Stream:	Internet Engineering Task Force (IETF)				
RFC:	8735				
Category:	Informational				
Published:	February 2020				
ISSN:	2070-1721				
Authors:	A. Wang	X. Huang	C. Kou	Z. Li	P. Mi
	China Telecom	*BUPT*	*BUPT*	*China Mobile*	*Huawei Technologies*

RFC 8735
Scenarios and Simulation Results of PCE in a Native IP Network

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 can cover 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 document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8735.

Copyright Notice

Table of Contents

1. Introduction

A service provider network is composed of thousands of routers that run distributed protocols to exchange 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; however, they 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 TE networks, but it introduces an MPLS network along with related technology, which would be an overlay of the IP network. MPLS-TE technology is often used for Label Switched Path (LSP) protection and setting up complex paths 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 central control technology, but it requires the underlying network, especially the provider edge router, to do an in-depth label push and pop action while adding 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 document 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 to the data-plane behavior on the router. Using knowledge of the Border Gateway Protocol (BGP) 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 underlying 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 document is the base document of the following two documents: the universal solution document, which is suitable for intra-domain and inter-domain TE scenario, is described in [PCE-NATIVE-IP]; and the related protocol extension contents is described in [PCEP-NATIVE-IP-EXT].

2. Terminology

In this document, PCE is used as defined in [RFC5440]. The following terms are used as described here:

BRAS: Broadband Remote Access Server

CD: Congestion Degree

CR: Core Router

CCDR: Centralized Control Dynamic Routing

E2E: End to End

IDC: Internet Data Center

MAN: Metro Area Network

QoS: Quality of Service

SR: Service Router

TE: Traffic Engineering

UID: Utilization Increment Degree

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 while keeping core business within their private cloud. The communication between the private and public cloud sites spans the WAN. 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 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:

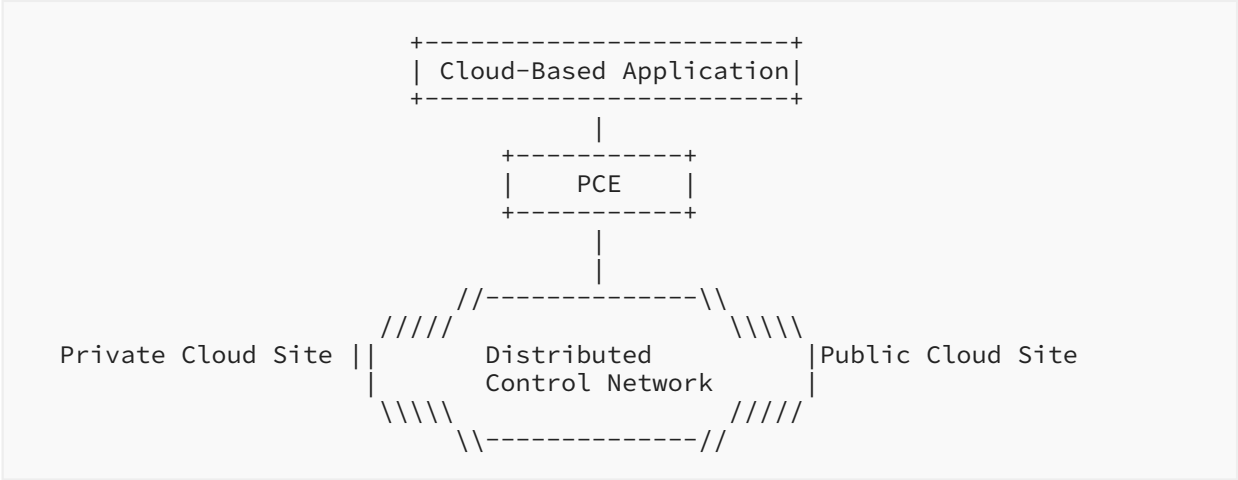

Figure 1: Hybrid Cloud Communication Scenario

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 an application requires 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 real traffic information, in order 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 due to subscribers under BRAS often using the network at night and the leased line users under SR often using the network during the daytime. The link between BRAS/SR and CR must satisfy the maximum traffic volume between them, respectively, which causes these links to often be underutilized.

