Internet Engineering Task Force (IETF)
Request for Comments: 9522
Obsoletes: 3272
Category: Informational
ISSN: 2070-1721

A. Farrel, Ed.
Old Dog Consulting

January 2024

# Overview and Principles of Internet Traffic Engineering

## Abstract

This document describes the principles of traffic engineering (TE) in the Internet. The document is intended to promote better understanding of the issues surrounding traffic engineering in IP networks and the networks that support IP networking and to provide a common basis for the development of traffic-engineering capabilities for the Internet. The principles, architectures, and methodologies for performance evaluation and performance optimization of operational networks are also discussed.

This work was first published as RFC 3272 in May 2002. This document obsoletes RFC 3272 by making a complete update to bring the text in line with best current practices for Internet traffic engineering and to include references to the latest relevant work in the IETF.

## 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/rfc9522.



## Table of Contents

## 1. Introduction

This document describes the principles of Internet traffic engineering (TE). The objective of the document is to articulate the general issues and principles for Internet TE and, where appropriate, to provide recommendations, guidelines, and options for the development of preplanned (offline) and dynamic (online) Internet TE capabilities and support systems.

Even though Internet TE is most effective when applied end-to-end, the focus of this document is TE within a given domain (such as an Autonomous System (AS)). However, because a preponderance of

Internet traffic tends to originate in one AS and terminate in another, this document also provides an overview of aspects pertaining to inter-domain TE.

This document provides terminology and a taxonomy for describing and understanding common Internet TE concepts.

This work was first published as [RFC3272] in May 2002. This document obsoletes [RFC3272] by making a complete update to bring the text in line with best current practices for Internet TE and to include references to the latest relevant work in the IETF. It is worth noting around three-fifths of the RFCs referenced in this document postdate the publication of [RFC3272]. Appendix A provides a summary of changes between [RFC3272] and this document.

### 1.1. What is Internet Traffic Engineering?

One of the most significant functions performed in the Internet is the routing and forwarding of traffic from ingress nodes to egress nodes. Therefore, one of the most distinctive functions performed by Internet traffic engineering is the control and optimization of these routing and forwarding functions, to steer traffic through the network.

Internet traffic engineering is defined as that aspect of Internet network engineering dealing with the issues of performance evaluation and performance optimization of operational IP networks. Traffic engineering encompasses the application of technology and scientific principles to the measurement, characterization, modeling, and control of Internet traffic [RFC2702] [AWD2].

It is the performance of the network as seen by end users of network services that is paramount. The characteristics visible to end users are the emergent properties of the network, which are the characteristics of the network when viewed as a whole. A central goal of the service provider, therefore, is to enhance the emergent properties of the network while taking economic considerations into account. This is accomplished by addressing traffic-oriented performance requirements while utilizing network resources without excessive waste and in a reliable way. Traffic-oriented performance measures include delay, delay variation, packet loss, and throughput.

Internet TE responds to network events (such as link or node failures, reported or predicted network congestion, planned maintenance, service degradation, planned changes in the traffic matrix, etc.). Aspects of capacity management respond at intervals ranging from days to years. Routing control functions operate at intervals ranging from milliseconds to days. Packet-level processing functions operate at very fine levels of temporal resolution (up to milliseconds) while reacting to statistical measures of the real-time behavior of traffic.

Thus, the optimization aspects of TE can be viewed from a control perspective and can be both proactive and reactive. In the proactive case, the TE control system takes preventive action to protect against predicted unfavorable future network states, for example, by engineering backup paths. It may also take action that will lead to a more desirable future network state. In the reactive case, the control system responds to correct issues and adapt to network events, such as routing after failure.

Another important objective of Internet TE is to facilitate reliable network operations [RFC2702]. Reliable network operations can be facilitated by providing mechanisms that enhance network integrity and by embracing policies emphasizing network survivability. This reduces the vulnerability of services to outages arising from errors,

faults, and failures occurring within the network infrastructure.

The optimization aspects of TE can be achieved through capacity management and traffic management. In this document, capacity management includes capacity planning, routing control, and resource management. Network resources of particular interest include link bandwidth, buffer space, and computational resources. In this document, traffic management includes:

1. Nodal traffic control functions, such as traffic conditioning, queue management, and scheduling.

2. Other functions that regulate the flow of traffic through the network or that arbitrate access to network resources between different packets or between different traffic flows.

One major challenge of Internet TE is the realization of automated control capabilities that adapt quickly and cost-effectively to significant changes in network state, while still maintaining stability of the network. Performance evaluation can assess the effectiveness of TE methods, and the results of this evaluation can be used to identify existing problems, guide network reoptimization, and aid in the prediction of potential future problems. However, this process can also be time-consuming and may not be suitable to act on short-lived changes in the network.

Performance evaluation can be achieved in many different ways. The most notable techniques include analytic methods, simulation, and empirical methods based on measurements.

Traffic engineering comes in two flavors:

* A background process that constantly monitors traffic and network conditions and optimizes the use of resources to improve performance.

* A form of a pre-planned traffic distribution that is considered optimal.

In the latter case, any deviation from the optimum distribution (e.g., caused by a fiber cut) is reverted upon repair without further optimization. However, this form of TE relies upon the notion that the planned state of the network is optimal. Hence, there are two levels of TE in such a mode:

* The TE-planning task to enable optimum traffic distribution.

* The routing and forwarding tasks that keep traffic flows attached to the pre-planned distribution.

As a general rule, TE concepts and mechanisms must be sufficiently specific and well-defined to address known requirements but simultaneously flexible and extensible to accommodate unforeseen future demands (see Section 6.1).

## 1.2. Components of Traffic Engineering

As mentioned in Section 1.1, Internet traffic engineering provides performance optimization of IP networks while utilizing network resources economically and reliably. Such optimization is supported at the control/controller level and within the data/forwarding plane.

The key elements required in any TE solution are as follows:

1. Policy

## 2. Path steering

## 3. Resource management

Some TE solutions rely on these elements to a lesser or greater extent. Debate remains about whether a solution can truly be called "TE" if it does not include all of these elements. For the sake of this document, we assert that all TE solutions must include some aspects of all of these elements. Other solutions can be classed as "partial TE" and also fall in scope of this document.

Policy allows for the selection of paths (including next hops) based on information beyond basic reachability. Early definitions of routing policy, e.g., [RFC1102] and [RFC1104], discuss routing policy being applied to restrict access to network resources at an aggregate level. BGP is an example of a commonly used mechanism for applying such policies; see [RFC4271] and [RFC8955]. In the TE context, policy decisions are made within the control plane or by controllers in the management plane and govern the selection of paths. Examples can be found in [RFC4655] and [RFC5394]. TE solutions may cover the mechanisms to distribute and/or enforce policies, but definition of specific policies is left to the network operator.

Path steering is the ability to forward packets using more information than just knowledge of the next hop. Examples of path steering include IPv4 source routes [RFC0791], RSVP-TE explicit routes [RFC3209], Segment Routing (SR) [RFC8402], and Service Function Chaining [RFC7665]. Path steering for TE can be supported via control plane protocols, by encoding in the data plane headers, or by a combination of the two. This includes when control is provided by a controller using a network-facing control protocol.

Resource management provides resource-aware control and forwarding. Examples of resources are bandwidth, buffers, and queues, all of which can be managed to control loss and latency.

Resource reservation is the control aspect of resource management. It provides for domain-wide consensus about which network resources are used by a particular flow. This determination may be made at a very coarse or very fine level. Note that this consensus exists at the network control or controller level but not within the data plane. It may be composed purely of accounting/bookkeeping, but it typically includes an ability to admit, reject, or reclassify a flow based on policy. Such accounting can be done based on any combination of a static understanding of resource requirements and the use of dynamic mechanisms to collect requirements (e.g., via RSVP-TE [RFC3209]) and resource availability (e.g., via OSPF extensions for GMPLS [RFC4203]).

Resource allocation is the data plane aspect of resource management. It provides for the allocation of specific node and link resources to specific flows. Example resources include buffers, policing, and rate-shaping mechanisms that are typically supported via queuing. Resource allocation also includes the matching of a flow (i.e., flow classification) to a particular set of allocated resources. The method of flow classification and granularity of resource management is technology-specific. Examples include Diffserv with dropping and remarking [RFC4594], MPLS-TE [RFC3209], GMPLS-based Label Switched Paths (LSPs) [RFC3945], as well as controller-based solutions [RFC8453]. This level of resource control, while optional, is important in networks that wish to support network congestion management policies to control or regulate the offered traffic to deliver different levels of service and alleviate network congestion problems. It is also important in networks that wish to control the latency experienced by specific traffic flows.

## 1.3. Scope

The scope of this document is intra-domain TE because this is the practical level of TE technology that exists in the Internet at the time of writing. That is, this document describes TE within a given AS in the Internet. This document discusses concepts pertaining to intra-domain traffic control, including such issues as routing control, micro and macro resource allocation, and control coordination problems that arise consequently.

This document describes and characterizes techniques already in use or in advanced development for Internet TE. The way these techniques fit together is discussed and scenarios in which they are useful are identified.

Although the emphasis in this document is on intra-domain traffic engineering, an overview of the high-level considerations pertaining to inter-domain TE is provided in Section 7. Inter-domain Internet TE is crucial to the performance enhancement of the world-wide Internet infrastructure.

Whenever possible, relevant requirements from existing IETF documents and other sources are incorporated by reference.

## 1.4. Terminology

This section provides terminology that is useful for Internet TE. The definitions presented apply to this document. These terms may have other meanings elsewhere.

Busy hour: A one-hour period within a specified interval of time (typically 24 hours) in which the traffic load in a network or sub-network is greatest.

Congestion: A state of a network resource in which the traffic incident on the resource exceeds its output capacity over an interval of time. A small amount of congestion may be beneficial to ensure that network resources are run at full capacity, and this may be particularly true at the network edge where it is desirable to ensure that user traffic is served as much as possible. Within the network, if congestion is allowed to build (such as when input traffic exceeds output traffic in a sustained way), it will have a negative effect on user traffic.

Congestion avoidance: An approach to congestion management that attempts to obviate the occurrence of congestion. It is chiefly relevant to network congestion, although it may form a part of demand-side congestion management.

Congestion response: An approach to congestion management that attempts to remedy congestion problems that have already occurred.

Constraint-based routing: A class of routing protocols that takes specified traffic attributes, network constraints, and policy constraints into account when making routing decisions. Constraint-based routing is applicable to traffic aggregates as well as flows. It is a generalization of QoS-based routing.

Demand-side congestion management: A congestion management scheme that addresses congestion problems by regulating or conditioning the offered load.

Effective bandwidth: The minimum amount of bandwidth that can be assigned to a flow or traffic aggregate in order to deliver "acceptable service quality" to the flow or traffic aggregate. See [KELLY] for a more mathematical definition.

