rfc-editor.org/info/rfc8578>>.

Authors' Addresses

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing, Beijing 102209
China

Email: wangaj3@chinatelecom.cn

Xiaohong Huang
Beijing University of Posts and Telecommunications
No.10 Xitucheng Road, Haidian District
Beijing
China

Email: huangxh@bupt.edu.cn

Caixia Kou
Beijing University of Posts and Telecommunications
No.10 Xitucheng Road, Haidian District
Beijing
China

Email: koucx@lsec.cc.ac.cn

Zhenqiang Li
China Mobile
32 Xuanwumen West Ave, Xicheng District
Beijing 100053
China

Email: li_zhenqiang@hotmail.com

Penghui Mi
Huawei Technologies
Tower C of Bldg.2, Cloud Park, No.2013 of Xuegang Road
Shenzhen, Bantian, Longgang District 518129
China

Email: mipenghui@huawei.com