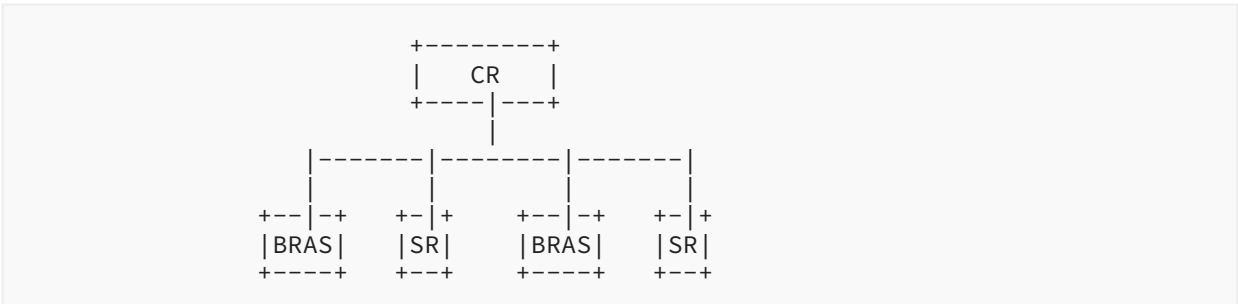

Figure 2: Star-Mode Network Topology within MAN

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).

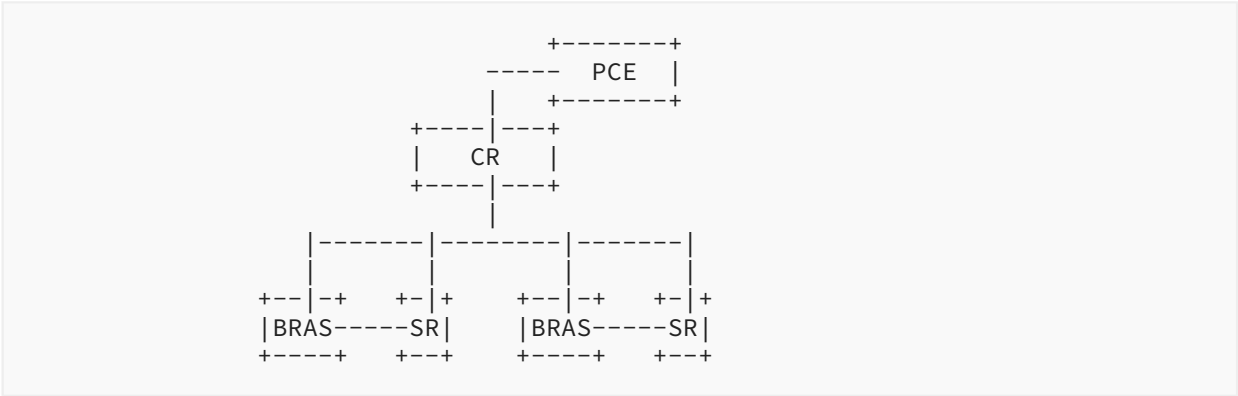

Figure 3: Link Utilization Maximization via CCDR

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.