Egress node: The device (router) at which traffic leaves a network toward a destination (host, server, etc.) or to another network.

End-to-end: This term is context-dependent and often applies to the life of a traffic flow from original source to final destination. In contrast, edge-to-edge is often used to describe the traffic flow from the entry of a domain or network to the exit of that domain or network. However, in some contexts (for example, where there is a service interface between a network and the client of that network or where a path traverses multiple domains under the control of a single process), end-to-end is used to refer to the full operation of the service that may be composed of concatenated edge-to-edge operations. Thus, in the context of TE, the term "end-to-end" may refer to the full TE path but not to the complete path of the traffic from source application to ultimate destination.

Hotspot: A network element or subsystem that is in a considerably higher state of congestion than others.

Ingress node: The device (router) at which traffic enters a network from a source (host) or from another network.

Metric: A parameter defined in terms of standard units of measurement.

Measurement methodology: A repeatable measurement technique used to derive one or more metrics of interest.

Network congestion: Congestion within the network at a specific node or a specific link that is sufficiently extreme that it results in unacceptable queuing delay or packet loss. Network congestion can negatively impact end-to-end or edge-to-edge traffic flows, so TE schemes may be deployed to balance traffic in the network and deliver congestion avoidance.

Network survivability: The capability to provide a prescribed level of QoS for existing services after a given number of failures occur within the network.

Offered load: Offered load is also sometimes called "offered traffic load". It is a measure of the amount of traffic being presented to be carried across a network compared to the capacity of the network to carry it. This term derives from queuing theory, and an offered load of 1 indicates that the network can carry, but only just manage to carry, all of the traffic presented to it.

Offline traffic engineering: A traffic engineering system that exists outside of the network.

Online traffic engineering: A traffic-engineering system that exists within the network, typically implemented on or as adjuncts to operational network elements.

Performance measures: Metrics that provide quantitative or qualitative measures of the performance of systems or subsystems of interest.

Performance metric: A performance parameter defined in terms of standard units of measurement.

Provisioning: The process of assigning or configuring network resources to meet certain requests.

Quality of Service (QoS): QoS [RFC3198] refers to the mechanisms

used within a network to achieve specific goals for the delivery of traffic for a particular service according to the parameters specified in a Service Level Agreement. "Quality" is characterized by service availability, delay, jitter, throughput, and packet loss ratio. At a network resource level, "Quality of Service" refers to a set of capabilities that allow a service provider to prioritize traffic, control bandwidth, and network latency.

QoS routing: Class of routing systems that selects paths to be used by a flow based on the QoS requirements of the flow.

Service Level Agreement (SLA): A contract between a provider and a customer that guarantees specific levels of performance and reliability at a certain cost.

Service Level Objective (SLO): A key element of an SLA between a provider and a customer. SLOs are agreed upon as a means of measuring the performance of the service provider and are outlined as a way of avoiding disputes between the two parties based on misunderstanding.

Stability: An operational state in which a network does not oscillate in a disruptive manner from one mode to another mode.

Supply-side congestion management: A congestion management scheme that provisions additional network resources to address existing and/or anticipated congestion problems.

Traffic characteristic: A description of the temporal behavior or a description of the attributes of a given traffic flow or traffic aggregate.

Traffic-engineering system: A collection of objects, mechanisms, and protocols that are used together to accomplish traffic-engineering objectives.

Traffic flow: A stream of packets between two endpoints that can be characterized in a certain way. A common classification for a traffic flow selects packets with the five-tuple of source and destination addresses, source and destination ports, and protocol ID. Flows may be very small and transient, ranging to very large. The TE techniques described in this document are likely to be more effective when applied to large flows. Traffic flows may be aggregated and treated as a single unit in some forms of TE, making it possible to apply TE to the smaller flows that comprise the aggregate.

Traffic mapping: Traffic mapping is the assignment of traffic workload onto (pre-established) paths to meet certain requirements.

Traffic matrix: A representation of the traffic demand between a set of origin and destination abstract nodes. An abstract node can consist of one or more network elements.

Traffic monitoring: The process of observing traffic characteristics at a given point in a network and collecting the traffic information for analysis and further action.

Traffic trunk: An aggregation of traffic flows belonging to the same class that are forwarded through a common path. A traffic trunk may be characterized by an ingress and egress node and a set of attributes that determine its behavioral characteristics and requirements from the network.

Workload: Workload is also sometimes called "traffic workload". It is an evaluation of the amount of work that must be done in a network in order to facilitate the traffic demand. Colloquially, it is the answer to, "How busy is the network?"

## 2. Background

The Internet aims to convey IP packets from ingress nodes to egress nodes efficiently, expeditiously, and economically. Furthermore, in a multi-class service environment (e.g., Diffserv capable networks; see Section 5.1.1.2), the resource-sharing parameters of the network must be appropriately determined and configured according to prevailing policies and service models to resolve resource contention issues arising from mutual interference between packets traversing the network. Thus, consideration must be given to resolving competition for network resources between traffic flows belonging to the same service class (intra-class contention resolution) and traffic flows belonging to different classes (inter-class contention resolution).

### 2.1. Context of Internet Traffic Engineering

The context of Internet traffic engineering includes the following sub-contexts:

1. A network domain context that defines the scope under consideration and, in particular, the situations in which the TE problems occur. The network domain context includes network structure, policies, characteristics, constraints, quality attributes, and optimization criteria.

2. A problem context defining the general and concrete issues that TE addresses. The problem context includes identification, abstraction of relevant features, representation, formulation, specification of the requirements on the solution space, and specification of the desirable features of acceptable solutions.

3. A solution context suggesting how to address the issues identified by the problem context. The solution context includes analysis, evaluation of alternatives, prescription, and resolution.

4. An implementation and operational context in which the solutions are instantiated. The implementation and operational context includes planning, organization, and execution.

The context of Internet TE and the different problem scenarios are discussed in the following subsections.

### 2.2. Network Domain Context

IP networks range in size from small clusters of routers situated within a given location to thousands of interconnected routers, switches, and other components distributed all over the world.

At the most basic level of abstraction, an IP network can be represented as a distributed dynamic system consisting of:

* a set of interconnected resources that provide transport services for IP traffic subject to certain constraints

* a demand system representing the offered load to be transported through the network

* a response system consisting of network processes, protocols, and related mechanisms that facilitate the movement of traffic through

the network (see also [AWD2])

The network elements and resources may have specific characteristics restricting the manner in which the traffic demand is handled. Additionally, network resources may be equipped with traffic control mechanisms managing the way in which the demand is serviced. Traffic control mechanisms may be used to:

* control packet processing activities within a given resource

* arbitrate contention for access to the resource by different packets

* regulate traffic behavior through the resource

A configuration management and provisioning system may allow the settings of the traffic control mechanisms to be manipulated by external or internal entities in order to exercise control over the way in which the network elements respond to internal and external stimuli.

The details of how the network carries packets are specified in the policies of the network administrators and are installed through network configuration management and policy-based provisioning systems. Generally, the types of service provided by the network also depend upon the technology and characteristics of the network elements and protocols, the prevailing service and utility models, and the ability of the network administrators to translate policies into network configurations.

Internet networks have two significant characteristics:

* They provide real-time services.

* Their operating environments are very dynamic.

The dynamic characteristics of IP and IP/MPLS networks can be attributed in part to fluctuations in demand, the interaction between various network protocols and processes, the rapid evolution of the infrastructure that demands the constant inclusion of new technologies and new network elements, and the transient and persistent faults that occur within the system.

Packets contend for the use of network resources as they are conveyed through the network. A network resource is considered to be congested if, for an interval of time, the arrival rate of packets exceeds the output capacity of the resource. Network congestion may result in some of the arriving packets being delayed or even dropped.

Network congestion increases transit delay and delay variation, may lead to packet loss, and reduces the predictability of network services. Clearly, while congestion may be a useful tool at ingress edge nodes, network congestion is highly undesirable. Combating network congestion at a reasonable cost is a major objective of Internet TE, although it may need to be traded with other objectives to keep the costs reasonable.

Efficient sharing of network resources by multiple traffic flows is a basic operational premise for the Internet. A fundamental challenge in network operation is to increase resource utilization while minimizing the possibility of congestion.

The Internet has to function in the presence of different classes of traffic with different service requirements. This requirement is clarified in the architecture for Differentiated Services (Diffserv) [RFC2475]. That document describes how packets can be grouped into

behavior aggregates such that each aggregate has a common set of behavioral characteristics or a common set of delivery requirements. Delivery requirements of a specific set of packets may be specified explicitly or implicitly. Two of the most important traffic delivery requirements are:

* Capacity constraints can be expressed statistically as peak rates, mean rates, burst sizes, or as some deterministic notion of effective bandwidth.

* QoS requirements can be expressed in terms of:

  - integrity constraints, such as packet loss

  - temporal constraints, such as timing restrictions for the delivery of each packet (delay) and timing restrictions for the delivery of consecutive packets belonging to the same traffic stream (delay variation)

## 2.3. Problem Context

There are several problems associated with operating a network like those described in the previous section. This section analyzes the problem context in relation to TE. The identification, abstraction, representation, and measurement of network features relevant to TE are significant issues.

A particular challenge is to formulate the problems that traffic engineering attempts to solve. For example:

* How to identify the requirements on the solution space

* How to specify the desirable features of solutions

* How to actually solve the problems

* How to measure and characterize the effectiveness of solutions

Another class of problems is how to measure and estimate relevant network state parameters. Effective TE relies on a good estimate of the offered traffic load as well as a view of the underlying topology and associated resource constraints. Offline planning requires a full view of the topology of the network or partial network that is being planned.

Still another class of problem is how to characterize the state of the network and how to evaluate its performance. The performance evaluation problem is two-fold: one aspect relates to the evaluation of the system-level performance of the network, and the other aspect relates to the evaluation of resource-level performance, which restricts attention to the performance analysis of individual network resources.

In this document, we refer to the system-level characteristics of the network as the "macro-states" and the resource-level characteristics as the "micro-states." The system-level characteristics are also known as the emergent properties of the network. Correspondingly, we refer to the TE schemes dealing with network performance optimization at the systems level as "macro-TE" and the schemes that optimize at the individual resource level as "micro-TE." Under certain circumstances, the system-level performance can be derived from the resource-level performance using appropriate rules of composition, depending upon the particular performance measures of interest.

Another fundamental class of problem concerns how to effectively optimize network performance. Performance optimization may entail

translating solutions for specific TE problems into network configurations. Optimization may also entail some degree of resource management control, routing control, and capacity augmentation.