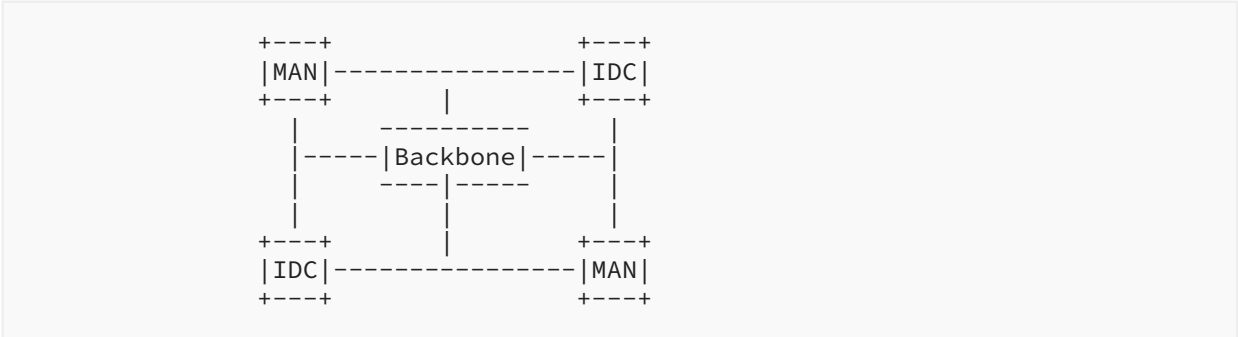

Figure 4: Traffic Engineering for Complex Multi-domain Topology

A solution for this scenario requires the gathering of NetFlow information, analysis of the source/destination autonomous system (AS), and determining what the main cause of the congested link (s) is. 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 the CCDR framework.

3.4. Network Temporal Congestion Elimination

In more general situations, there is 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 operation capabilities and increase satisfaction for customers. CCDR is also suitable for such scenarios, as the controller can schedule traffic out of the congested links, lowering their utilization 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. 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 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 this case. There are eight 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 10 Mb/s. The flow rate of f1 is 1 Mb/s and the flow rate of f2 is 2 Mb/s. The threshold of the link in congestion is 90%.

If the OSPF protocol, which adopts Dijkstra's algorithm (IS-IS is similar because it also uses Dijkstra's algorithm), is applied in the network, the two flows from node 1 to node 8 can only use the OSPF path (p1: 1->2->3->8). This is because Dijkstra's algorithm mainly considers the 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 the OSPF path. Moreover, the path p2 is also better than the path (p3: 1->2->4->7->8) for flow f1. However, f2 will not select the same path since it will cause new congestion in the link (6, 7). As a result, f2 will select the path (p3: 1->2->4->7->8).

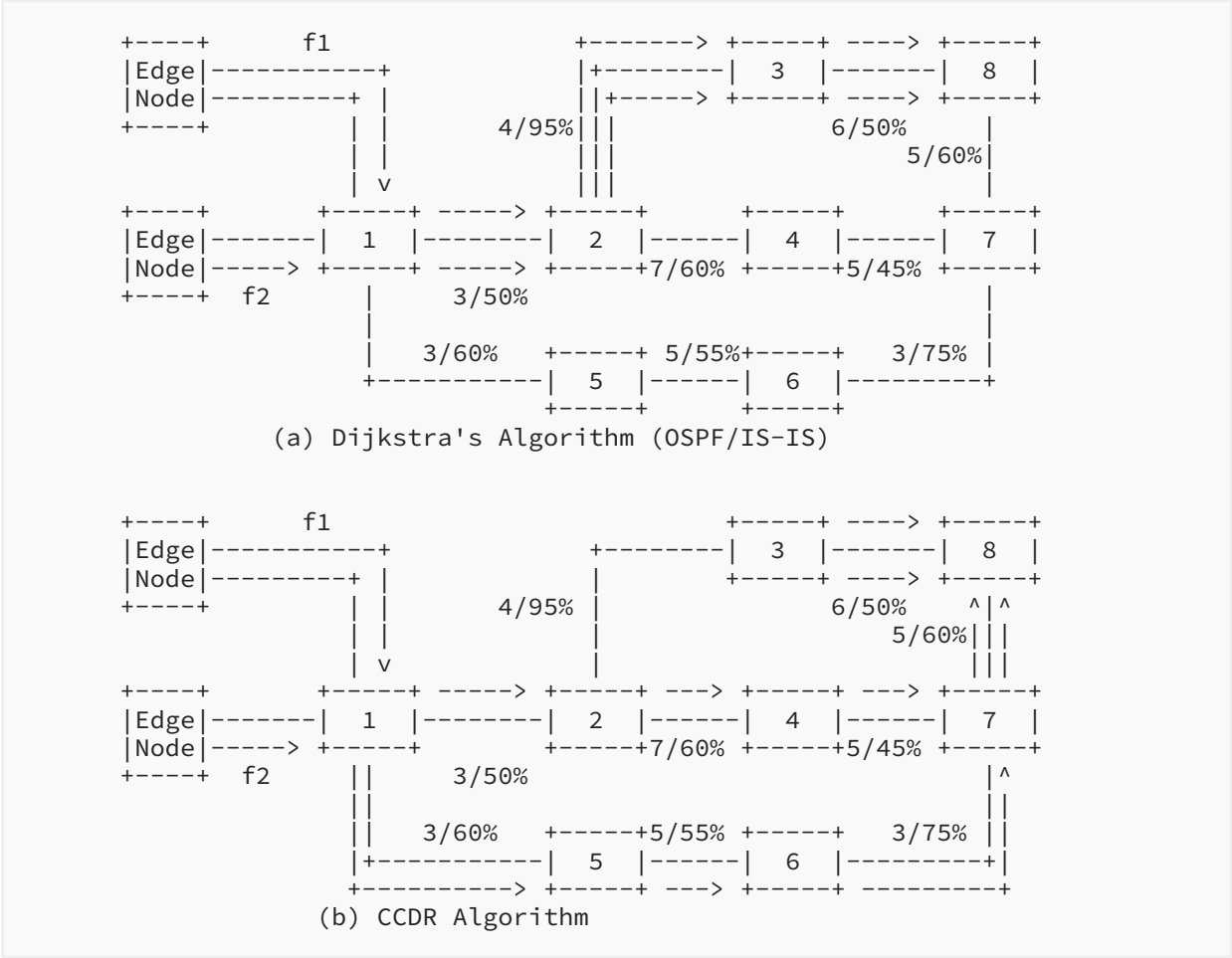

Figure 5: Case Study for CCDR's Algorithm

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.