### 2.3.1. Congestion and Its Ramifications

Network congestion is one of the most significant problems in an operational IP context. A network element is said to be congested if it experiences sustained overload over an interval of time. Although congestion at the edge of the network may be beneficial in ensuring that the network delivers as much traffic as possible, network congestion almost always results in degradation of service quality to end users. Congestion avoidance and response schemes can include demand-side policies and supply-side policies. Demand-side policies may restrict access to congested resources or dynamically regulate the demand to alleviate the overload situation. Supply-side policies may expand or augment network capacity to better accommodate offered traffic. Supply-side policies may also reallocate network resources by redistributing traffic over the infrastructure. Traffic redistribution and resource reallocation serve to increase the effective capacity of the network.

The emphasis of this document is primarily on congestion management schemes falling within the scope of the network, rather than on congestion management systems dependent upon sensitivity and adaptivity from end systems. That is, the aspects that are considered in this document with respect to congestion management are those solutions that can be provided by control entities operating on the network and by the actions of network administrators and network operations systems.

## 2.4. Solution Context

The solution context for Internet TE involves analysis, evaluation of alternatives, and choice between alternative courses of action. Generally, the solution context is based on making inferences about the current or future state of the network and making decisions that may involve a preference between alternative sets of action. More specifically, the solution context demands reasonable estimates of traffic workload, characterization of network state, derivation of solutions that may be implicitly or explicitly formulated, and possibly instantiation of a set of control actions. Control actions may involve the manipulation of parameters associated with routing, control over tactical capacity acquisition, and control over the traffic management functions.

The following list of instruments may be applicable to the solution context of Internet TE:

* A set of policies, objectives, and requirements (which may be context dependent) for network performance evaluation and performance optimization.

* A collection of online and, in some cases, possibly offline tools and mechanisms for measurement, characterization, modeling, control of traffic, control over the placement and allocation of network resources, as well as control over the mapping or distribution of traffic onto the infrastructure.

* A set of constraints on the operating environment, the network protocols, and the TE system itself.

* A set of quantitative and qualitative techniques and methodologies for abstracting, formulating, and solving TE problems.

* A set of administrative control parameters that may be manipulated

through a configuration management system. Such a system may itself include a configuration control subsystem, a configuration repository, a configuration accounting subsystem, and a configuration auditing subsystem.

* A set of guidelines for network performance evaluation, performance optimization, and performance improvement.

Determining traffic characteristics through measurement or estimation is very useful within the realm of the TE solution space. Traffic estimates can be derived from customer subscription information, traffic projections, traffic models, and actual measurements. The measurements may be performed at different levels, e.g., at the traffic-aggregate level or at the flow level. Measurements at the flow level or on small traffic aggregates may be performed at edge nodes, when traffic enters and leaves the network. Measurements for large traffic aggregates may be performed within the core of the network.

To conduct performance studies and to support planning of existing and future networks, a routing analysis may be performed to determine the paths the routing protocols will choose for various traffic demands and to ascertain the utilization of network resources as traffic is routed through the network. Routing analysis captures the selection of paths through the network, the assignment of traffic across multiple feasible routes, and the multiplexing of IP traffic over traffic trunks (if such constructs exist) and over the underlying network infrastructure. A model of network topology is necessary to perform routing analysis. A network topology model may be extracted from:

* network architecture documents

* network designs

* information contained in router configuration files

* routing databases such as the link-state database of an Interior Gateway Protocol (IGP)

* routing tables

* automated tools that discover and collate network topology information

Topology information may also be derived from servers that monitor network state and from servers that perform provisioning functions.

Routing in operational IP networks can be administratively controlled at various levels of abstraction, including the manipulation of BGP attributes and IGP metrics. For path-oriented technologies such as MPLS, routing can be further controlled by the manipulation of relevant TE parameters, resource parameters, and administrative policy constraints. Within the context of MPLS, the path of an explicitly routed LSP can be computed and established in various ways, including:

* manually

* automatically and online using constraint-based routing processes implemented on Label Switching Routers (LSRs)

* automatically and offline using constraint-based routing entities implemented on external TE support systems

### 2.4.1. Combating the Congestion Problem

Minimizing congestion is a significant aspect of Internet traffic engineering. This subsection gives an overview of the general approaches that have been used or proposed to combat congestion.

Congestion management policies can be categorized based upon the following criteria (see [YARE95] for a more detailed taxonomy of congestion control schemes):

1. Congestion Management Based on Response Timescales

   * Long (weeks to months): Expanding network capacity by adding new equipment, routers, and links takes time and is comparatively costly. Capacity planning needs to take this into consideration. Network capacity is expanded based on estimates or forecasts of future traffic development and traffic distribution. These upgrades are typically carried out over weeks, months, or maybe even years.

   * Medium (minutes to days): Several control policies fall within the medium timescale category. Examples include:

     a. Adjusting routing protocol parameters to route traffic away from or towards certain segments of the network.

     b. Setting up or adjusting explicitly routed LSPs in MPLS networks to route traffic trunks away from possibly congested resources or toward possibly more favorable routes.

     c. Reconfiguring the logical topology of the network to make it correlate more closely with the spatial traffic distribution using, for example, an underlying path-oriented technology such as MPLS LSPs or optical channel trails.

     When these schemes are adaptive, they rely on measurement systems. A measurement system monitors changes in traffic distribution, traffic loads, and network resource utilization and then provides feedback to the online or offline TE mechanisms and tools so that they can trigger control actions within the network. The TE mechanisms and tools can be implemented in a distributed or centralized fashion. A centralized scheme may have full visibility into the network state and may produce more optimal solutions. However, centralized schemes are prone to single points of failure and may not scale as well as distributed schemes. Moreover, the information utilized by a centralized scheme may be stale and might not reflect the actual state of the network. It is not an objective of this document to make a recommendation between distributed and centralized schemes; that is a choice that network administrators must make based on their specific needs.

   * Short (minutes or less): This category includes packet-level processing functions and events that are recorded on the order of several round-trip times. It also includes router mechanisms such as passive and active buffer management. All of these mechanisms are used to control congestion or signal congestion to end systems so that they can adaptively regulate the rate at which traffic is injected into the network. A well-known active queue management scheme, especially for responsive traffic such as TCP, is Random Early Detection (RED) [FLJA93]. During congestion (but before the queue is filled), the RED scheme chooses arriving packets to "mark" according to a probabilistic algorithm that takes into account

the average queue size. A router that does not utilize Explicit Congestion Notification (ECN) [RFC3168] can simply drop marked packets to alleviate congestion and implicitly notify the receiver about the congestion. On the other hand, if the router and the end hosts support ECN, they can set the ECN field in the packet header, and the end host can act on this information. Several variations of RED have been proposed to support different drop precedence levels in multi-class environments [RFC2597]. RED provides congestion avoidance that is better than or equivalent to Tail-Drop (TD) queue management (drop arriving packets only when the queue is full). Importantly, RED reduces the possibility of retransmission bursts becoming synchronized within the network and improves fairness among different responsive traffic sessions. However, RED by itself cannot prevent congestion and unfairness caused by sources unresponsive to RED, e.g., some misbehaved greedy connections. Other schemes have been proposed to improve performance and fairness in the presence of unresponsive traffic. Some of those schemes (such as Longest Queue Drop (LQD) and Dynamic Soft Partitioning with Random Drop (RND) [SLDC98]) were proposed as theoretical frameworks and are typically not available in existing commercial products, while others (such as Approximate Fair Dropping (AFD) [AFD03]) have seen some implementation. Advice on the use of Active Queue Management (AQM) schemes is provided in [RFC7567]. [RFC7567] recommends self-tuning AQM algorithms like those that the IETF has published in [RFC8290], [RFC8033], [RFC8034], and [RFC9332], but RED is still appropriate for links with stable bandwidth, if configured carefully.

2. Reactive versus Preventive Congestion Management Schemes

   * Reactive (recovery) congestion management policies react to existing congestion problems. All the policies described above for the short and medium timescales can be categorized as being reactive. They are based on monitoring and identifying congestion problems that exist in the network and on the initiation of relevant actions to ease a situation. Reactive congestion management schemes may also be preventive.

   * Preventive (predictive/avoidance) policies take proactive action to prevent congestion based on estimates and predictions of future congestion problems (e.g., traffic matrix forecasts). Some of the policies described for the long and medium timescales fall into this category. Preventive policies do not necessarily respond immediately to existing congestion problems. Instead, forecasts of traffic demand and workload distribution are considered, and action may be taken to prevent potential future congestion problems. The schemes described for the short timescale can also be used for congestion avoidance because dropping or marking packets before queues actually overflow would trigger corresponding responsive traffic sources to slow down. Preventive congestion management schemes may also be reactive.

3. Supply-Side versus Demand-Side Congestion Management Schemes

   * Supply-side congestion management policies increase the effective capacity available to traffic in order to control or reduce congestion. This can be accomplished by increasing capacity or by balancing distribution of traffic over the network. Capacity planning aims to provide a physical topology and associated link bandwidths that match or exceed estimated traffic workload and traffic distribution, subject to traffic forecasts and budgetary (or other) constraints. If

the actual traffic distribution does not fit the topology derived from capacity planning, then the traffic can be mapped onto the topology by using routing control mechanisms, by applying path-oriented technologies (e.g., MPLS LSPs and optical channel trails) to modify the logical topology or by employing some other load redistribution mechanisms.

* Demand-side congestion management policies control or regulate the offered traffic to alleviate congestion problems. For example, some of the short timescale mechanisms described earlier as well as policing and rate-shaping mechanisms attempt to regulate the offered load in various ways.

### 2.5. Implementation and Operational Context

The operational context of Internet TE is characterized by constant changes that occur at multiple levels of abstraction. The implementation context demands effective planning, organization, and execution. The planning aspects may involve determining prior sets of actions to achieve desired objectives. Organizing involves arranging and assigning responsibility to the various components of the TE system and coordinating the activities to accomplish the desired TE objectives. Execution involves measuring and applying corrective or perfective actions to attain and maintain desired TE goals.

## 3. Traffic-Engineering Process Models

This section describes a generic process model that captures the high-level practical aspects of Internet traffic engineering in an operational context. The process model is described as a sequence of actions that must be carried out to optimize the performance of an operational network (see also [RFC2702] and [AWD2]). This process model may be enacted explicitly or implicitly, by a software process or by a human.

The TE process model is iterative [AWD2]. The four phases of the process model described below are repeated as a continual sequence:

1. Define the relevant control policies that govern the operation of the network.

2. Acquire measurement data from the operational network.

3. Analyze the network state and characterize the traffic workload. Proactive analysis identifies potential problems that could manifest in the future. Reactive analysis identifies existing problems and determines their causes.

4. Optimize the performance of the network. This involves a decision process that selects and implements a set of actions from a set of alternatives given the results of the three previous steps. Optimization actions may include the use of techniques to control the offered traffic and to control the distribution of traffic across the network.

### 3.1. Components of the Traffic-Engineering Process Model

The key components of the traffic-engineering process model are as follows:

1. Measurement is crucial to the TE function. The operational state of a network can only be conclusively determined through measurement. Measurement is also critical to the optimization function because it provides feedback data that is used by TE control subsystems. This data is used to adaptively optimize

network performance in response to events and stimuli originating within and outside the network. Measurement in support of the TE function can occur at different levels of abstraction. For example, measurement can be used to derive packet-level characteristics, flow-level characteristics, user- or customer-level characteristics, traffic-aggregate characteristics, component-level characteristics, and network-wide characteristics.

2. Modeling, analysis, and simulation are important aspects of Internet TE. Modeling involves constructing an abstract or physical representation that depicts relevant traffic characteristics and network attributes. A network model is an abstract representation of the network that captures relevant network features, attributes, and characteristics. Network simulation tools are extremely useful for TE. Because of the complexity of realistic quantitative analysis of network behavior, certain aspects of network performance studies can only be conducted effectively using simulation.

3. Network performance optimization involves resolving network issues by transforming such issues into concepts that enable a solution, identification of a solution, and implementation of the solution. Network performance optimization can be corrective or perfective. In corrective optimization, the goal is to remedy a problem that has occurred or that is incipient. In perfective optimization, the goal is to improve network performance even when explicit problems do not exist and are not anticipated.

## 4. Taxonomy of Traffic-Engineering Systems

This section presents a short taxonomy of traffic-engineering systems constructed based on TE styles and views as listed below and described in greater detail in the following subsections of this document:

* Time-Dependent versus State-Dependent versus Event-Dependent

* Offline versus Online

* Centralized versus Distributed

* Local versus Global Information

* Prescriptive versus Descriptive

* Open-Loop versus Closed-Loop

* Tactical versus Strategic

### 4.1. Time-Dependent versus State-Dependent versus Event-Dependent

Traffic-engineering methodologies can be classified as time-dependent, state-dependent, or event-dependent. All TE schemes are considered to be dynamic in this document. Static TE implies that no TE methodology or algorithm is being applied -- it is a feature of network planning but lacks the reactive and flexible nature of TE.

In time-dependent TE, historical information based on periodic variations in traffic (such as time of day) is used to pre-program routing and other TE control mechanisms. Additionally, customer subscription or traffic projection may be used. Pre-programmed routing plans typically change on a relatively long timescale (e.g., daily). Time-dependent algorithms do not attempt to adapt to short-term variations in traffic or changing network conditions. An example of a time-dependent algorithm is a centralized optimizer

where the input to the system is a traffic matrix and multi-class QoS requirements as described in [MR99]. Another example of such a methodology is the application of data mining to Internet traffic [AJ19], which enables the use of various machine learning algorithms to identify patterns within historically collected datasets about Internet traffic and to extract information in order to guide decision-making and improve efficiency and productivity of operational processes.

State-dependent TE adapts the routing plans based on the current state of the network, which provides additional information on variations in actual traffic (i.e., perturbations from regular variations) that could not be predicted using historical information. Constraint-based routing is an example of state-dependent TE operating in a relatively long timescale. An example of operating in a relatively short timescale is a load-balancing algorithm described in [MATE]. The state of the network can be based on parameters flooded by the routers. Another approach is for a particular router performing adaptive TE to send probe packets along a path to gather the state of that path. [RFC6374] defines protocol extensions to collect performance measurements from MPLS networks. Another approach is for a management system to gather the relevant information directly from network elements using telemetry data collection publication/subscription techniques [RFC7923]. Timely gathering and distribution of state information is critical for adaptive TE. While time-dependent algorithms are suitable for predictable traffic variations, state-dependent algorithms may be needed to increase network efficiency and to provide resilience to adapt to changes in network state.

Event-dependent TE methods can also be used for TE path selection. Event-dependent TE methods are distinct from time-dependent and state-dependent TE methods in the manner in which paths are selected. These algorithms are adaptive and distributed in nature, and they typically use learning models to find good paths for TE in a network. While state-dependent TE models typically use available-link-bandwidth (ALB) flooding [E.360.1] for TE path selection, event-dependent TE methods do not require ALB flooding. Rather, event-dependent TE methods typically search out capacity by learning models, as in the success-to-the-top (STT) method [RFC6601]. ALB flooding can be resource intensive, since it requires link bandwidth to carry routing protocol link-state advertisements and processor capacity to process those advertisements; in addition, the overhead of the ALB advertisements and their processing can limit the size of the area and AS. Modeling results suggest that event-dependent TE methods could lead to a reduction in ALB flooding overhead without loss of network throughput performance [TE-QoS-ROUTING].

A fully functional TE system is likely to use all aspects of time-dependent, state-dependent, and event-dependent methodologies as described in Section 4.3.1.

## 4.2. Offline versus Online

Traffic engineering requires the computation of routing plans. The computation may be performed offline or online. The computation can be done offline for scenarios where routing plans need not be executed in real time. For example, routing plans computed from forecast information may be computed offline. Typically, offline computation is also used to perform extensive searches on multi-dimensional solution spaces.

Online computation is required when the routing plans must adapt to changing network conditions as in state-dependent algorithms. Unlike offline computation (which can be computationally demanding), online computation is geared toward relatively simple and fast calculations

to select routes, fine-tune the allocations of resources, and perform load balancing.

### 4.3. Centralized versus Distributed

Under centralized control, there is a central authority that determines routing plans and perhaps other TE control parameters on behalf of each router. The central authority periodically collects network-state information from all routers and sends routing information to the routers. The update cycle for information exchange in both directions is a critical parameter directly impacting the performance of the network being controlled. Centralized control may need high processing power and high bandwidth control channels.

Distributed control determines route selection by each router autonomously based on the router's view of the state of the network. The network state information may be obtained by the router using a probing method or distributed by other routers on a periodic basis using link-state advertisements. Network state information may also be disseminated under exception conditions. Examples of protocol extensions used to advertise network link-state information are defined in [RFC5305], [RFC6119], [RFC7471], [RFC8570], and [RFC8571]. See also Section 5.1.3.9.

#### 4.3.1. Hybrid Systems

In practice, most TE systems will be a hybrid of central and distributed control. For example, a popular MPLS approach to TE is to use a central controller based on an active, stateful Path Computation Element (PCE) but to use routing and signaling protocols to make local decisions at routers within the network. Local decisions may be able to respond more quickly to network events but may result in conflicts with decisions made by other routers.

Network operations for TE systems may also use a hybrid of offline and online computation. TE paths may be precomputed based on stable-state network information and planned traffic demands but may then be modified in the active network depending on variations in network state and traffic load. Furthermore, responses to network events may be precomputed offline to allow rapid reactions without further computation or may be derived online depending on the nature of the events.

#### 4.3.2. Considerations for Software-Defined Networking

As discussed in Section 5.1.2.2, one of the main drivers for Software-Defined Networking (SDN) is a decoupling of the network control plane from the data plane [RFC7149]. However, SDN may also combine centralized control of resources and facilitate application-to-network interaction via an Application Programming Interface (API), such as the one described in [RFC8040]. Combining these features provides a flexible network architecture that can adapt to the network requirements of a variety of higher-layer applications, a concept often referred to as the "programmable network" [RFC7426].

The centralized control aspect of SDN helps improve network resource utilization compared with distributed network control, where local policy may often override network-wide optimization goals. In an SDN environment, the data plane forwards traffic to its desired destination. However, before traffic reaches the data plane, the logically centralized SDN control plane often determines the path the application traffic will take in the network. Therefore, the SDN control plane needs to be aware of the underlying network topology, capabilities, and current node and link resource state.

Using a PCE-based SDN control framework [RFC7491], the available network topology may be discovered by running a passive instance of OSPF or IS-IS, or via BGP Link State (BGP-LS) [RFC9552]), to generate a Traffic Engineering Database (TED) (see Section 5.1.3.14). The PCE is used to compute a path (see Section 5.1.3.11) based on the TED and available bandwidth, and further path optimization may be based on requested objective functions [RFC5541]. When a suitable path has been computed, the programming of the explicit network path may be either performed using a signaling protocol that traverses the length of the path [RFC3209] or performed per-hop with each node being directly programmed [RFC8283] by the SDN controller.

By utilizing a centralized approach to network control, additional network benefits are also available, including Global Concurrent Optimization (GCO) [RFC5557]. A GCO path computation request will simultaneously use the network topology and a set of new path signaling requests, along with their respective constraints, for optimal placement in the network. Correspondingly, a GCO-based computation may be applied to recompute existing network paths to groom traffic and to mitigate congestion.

### 4.4. Local versus Global

Traffic-engineering algorithms may require local and global network-state information.

Local information is the state of a portion of the domain. Examples include the bandwidth and packet loss rate of a particular path or the state and capabilities of a network link. Local state information may be sufficient for certain instances of distributed control TE.

Global information is the state of the entire TE domain. Examples include a global traffic matrix and loading information on each link throughout the domain of interest. Global state information is typically required with centralized control. Distributed TE systems may also need global information in some cases.

### 4.5. Prescriptive versus Descriptive

TE systems may also be classified as prescriptive or descriptive.

Prescriptive traffic engineering evaluates alternatives and recommends a course of action. Prescriptive TE can be further categorized as either corrective or perfective. Corrective TE prescribes a course of action to address an existing or predicted anomaly. Perfective TE prescribes a course of action to evolve and improve network performance even when no anomalies are evident.

Descriptive traffic engineering, on the other hand, characterizes the state of the network and assesses the impact of various policies without recommending any particular course of action.

#### 4.5.1. Intent-Based Networking

One way to express a service request is through "intent". Intent-Based Networking aims to produce networks that are simpler to manage and operate, requiring only minimal intervention. Intent is defined in [RFC9315] as follows:

> A set of operational goals (that a network should meet) and outcomes (that a network is supposed to deliver) defined in a declarative manner without specifying how to achieve or implement them.

Intent provides data and functional abstraction so that users and

operators do not need to be concerned with low-level device configuration or the mechanisms used to achieve a given intent. This approach can be conceptually easier for a user but may be less expressive in terms of constraints and guidelines.

Intent-Based Networking is applicable to TE because many of the high-level objectives may be expressed as intent (for example, load balancing, delivery of services, and robustness against failures). The intent is converted by the management system into TE actions within the network.

### 4.6. Open-Loop versus Closed-Loop

Open-loop traffic-engineering control is where control action does not use feedback information from the current network state. However, the control action may use its own local information for accounting purposes.

Closed-loop traffic-engineering control is where control action utilizes feedback information from the network state. The feedback information may be in the form of current measurement or recent historical records.