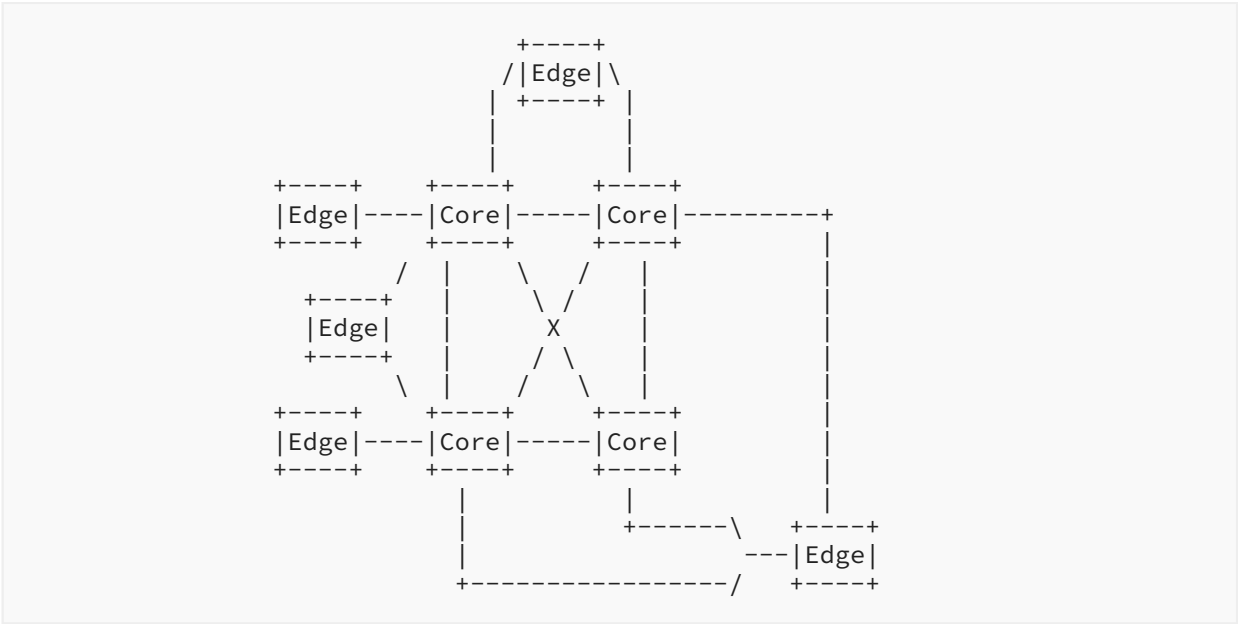

Figure 6: Topology of Simulation

For the simulations, the number of links connecting one edge node to the set of core nodes is randomly chosen between two and thirty, and the total number of links is more than 20,000. Each link has a congestion threshold, which can be arbitrarily set, for example, to 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 the 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 entails finding the best path, which is the lowest in metric value, as well as having utilization far below the congestion threshold for each link of the path. 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 suggest that real-world deployments will also experience significant improvement in UID results.