### 4.7. Tactical versus Strategic

Tactical traffic engineering aims to address specific performance problems (such as hotspots) that occur in the network from a tactical perspective, without consideration of overall strategic imperatives. Without proper planning and insights, tactical TE tends to be ad hoc in nature.

Strategic traffic-engineering approaches the TE problem from a more organized and systematic perspective, taking into consideration the immediate and longer-term consequences of specific policies and actions.

## 5. Review of TE Techniques

This section briefly reviews different TE-related approaches proposed and implemented in telecommunications and computer networks using IETF protocols and architectures. These approaches are organized into three categories:

* TE mechanisms that adhere to the definition provided in Section 1.2

* Approaches that rely upon those TE mechanisms

* Techniques that are used by those TE mechanisms and approaches

The discussion is not intended to be comprehensive. It is primarily intended to illuminate existing approaches to TE in the Internet. A historic overview of TE in telecommunications networks was provided in Section 4 of [RFC3272], and Section 4.6 of that document presented an outline of some early approaches to TE conducted in other standards bodies. It is out of the scope of this document to provide an analysis of the history of TE or an inventory of TE-related efforts conducted by other Standards Development Organizations (SDOs).

### 5.1. Overview of IETF Projects Related to Traffic Engineering

This subsection reviews a number of IETF activities pertinent to Internet traffic engineering. Some of these technologies are widely deployed, others are mature but have seen less deployment, and some are unproven or are still under development.

### 5.1.1. IETF TE Mechanisms

#### 5.1.1.1. Integrated Services

The IETF developed the Integrated Services (Intserv) model that requires resources, such as bandwidth and buffers, to be reserved a priori for a given traffic flow to ensure that the QoS requested by the traffic flow is satisfied. The Intserv model includes additional components beyond those used in the best-effort model such as packet classifiers, packet schedulers, and admission control. A packet classifier is used to identify flows that are to receive a certain level of service. A packet scheduler handles the scheduling of service to different packet flows to ensure that QoS commitments are met. Admission control is used to determine whether a router has the necessary resources to accept a new flow.

The main issue with the Intserv model has been scalability [RFC2998], especially in large public IP networks that may potentially have millions of active traffic flows in transit concurrently. Pre-Congestion Notification (PCN) [RFC5559] solves the scaling problems of Intserv by using measurement-based admission control (and flow termination to handle failures) between edge nodes. Nodes between the edges of the internetwork have no per-flow operations, and the edge nodes can use the Resource Reservation Protocol (RSVP) per-flow or per-aggregate.

A notable feature of the Intserv model is that it requires explicit signaling of QoS requirements from end systems to routers [RFC2753]. RSVP performs this signaling function and is a critical component of the Intserv model. RSVP is described in Section 5.1.3.2.

#### 5.1.1.2. Differentiated Services

The goal of Differentiated Services (Diffserv) within the IETF was to devise scalable mechanisms for categorization of traffic into behavior aggregates, which ultimately allows each behavior aggregate to be treated differently, especially when there is a shortage of resources, such as link bandwidth and buffer space [RFC2475]. One of the primary motivations for Diffserv was to devise alternative mechanisms for service differentiation in the Internet that mitigate the scalability issues encountered with the Intserv model.

Diffserv uses the Differentiated Services field in the IP header (the DS field) consisting of six bits in what was formerly known as the Type of Service (TOS) octet. The DS field is used to indicate the forwarding treatment that a packet should receive at a transit node [RFC2474]. Diffserv includes the concept of Per-Hop Behavior (PHB) groups. Using the PHBs, several classes of services can be defined using different classification, policing, shaping, and scheduling rules.

For an end user of network services to utilize Diffserv provided by its Internet Service Provider (ISP), it may be necessary for the user to have an SLA with the ISP. An SLA may explicitly or implicitly specify a Traffic Conditioning Agreement (TCA) that defines classifier rules as well as metering, marking, discarding, and shaping rules.

Packets are classified and possibly policed and shaped at the ingress to a Diffserv network. When a packet traverses the boundary between different Diffserv domains, the DS field of the packet may be re-marked according to existing agreements between the domains.

Diffserv allows only a finite number of service classes to be specified by the DS field. The main advantage of the Diffserv

approach relative to the Intserv model is scalability. Resources are allocated on a per-class basis, and the amount of state information is proportional to the number of classes rather than to the number of application flows.

Once the network has been planned and the packets have been marked at the network edge, the Diffserv model deals with traffic management issues on a per-hop basis. The Diffserv control model consists of a collection of micro-TE control mechanisms. Other TE capabilities, such as capacity management (including routing control), are also required in order to deliver acceptable service quality in Diffserv networks. The concept of "Per-Domain Behaviors" has been introduced to better capture the notion of Diffserv across a complete domain [RFC3086].

Diffserv procedures can also be applied in an MPLS context. See Section 6.8 for more information.

#### 5.1.1.3. SR Policy

SR Policy [RFC9256] is an evolution of SR (see Section 5.1.3.12) to enhance the TE capabilities of SR. It is a framework that enables instantiation of an ordered list of segments on a node for implementing a source routing policy with a specific intent for traffic steering from that node.

An SR Policy is identified through the tuple <headend, color, endpoint>. The headend is the IP address of the node where the policy is instantiated. The endpoint is the IP address of the destination of the policy. The color is an index that associates the SR Policy with an intent (e.g., low latency).

The headend node is notified of SR Policies and associated SR paths via configuration or by extensions to protocols such as the Path Computation Element Communication Protocol (PCEP) [RFC8664] or BGP [SR-TE-POLICY]. Each SR path consists of a segment list (an SR source-routed path), and the headend uses the endpoint and color parameters to classify packets to match the SR Policy and so determine along which path to forward them. If an SR Policy is associated with a set of SR paths, each is associated with a weight for weighted load balancing. Furthermore, multiple SR Policies may be associated with a set of SR paths to allow multiple traffic flows to be placed on the same paths.

An SR Binding SID (BSID) may also be associated with each candidate path associated with an SR Policy or with the SR Policy itself. The headend node installs a BSID-keyed entry in the forwarding plane and assigns it the action of steering packets that match the entry to the selected path of the SR Policy. This steering can be done in various ways:

SID Steering: Incoming packets have an active Segment Identifier (SID) matching a local BSID at the headend.

Per-destination Steering: Incoming packets match a BGP/Service route, which indicates an SR Policy.

Per-flow Steering: Incoming packets match a forwarding array (for example, the classic 5-tuple), which indicates an SR Policy.

Policy-based Steering: Incoming packets match a routing policy, which directs them to an SR Policy.

#### 5.1.1.4. Layer 4 Transport-Based TE

In addition to IP-based TE mechanisms, Layer 4 transport-based TE

approaches can be considered in specific deployment contexts (e.g., data centers and multi-homing). For example, the 3GPP defines the Access Traffic Steering, Switching, and Splitting (ATSSS) [ATSSS] service functions as follows:

Access Traffic Steering: This is the selection of an access network for a new flow and the transfer of the traffic of that flow over the selected access network.

Access Traffic Switching: This is the migration of all packets of an ongoing flow from one access network to another access network. Only one access network is in use at a time.

Access Traffic Splitting: This is about forwarding the packets of a flow across multiple access networks simultaneously.

The control plane is used to provide hosts and specific network devices with a set of policies that specify which flows are eligible to use the ATSSS service. The traffic that matches an ATSSS policy can be distributed among the available access networks following one of the following four modes:

Active-Standby: The traffic is forwarded via a specific access (called "active access") and switched to another access (called "standby access") when the active access is unavailable.

Priority-based: Network accesses are assigned priority levels that indicate which network access is to be used first. The traffic associated with the matching flow will be steered onto the network access with the highest priority until congestion is detected. Then, the overflow will be forwarded over the next highest priority access.

Load-Balancing: The traffic is distributed among the available access networks following a distribution ratio (e.g., 75% to 25%).

Smallest Delay: The traffic is forwarded via the access that presents the smallest round-trip time (RTT).

For resource management purposes, hosts and network devices support means such as congestion control, RTT measurement, and packet scheduling.

For TCP traffic, Multipath TCP [RFC8684] and the 0-RTT Convert Protocol [RFC8803] are used to provide the ATSSS service.

Multipath QUIC [QUIC-MULTIPATH] and Proxying UDP in HTTP [RFC9298] are used to provide the ATSSS service for UDP traffic. Note that QUIC [RFC9000] supports the switching and steering functions. Indeed, QUIC supports a connection migration procedure that allows peers to change their Layer 4 transport coordinates (IP addresses, port numbers) without breaking the underlying QUIC connection.

Extensions to the Datagram Congestion Control Protocol (DCCP) [RFC4340] to support multipath operations are defined in [MULTIPATH-DCCP].

#### 5.1.1.5. Deterministic Networking

Deterministic Networking (DetNet) [RFC8655] is an architecture for applications with critical timing and reliability requirements. The layered architecture particularly focuses on developing DetNet service capabilities in the data plane [RFC8938]. The DetNet service sub-layer provides a set of Packet Replication, Elimination, and Ordering Functions (PREOF) to provide end-to-end service assurance. The DetNet forwarding sub-layer provides corresponding forwarding

assurance (low packet loss, bounded latency, and in-order delivery) functions using resource allocations and explicit route mechanisms.

The separation into two sub-layers allows a greater flexibility to adapt DetNet capability over a number of TE data plane mechanisms, such as IP, MPLS, and SR. More importantly, it interconnects IEEE 802.1 Time Sensitive Networking (TSN) [RFC9023] deployed in Industry Control and Automation Systems (ICAS).

DetNet can be seen as a specialized branch of TE, since it sets up explicit optimized paths with allocation of resources as requested. A DetNet application can express its QoS attributes or traffic behavior using any combination of DetNet functions described in sub-layers. They are then distributed and provisioned using well-established control and provisioning mechanisms adopted for traffic engineering.

In DetNet, a considerable amount of state information is required to maintain per-flow queuing disciplines and resource reservation for a large number of individual flows. This can be quite challenging for network operations during network events, such as faults, change in traffic volume, or reprovisioning. Therefore, DetNet recommends support for aggregated flows; however, it still requires a large amount of control signaling to establish and maintain DetNet flows.

Note that DetNet might suffer from some of the scalability concerns described for Intserv in Section 5.1.1.1, but the scope of DetNet’s deployment scenarios is smaller and therefore less exposed to scaling issues.

### 5.1.2. IETF Approaches Relying on TE Mechanisms

#### 5.1.2.1. Application-Layer Traffic Optimization

This document describes various TE mechanisms available in the network. However, in general, distributed applications (particularly, bandwidth-greedy P2P applications that are used for file sharing, for example) cannot directly use those techniques. As per [RFC5693], applications could greatly improve traffic distribution and quality by cooperating with external services that are aware of the network topology. Addressing the Application-Layer Traffic Optimization (ALTO) problem means, on the one hand, deploying an ALTO service to provide applications with information regarding the underlying network (e.g., basic network location structure and preferences of network paths) and, on the other hand, enhancing applications in order to use such information to perform better-than-random selection of the endpoints with which they establish connections.

The basic function of ALTO is based on abstract maps of a network. These maps provide a simplified view, yet enough information about a network for applications to effectively utilize them. Additional services are built on top of the maps. [RFC7285] describes a protocol implementing the ALTO services as an information-publishing interface that allows a network to publish its network information to network applications. This information can include network node locations, groups of node-to-node connectivity arranged by cost according to configurable granularities, and end-host properties. The information published by the ALTO Protocol should benefit both the network and the applications. The ALTO Protocol uses a REST-ful design and encodes its requests and responses using JSON [RFC8259] with a modular design by dividing ALTO information publication into multiple ALTO services (e.g., the Map Service, the Map-Filtering Service, the Endpoint Property Service, and the Endpoint Cost Service).

[RFC8189] defines a new service that allows an ALTO Client to retrieve several cost metrics in a single request for an ALTO filtered cost map and endpoint cost map. [RFC8896] extends the ALTO cost information service so that applications decide not only "where" to connect but also "when". This is useful for applications that need to perform bulk data transfer and would like to schedule these transfers during an off-peak hour, for example. [RFC9439] introduces network performance metrics, including network delay, jitter, packet loss rate, hop count, and bandwidth. The ALTO server may derive and aggregate such performance metrics from BGP-LS (see Section 5.1.3.10), IGP-TE (see Section 5.1.3.9), or management tools and then expose the information to allow applications to determine "where" to connect based on network performance criteria. The ALTO Working Group is evaluating the use of network TE properties while making application decisions for new use cases such as edge computing and data-center interconnect.

#### 5.1.2.2. Network Virtualization and Abstraction

One of the main drivers for SDN [RFC7149] is a decoupling of the network control plane from the data plane. This separation has been achieved for TE networks with the development of MPLS and GMPLS (see Sections 5.1.3.3 and 5.1.3.5, respectively) and the PCE (see Section 5.1.3.11). One of the advantages of SDN is its logically centralized control regime that allows a full view of the underlying networks. Centralized control in SDN helps improve network resource utilization compared with distributed network control.

Abstraction and Control of TE Networks (ACTN) [RFC8453] defines a hierarchical SDN architecture that describes the functional entities and methods for the coordination of resources across multiple domains, to provide composite traffic-engineered services. ACTN facilitates composed, multi-domain connections and provides them to the user. ACTN is focused on:

* Abstraction of the underlying network resources and how they are provided to higher-layer applications and customers.

* Virtualization of underlying resources for use by the customer, application, or service. The creation of a virtualized environment allows operators to view and control multi-domain networks as a single virtualized network.

* Presentation to customers of networks as a virtual network via open and programmable interfaces.

The ACTN managed infrastructure is built from traffic-engineered network resources, which may include statistical packet bandwidth, physical forwarding-plane sources (such as wavelengths and time slots), and forwarding and cross-connect capabilities. The type of network virtualization seen in ACTN allows customers and applications (tenants) to utilize and independently control allocated virtual network resources as if they were physically their own resource. The ACTN network is sliced, with tenants being given a different partial and abstracted topology view of the physical underlying network.

#### 5.1.2.3. Network Slicing

An IETF Network Slice is a logical network topology connecting a number of endpoints using a set of shared or dedicated network resources [NETWORK-SLICES]. The resources are used to satisfy specific SLOs specified by the consumer.

IETF Network Slices are not, of themselves, TE constructs. However, a network operator that offers IETF Network Slices is likely to use many TE tools in order to manage their network and provide the

services.

IETF Network Slices are defined such that they are independent of the underlying infrastructure connectivity and technologies used. From a customer's perspective, an IETF Network Slice looks like a VPN connectivity matrix with additional information about the level of service that the customer requires between the endpoints. From an operator's perspective, the IETF Network Slice looks like a set of routing or tunneling instructions with the network resource reservations necessary to provide the required service levels as specified by the SLOs. The concept of an IETF Network Slice is consistent with an enhanced VPN [ENHANCED-VPN].

#### 5.1.3. IETF Techniques Used by TE Mechanisms

##### 5.1.3.1. Constraint-Based Routing

Constraint-based routing refers to a class of routing systems that compute routes through a network subject to the satisfaction of a set of constraints and requirements. In the most general case, constraint-based routing may also seek to optimize overall network performance while minimizing costs.

The constraints and requirements may be imposed by the network itself or by administrative policies. Constraints may include bandwidth, hop count, delay, and policy instruments such as resource class attributes. Constraints may also include domain-specific attributes of certain network technologies and contexts that impose restrictions on the solution space of the routing function. Path-oriented technologies such as MPLS have made constraint-based routing feasible and attractive in public IP networks.

The concept of constraint-based routing within the context of MPLS-TE requirements in IP networks was first described in [RFC2702] and led to developments such as MPLS-TE [RFC3209] as described in Section 5.1.3.3.

Unlike QoS-based routing (for example, see [RFC2386], [MA], and [PERFORMANCE-ROUTING]) that generally addresses the issue of routing individual traffic flows to satisfy prescribed flow-based QoS requirements subject to network resource availability, constraint-based routing is applicable to traffic aggregates as well as flows and may be subject to a wide variety of constraints that may include policy restrictions.

###### 5.1.3.1.1. IGP Flexible Algorithms

The normal approach to routing in an IGP network relies on the IGPs deriving "shortest paths" over the network based solely on the IGP metric assigned to the links. Such an approach is often limited: traffic may tend to converge toward the destination, possibly causing congestion, and it is not possible to steer traffic onto paths depending on the end-to-end qualities demanded by the applications.

To overcome this limitation, various sorts of TE have been widely deployed (as described in this document), where the TE component is responsible for computing the path based on additional metrics and/or constraints. Such paths (or tunnels) need to be installed in the routers' forwarding tables in addition to, or as a replacement for, the original paths computed by IGPs. The main drawbacks of these TE approaches are the additional complexity of protocols and management and the state that may need to be maintained within the network.

IGP Flexible Algorithms [RFC9350] allow IGPs to construct constraint-based paths over the network by computing constraint-based next hops. The intent of Flexible Algorithms is to reduce TE complexity by

letting an IGP perform some basic TE computation capabilities. Flexible Algorithm includes a set of extensions to the IGPs that enable a router to send TLVs that:

* describe a set of constraints on the topology

* identify calculation-type

* describe a metric-type that is to be used to compute the best paths through the constrained topology

A given combination of calculation-type, metric-type, and constraints is known as a Flexible Algorithm Definition (FAD). A router that sends such a set of TLVs also assigns a specific identifier (the Flexible Algorithm) to the specified combination of calculation-type, metric-type, and constraints.

There are two use cases for Flexible Algorithm: in IP networks [RFC9502] and in SR networks [RFC9350]. In the first case, Flexible Algorithm computes paths to an IPv4 or IPv6 address; in the second case, Flexible Algorithms computes paths to a Prefix SID (see Section 5.1.3.12).

Examples of where Flexible Algorithms can be useful include:

* Expansion of the function of IP performance metrics [RFC5664] where specific constraint-based routing (Flexible Algorithm) can be instantiated within the network based on the results of performance measurement.

* The formation of an "underlay" network using Flexible Algorithms, and the realization of an "overlay" network using TE techniques. This approach can leverage the nested combination of Flexible Algorithm and TE extensions for IGP (see Section 5.1.3.9).

* Flexible Algorithms in SR-MPLS (Section 5.1.3.12) can be used as a base to easily build a TE-like topology without TE components on routers or the use of a PCE (see Section 5.1.3.11).

* The support for network slices [NETWORK-SLICES] where the SLOs of a particular IETF Network Slice can be guaranteed by a Flexible Algorithm or where a Filtered Topology [NETWORK-SLICES] can be created as a TE-like topology using a Flexible Algorithm.

#### 5.1.3.2. RSVP

RSVP is a soft-state signaling protocol [RFC2205]. It supports receiver-initiated establishment of resource reservations for both multicast and unicast flows. RSVP was originally developed as a signaling protocol within the Integrated Services framework (see Section 5.1.1.1) for applications to communicate QoS requirements to the network and for the network to reserve relevant resources to satisfy the QoS requirements [RFC2205].

In RSVP, the traffic sender or source node sends a Path message to the traffic receiver with the same source and destination addresses as the traffic that the sender will generate. The Path message contains:

* A sender traffic specification describing the characteristics of the traffic

* A sender template specifying the format of the traffic

* An optional advertisement specification that is used to support the concept of One Pass With Advertising (OPWA) [RFC2205]

Every intermediate router along the path forwards the Path message to the next hop determined by the routing protocol. Upon receiving a Path message, the receiver responds with a Resv message that includes a flow descriptor used to request resource reservations. The Resv message travels to the sender or source node in the opposite direction along the path that the Path message traversed. Every intermediate router along the path can reject or accept the reservation request of the Resv message. If the request is rejected, the rejecting router will send an error message to the receiver, and the signaling process will terminate. If the request is accepted, link bandwidth and buffer space are allocated for the flow, and the related flow state information is installed in the router.

One of the issues with the original RSVP specification [RFC2205] was scalability. This was because reservations were required for micro-flows, so that the amount of state maintained by network elements tended to increase linearly with the number of traffic flows. These issues are described in [RFC2961], which also modifies and extends RSVP to mitigate the scaling problems to make RSVP a versatile signaling protocol for the Internet. For example, RSVP has been extended to reserve resources for aggregation of flows [RFC3175], to set up MPLS explicit LSPs (see Section 5.1.3.3), and to perform other signaling functions within the Internet. [RFC2961] also describes a mechanism to reduce the amount of Refresh messages required to maintain established RSVP sessions.

#### 5.1.3.3. MPLS

MPLS is a forwarding scheme that also includes extensions to conventional IP control plane protocols. MPLS extends the Internet routing model and enhances packet forwarding and path control [RFC3031].

At the ingress to an MPLS domain, LSRs classify IP packets into Forwarding Equivalence Classes (FECs) based on a variety of factors, including, e.g., a combination of the information carried in the IP header of the packets and the local routing information maintained by the LSRs. An MPLS label stack entry is then prepended to each packet according to their FECs. The MPLS label stack entry is 32 bits long and contains a 20-bit label field.