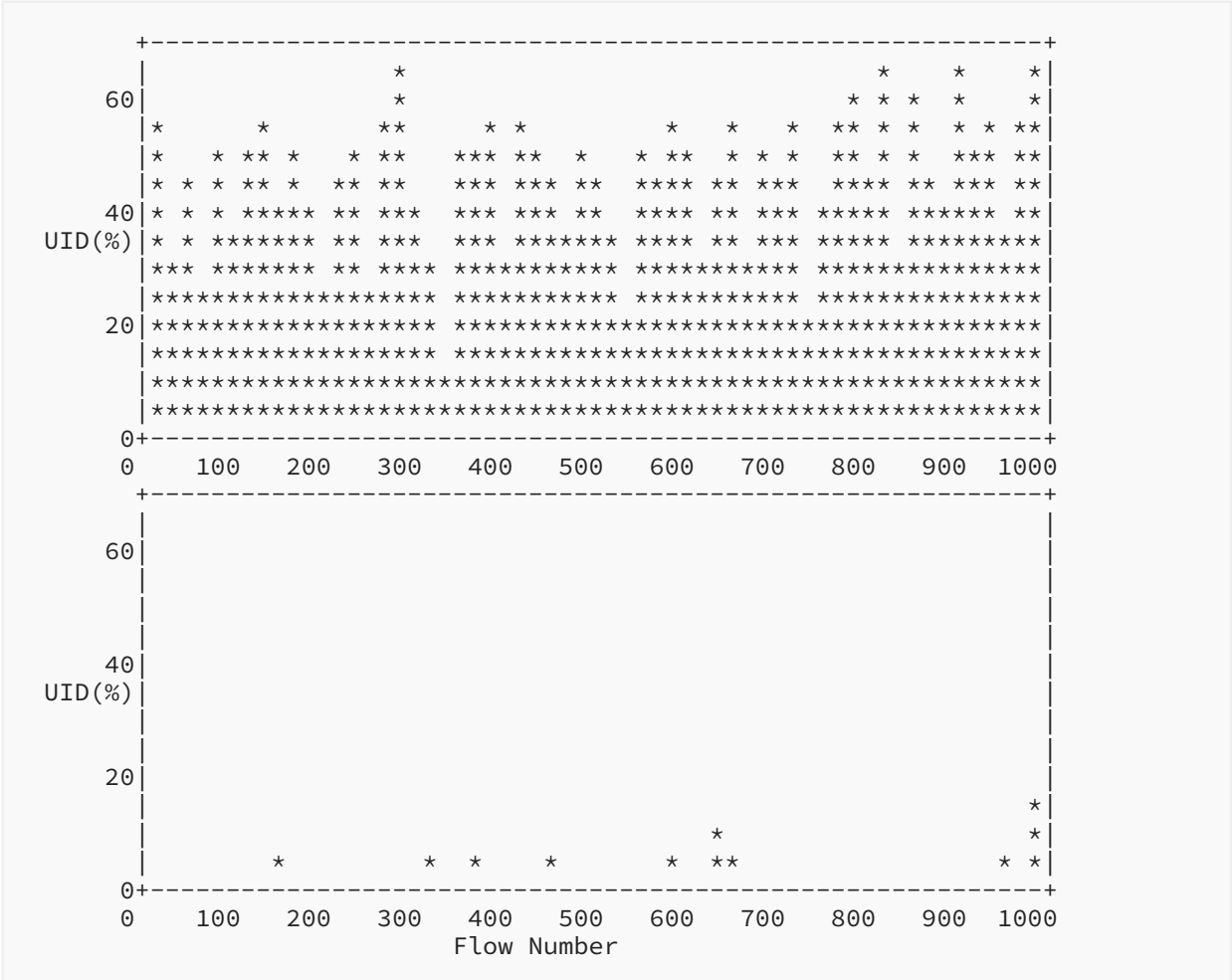

Figure 7: Simulation Results with Congestion Elimination

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 twelve links among the total 20,000 exceed the threshold, and all the CD values are less than 3%. Thus, after scheduling the traffic away from the congested paths, the degree of network congestion is greatly eliminated and the network utilization is in balance.