An LSR makes forwarding decisions by using the label prepended to packets as the index into a local Next Hop Label Forwarding Entry (NHLFE). The packet is then processed as specified in the NHLFE. The incoming label may be replaced by an outgoing label (label swap), and the packet may be forwarded to the next LSR. Before a packet leaves an MPLS domain, its MPLS label may be removed (label pop). An LSP is the path between an ingress LSR and an egress LSR through which a labeled packet traverses. The path of an explicit LSP is defined at the originating (ingress) node of the LSP. MPLS can use a signaling protocol such as RSVP or the Label Distribution Protocol (LDP) to set up LSPs.

MPLS is a powerful technology for Internet TE because it supports explicit LSPs that allow constraint-based routing to be implemented efficiently in IP networks [AWD2]. The requirements for TE over MPLS are described in [RFC2702]. Extensions to RSVP to support instantiation of explicit LSP are discussed in [RFC3209] and Section 5.1.3.4.

#### 5.1.3.4. RSVP-TE

RSVP-TE is a protocol extension of RSVP (Section 5.1.3.2) for traffic engineering. The base specification is found in [RFC3209]. RSVP-TE enables the establishment of traffic-engineered MPLS LSPs (TE LSPs),

using loose or strict paths and taking into consideration network constraints such as available bandwidth. The extension supports signaling LSPs on explicit paths that could be administratively specified or computed by a suitable entity (such as a PCE, Section 5.1.3.11) based on QoS and policy requirements, taking into consideration the prevailing network state as advertised by the IGP extension for IS-IS in [RFC5305], for OSPFv2 in [RFC3630], and for OSPFv3 in [RFC5329]. RSVP-TE enables the reservation of resources (for example, bandwidth) along the path.

RSVP-TE includes the ability to preempt LSPs based on priorities and uses link affinities to include or exclude links from the LSPs. The protocol is further extended to support Fast Reroute (FRR) [RFC4090], Diffserv [RFC4124], and bidirectional LSPs [RFC7551]. RSVP-TE extensions for support for GMPLS (see Section 5.1.3.5) are specified in [RFC3473].

Requirements for point-to-multipoint (P2MP) MPLS-TE LSPs are documented in [RFC4461], and signaling protocol extensions for setting up P2MP MPLS-TE LSPs via RSVP-TE are defined in [RFC4875], where a P2MP LSP comprises multiple source-to-leaf (S2L) sub-LSPs. To determine the paths for P2MP LSPs, selection of the branch points (based on capabilities, network state, and policies) is key [RFC5671]

RSVP-TE has evolved to provide real-time dynamic metrics for path selection for low-latency paths using extensions to IS-IS [RFC8570] and OSPF [RFC7471] based on performance measurements for the Simple Two-Way Active Measurement Protocol (STAMP) [RFC8972] and the Two-Way Active Measurement Protocol (TWAMP) [RFC5357].

RSVP-TE has historically been used when bandwidth was constrained; however, as bandwidth has increased, RSVP-TE has developed into a bandwidth management tool to provide bandwidth efficiency and proactive resource management.

#### 5.1.3.5. Generalized MPLS (GMPLS)

GMPLS extends MPLS control protocols to encompass time-division (e.g., Synchronous Optical Network / Synchronous Digital Hierarchy (SONET/SDH), Plesiochronous Digital Hierarchy (PDH), and Optical Transport Network (OTN)), wavelength (lambdas), and spatial switching (e.g., incoming port or fiber to outgoing port or fiber) and continues to support packet switching. GMPLS provides a common set of control protocols for all of these layers (including some technology-specific extensions), each of which has a distinct data or forwarding plane. GMPLS covers both the signaling and the routing part of that control plane and is based on the TE extensions to MPLS (see Section 5.1.3.4).

In GMPLS [RFC3945], the original MPLS architecture is extended to include LSRs whose forwarding planes rely on circuit switching and therefore cannot forward data based on the information carried in either packet or cell headers. Specifically, such LSRs include devices where the switching is based on time slots, wavelengths, or physical ports. These additions impact basic LSP properties: how labels are requested and communicated, the unidirectional nature of MPLS LSPs, how errors are propagated, and information provided for synchronizing the ingress and egress LSRs [RFC3473].

#### 5.1.3.6. IP Performance Metrics (IPPM)

The IETF IP Performance Metrics (IPPM) Working Group has developed a set of standard metrics that can be used to monitor the quality, performance, and reliability of Internet services. These metrics can be applied by network operators, end users, and independent testing groups to provide users and service providers with a common

understanding of the performance and reliability of the Internet component clouds they use/provide [RFC2330]. The criteria for performance metrics developed by the IPPM Working Group are described in [RFC2330]. Examples of performance metrics include one-way packet loss [RFC7680], one-way delay [RFC7679], and connectivity measures between two nodes [RFC2678]. Other metrics include second-order measures of packet loss and delay.

Some of the performance metrics specified by the IPPM Working Group are useful for specifying SLAs. SLAs are sets of SLOs negotiated between users and service providers, wherein each objective is a combination of one or more performance metrics, possibly subject to certain constraints.

The IPPM Working Group also designs measurement techniques and protocols to obtain these metrics.

#### 5.1.3.7. Flow Measurement

The IETF Real Time Flow Measurement (RTFM) Working Group produced an architecture that defines a method to specify traffic flows as well as a number of components for flow measurement (meters, meter readers, and managers) [RFC2722]. A flow measurement system enables network traffic flows to be measured and analyzed at the flow level for a variety of purposes. As noted in [RFC2722], a flow measurement system can be very useful in the following contexts:

* understanding the behavior of existing networks
* planning for network development and expansion
* quantification of network performance
* verifying the quality of network service
* attribution of network usage to users

A flow measurement system consists of meters, meter readers, and managers. A meter observes packets passing through a measurement point, classifies them into groups, accumulates usage data (such as the number of packets and bytes for each group), and stores the usage data in a flow table. A group may represent any collection of user applications, hosts, networks, etc. A meter reader gathers usage data from various meters so it can be made available for analysis. A manager is responsible for configuring and controlling meters and meter readers. The instructions received by a meter from a manager include flow specifications, meter control parameters, and sampling techniques. The instructions received by a meter reader from a manager include the address of the meter whose data are to be collected, the frequency of data collection, and the types of flows to be collected.

IP Flow Information Export (IPFIX) [RFC5470] defines an architecture that is very similar to the RTFM architecture and includes Metering, Exporting, and Collecting Processes. [RFC5472] describes the applicability of IPFIX and makes a comparison with RTFM, pointing out that, architecturally, while RTM talks about devices, IPFIX deals with processes to clarify that multiple of those processes may be co-located on the same machine. The IPFIX protocol [RFC7011] is widely implemented.

#### 5.1.3.8. Endpoint Congestion Management

[RFC3124] provides a set of congestion control mechanisms for the use of transport protocols. It also allows the development of mechanisms for unifying congestion control across a subset of an endpoint’s

active unicast connections (called a "congestion group"). A congestion manager continuously monitors the state of the path for each congestion group under its control. The manager uses that information to instruct a scheduler on how to partition bandwidth among the connections of that congestion group.

The concepts described in [RFC3124] and the lessons that can be learned from that work found a home in HTTP/2 [RFC9113] and QUIC [RFC9000], while [RFC9040] describes TCP control block interdependence that is a core construct underpinning the congestion manager defined in [RFC3124].

#### 5.1.3.9. TE Extensions to the IGPs

[RFC5305] describes the extensions to the Intermediate System to Intermediate System (IS-IS) protocol to support TE. Similarly, [RFC3630] specifies TE extensions for OSPFv2, and [RFC5329] has the same description for OSPFv3.

IS-IS and OSPF share the common concept of TE extensions to distribute TE parameters, such as link type and ID, local and remote IP addresses, TE metric, maximum bandwidth, maximum reservable bandwidth, unreserved bandwidth, and admin group. The information distributed by the IGPs in this way can be used to build a view of the state and capabilities of a TE network (see Section 5.1.3.14).

The difference between IS-IS and OSPF is in the details of how they encode and transmit the TE parameters:

* IS-IS uses the Extended IS Reachability TLV (type 22), the Extended IP Reachability TLV (type 135), and the Traffic Engineering router ID TLV (type 134). These TLVs use specific sub-TLVs described in [RFC8570] to carry the TE parameters.

* OSPFv2 uses Opaque LSA [RFC5250] type 10, and OSPFv3 uses the Intra-Area-TE-LSA. In both OSPF cases, two top-level TLVs are used (Router Address and Link TLVs), and these use sub-TLVs to carry the TE parameters (as defined in [RFC7471] for OSPFv2 and [RFC5329] for OSPFv3).

#### 5.1.3.10. BGP - Link State

In a number of environments, a component external to a network is called upon to perform computations based on the network topology and current state of the connections within the network, including TE information. This is information typically distributed by IGP routing protocols within the network (see Section 5.1.3.9).

BGP (see also Section 7) is one of the essential routing protocols that glues the Internet together. BGP-LS [RFC9552] is a mechanism by which link-state and TE information can be collected from networks and shared with external components using the BGP routing protocol. The mechanism is applicable to physical and virtual IGP links and is subject to policy control.

Information collected by BGP-LS can be used, for example, to construct the TED (Section 5.1.3.14) for use by the PCE (see Section 5.1.3.11) or may be used by ALTO servers (see Section 5.1.2.1).

#### 5.1.3.11. Path Computation Element

Constraint-based path computation is a fundamental building block for TE in MPLS and GMPLS networks. Path computation in large, multi-domain networks is complex and may require special computational components and cooperation between the elements in different domains.

The PCE [RFC4655] is an entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

Thus, a PCE can provide a central component in a TE system operating on the TED (see Section 5.1.3.14) with delegated responsibility for determining paths in MPLS, GMPLS, or SR networks. The PCE uses the Path Computation Element Communication Protocol (PCEP) [RFC5440] to communicate with Path Computation Clients (PCCs), such as MPLS LSRs, to answer their requests for computed paths or to instruct them to initiate new paths [RFC8281] and maintain state about paths already installed in the network [RFC8231].

PCEs form key components of a number of TE systems. More information about the applicability of PCEs can be found in [RFC8051], while [RFC6805] describes the application of PCEs to determining paths across multiple domains. PCEs also have potential uses in Abstraction and Control of TE Networks (ACTN) (see Section 5.1.2.2), Centralized Network Control [RFC8283], and SDN (see Section 4.3.2).

#### 5.1.3.12. Segment Routing (SR)

The SR architecture [RFC8402] leverages the source routing and tunneling paradigms. The path a packet takes is defined at the ingress, and the packet is tunneled to the egress.

In a protocol realization, an ingress node steers a packet using a set of instructions, called "segments", that are included in an SR header prepended to the packet: a label stack in MPLS case, and a series of 128-bit SIDs in the IPv6 case.

Segments are identified by SIDs. There are four types of SIDs that are relevant for TE.

* Prefix SID: A SID that is unique within the routing domain and is used to identify a prefix.

* Node SID: A Prefix SID with the "N" bit set to identify a node.