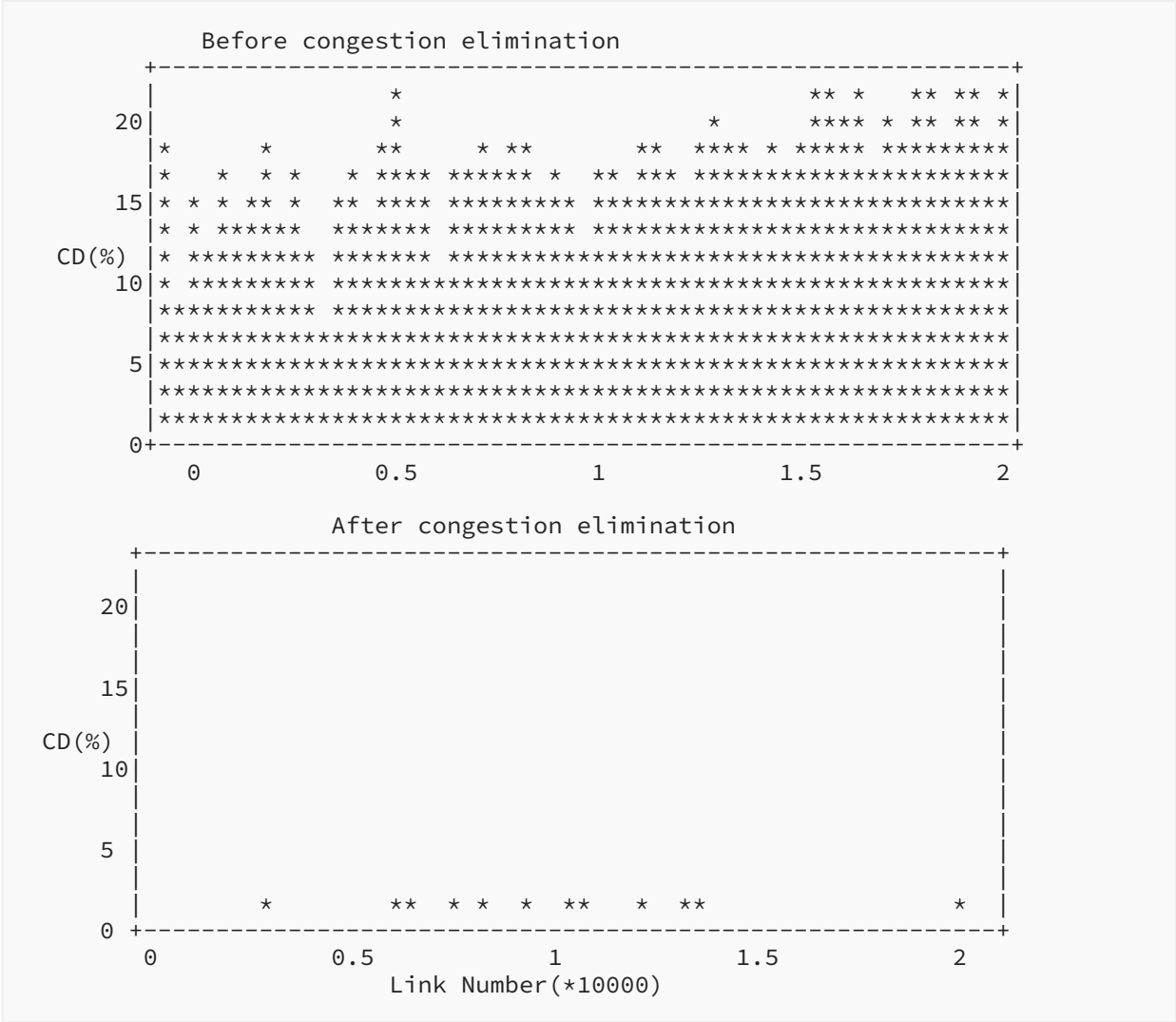

Figure 8: Simulation Results with Congestion Elimination

It is clear that by 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 underlying 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 along with no changes being 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 underlying network topology dynamically, for example, via Border Gateway Protocol - Link State (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 a change to the underlying network in a matter of minutes. More agile requirements would need to increase the sample rate of the underlying network and decrease the detection and notification interval of the underlying network. The methods of gathering this information as well as decreasing its latency are out of the scope of this document.

6. Security Considerations

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

[RFC5440] and [RFC8253] provide additional information.

The control priority and interaction process should also be carefully designed for the combination of the distributed protocol and central control. Generally, the central control instructions should have higher priority than the forwarding actions determined by the distributed protocol. When communication between PCE and the underlying devices is disrupted, the distributed protocol should take control of the underlying network. [PCE-NATIVE-IP] provides more considerations corresponding to the solution.

7. IANA Considerations

This document has no IANA actions.

8. References

8.1. 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>.

8.2. Informative References

[PCE-NATIVE-IP] Wang, A., Zhao, Q., Khasanov, B., and H. Chen, "PCE in Native IP Network", Work in Progress, Internet-Draft, draft-ietf-teas-pce-native-ip-05, 9 January 2020, <https://tools.ietf.org/html/draft-ietf-teas-pce-native-ip-05>.

[PCEP-NATIVE-IP-EXT] Wang, A., Khasanov, B., Fang, S., and C. Zhu, "PCEP Extension for Native IP Network", Work in Progress, Internet-Draft, draft-ietf-pce-pcep-extension-native-ip-05, 17 February 2020, <https://tools.ietf.org/html/draft-ietf-pce-pcep-extension-native-ip-05>.

[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", DOI 10.1109/ACCESS.2019.2902610, IEEE Access 18526773, March 2019, <https://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>.

Acknowledgements

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

Thanks to Benjamin Kaduk for his careful review and valuable suggestions on this document. Also, thanks to Roman Danyliw, Alvaro Retana, and Éric Vyncke for their reviews and comments.

Contributors

Lu Huang contributed to the content of this document.

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