* Adjacency SID: Identifies a unidirectional adjacency.

* Binding SID: A Binding SID has two purposes:

  1. To advertise the mappings of prefixes to SIDs/Labels

  2. To advertise a path available for a Forwarding Equivalence Class (FEC)

A segment can represent any instruction, topological or service-based. SIDs can be looked up in a global context (domain-wide) as well as in some other contexts (see, for example, "context labels" in Section 3 of [RFC5331]).

The application of policy to SR can make SR into a TE mechanism, as described in Section 5.1.1.3.

#### 5.1.3.13. Tree Engineering for Bit Index Explicit Replication

Bit Index Explicit Replication (BIER) [RFC8279] specifies an encapsulation for multicast forwarding that can be used on MPLS or Ethernet transports. A mechanism known as Tree Engineering for Bit Index Explicit Replication (BIER-TE) [RFC9262] provides a component that could be used to build a traffic-engineered multicast system. BIER-TE does not of itself offer full traffic engineering, and the abbreviation "TE" does not, in this case, refer to traffic engineering.

In BIER-TE, path steering is supported via the definition of a bitstring attached to each packet that determines how the packet is forwarded and replicated within the network. Thus, this bitstring steers the traffic within the network and forms an element of a traffic-engineering system. A central controller that is aware of the capabilities and state of the network as well as the demands of the various traffic flows is able to select multicast paths that take account of the available resources and demands. Therefore, this controller is responsible for the policy elements of traffic engineering.

Resource management has implications for the forwarding plane beyond the steering of packets defined for BIER-TE. These include the allocation of buffers to meet the requirements of admitted traffic and may include policing and/or rate-shaping mechanisms achieved via various forms of queuing. This level of resource control, while optional, is important in networks that wish to support congestion management policies to control or regulate the offered traffic to deliver different levels of service and alleviate congestion problems. It is also important in networks that wish to control latencies experienced by specific traffic flows.

#### 5.1.3.14. Network TE State Definition and Presentation

The network states that are relevant to TE need to be stored in the system and presented to the user. The TED is a collection of all TE information about all TE nodes and TE links in the network. It is an essential component of a TE system, such as MPLS-TE [RFC2702] or GMPLS [RFC3945]. In order to formally define the data in the TED and to present the data to the user, the data modeling language YANG [RFC7950] can be used as described in [RFC8795].

#### 5.1.3.15. System Management and Control Interfaces

The TE control system needs to have a management interface that is human-friendly and a control interface that is programmable for automation. The Network Configuration Protocol (NETCONF) [RFC6241] and the RESTCONF protocol [RFC8040] provide programmable interfaces that are also human-friendly. These protocols use XML- or JSON-encoded messages. When message compactness or protocol bandwidth consumption needs to be optimized for the control interface, other protocols, such as Group Communication for the Constrained Application Protocol (CoAP) [RFC7390] or gRPC [GRPC], are available, especially when the protocol messages are encoded in a binary format. Along with any of these protocols, the data modeling language YANG [RFC7950] can be used to formally and precisely define the interface data.

PCEP [RFC5440] is another protocol that has evolved to be an option for the TE system control interface. PCEP messages are TLV based; they are not defined by a data-modeling language such as YANG.

### 5.2. Content Distribution

The Internet is dominated by client-server interactions, principally web traffic and multimedia streams, although in the future, more sophisticated media servers may become dominant. The location and performance of major information servers have a significant impact on the traffic patterns within the Internet as well as on the perception of service quality by end users.

A number of dynamic load-balancing techniques have been devised to improve the performance of replicated information servers. These techniques can cause spatial traffic characteristics to become more dynamic in the Internet because information servers can be

dynamically picked based upon the location of the clients, the location of the servers, the relative utilization of the servers, the relative performance of different networks, and the relative performance of different parts of a network. This process of assignment of distributed servers to clients is called "traffic directing". It is an application-layer function.

Traffic-directing schemes that allocate servers in multiple geographically dispersed locations to clients may require empirical network performance statistics to make more effective decisions. In the future, network measurement systems may need to provide this type of information.

When congestion exists in the network, traffic-directing and traffic-engineering systems should act in a coordinated manner. This topic is for further study.

The issues related to location and replication of information servers, particularly web servers, are important for Internet traffic engineering because these servers contribute a substantial proportion of Internet traffic.

## 6. Recommendations for Internet Traffic Engineering

This section describes high-level recommendations for traffic engineering in the Internet in general terms.

The recommendations describe the capabilities needed to solve a TE problem or to achieve a TE objective. Broadly speaking, these recommendations can be categorized as either functional or non-functional recommendations:

* Functional recommendations describe the functions that a traffic-engineering system should perform. These functions are needed to realize TE objectives by addressing traffic-engineering problems.

* Non-functional recommendations relate to the quality attributes or state characteristics of a TE system. These recommendations may contain conflicting assertions and may sometimes be difficult to quantify precisely.

The subsections that follow first summarize the non-functional requirements and then detail the functional requirements.

### 6.1. Generic Non-functional Recommendations

The generic non-functional recommendations for Internet traffic engineering are listed in the paragraphs that follow. In a given context, some of these recommendations may be critical while others may be optional. Therefore, prioritization may be required during the development phase of a TE system to tailor it to a specific operational context.

Automation: Whenever feasible, a TE system should automate as many TE functions as possible to minimize the amount of human effort needed to analyze and control operational networks. Automation is particularly important in large-scale public networks because of the high cost of the human aspects of network operations and the high risk of network problems caused by human errors. Automation may additionally benefit from feedback from the network that indicates the state of network resources and the current load in the network. Further, placing intelligence into components of the TE system could enable automation to be more dynamic and responsive to changes in the network.

Flexibility: A TE system should allow for changes in optimization

policy. In particular, a TE system should provide sufficient configuration options so that a network administrator can tailor the system to a particular environment. It may also be desirable to have both online and offline TE subsystems that can be independently enabled and disabled. TE systems that are used in multi-class networks should also have options to support class-based performance evaluation and optimization.

Interoperability: Whenever feasible, TE systems and their components should be developed with open standards-based interfaces to allow interoperation with other systems and components.

Scalability: Public networks continue to grow rapidly with respect to network size and traffic volume. Therefore, to remain applicable as the network evolves, a TE system should be scalable. In particular, a TE system should remain functional as the network expands with regard to the number of routers and links and with respect to the number of flows and the traffic volume. A TE system should have a scalable architecture, should not adversely impair other functions and processes in a network element, and should not consume too many network resources when collecting and distributing state information or when exerting control.

Security: Security is a critical consideration in TE systems. Such systems typically exert control over functional aspects of the network to achieve the desired performance objectives. Therefore, adequate measures must be taken to safeguard the integrity of the TE system. Adequate measures must also be taken to protect the network from vulnerabilities that originate from security breaches and other impairments within the TE system.

Simplicity: A TE system should be as simple as possible. Simplicity in user interface does not necessarily imply that the TE system will use naive algorithms. When complex algorithms and internal structures are used, the user interface should hide such complexities from the network administrator as much as possible.

Stability: Stability refers to the resistance of the network to oscillate (flap) in a disruptive manner from one state to another, which may result in traffic being routed first one way and then another without satisfactory resolution of the underlying TE issues and with continued changes that do not settle down. Stability is a very important consideration in TE systems that respond to changes in the state of the network. State-dependent TE methodologies typically include a trade-off between responsiveness and stability. It is strongly recommended that when a trade-off between responsiveness and stability is needed, it should be made in favor of stability (especially in public IP backbone networks).

Usability: Usability is a human aspect of TE systems. It refers to the ease with which a TE system can be deployed and operated. In general, it is desirable to have a TE system that can be readily deployed in an existing network. It is also desirable to have a TE system that is easy to operate and maintain.

Visibility: Mechanisms should exist as part of the TE system to collect statistics from the network and to analyze these statistics to determine how well the network is functioning. Derived statistics (such as traffic matrices, link utilization, latency, packet loss, and other performance measures of interest) that are determined from network measurements can be used as indicators of prevailing network conditions. The capabilities of the various components of the routing system are other examples of status information that should be observable.

## 6.2. Routing Recommendations

Routing control is a significant aspect of Internet traffic engineering. Routing impacts many of the key performance measures associated with networks, such as throughput, delay, and utilization. Generally, it is very difficult to provide good service quality in a wide area network without effective routing control. A desirable TE routing system is one that takes traffic characteristics and network constraints into account during route selection while maintaining stability.

Shortest Path First (SPF) IGPs are based on shortest path algorithms and have limited control capabilities for TE [RFC2702] [AWD2]. These limitations include:

1. Pure SPF protocols do not take network constraints and traffic characteristics into account during route selection. For example, IGPs always select the shortest paths based on link metrics assigned by administrators, so load sharing cannot be performed across paths of different costs. Note that link metrics are assigned following a range of operator-selected policies that might reflect preference for the use of some links over others; therefore, "shortest" might not be purely a measure of distance. Using shortest paths to forward traffic may cause the following problems:

    * If traffic from a source to a destination exceeds the capacity of a link along the shortest path, the link (and hence the shortest path) becomes congested while a longer path between these two nodes may be under-utilized.

    * The shortest paths from different sources can overlap at some links. If the total traffic from the sources exceeds the capacity of any of these links, congestion will occur.

    * Problems can also occur because traffic demand changes over time, but network topology and routing configuration cannot be changed as rapidly. This causes the network topology and routing configuration to become sub-optimal over time, which may result in persistent congestion problems.

2. The Equal-Cost Multipath (ECMP) capability of SPF IGPs supports sharing of traffic among equal-cost paths. However, ECMP attempts to divide the traffic as equally as possible among the equal-cost shortest paths. Generally, ECMP does not support configurable load-sharing ratios among equal-cost paths. The result is that one of the paths may carry significantly more traffic than other paths because it may also carry traffic from other sources. This situation can result in congestion along the path that carries more traffic. Weighted ECMP (WECMP) (see, for example, [EVPN-UNEQUAL-LB]) provides some mitigation.

3. Modifying IGP metrics to control traffic routing tends to have network-wide effects. Consequently, undesirable and unanticipated traffic shifts can be triggered as a result. Work described in Section 8 may be capable of better control [FT00] [FT01].

Because of these limitations, capabilities are needed to enhance the routing function in IP networks. Some of these capabilities are summarized below:

* Constraint-based routing computes routes to fulfill requirements subject to constraints. This can be useful in public IP backbones with complex topologies. Constraints may include bandwidth, hop count, delay, and administrative policy instruments, such as

resource class attributes [RFC2702] [RFC2386]. This makes it possible to select routes that satisfy a given set of requirements. Routes computed by constraint-based routing are not necessarily the shortest paths. Constraint-based routing works best with path-oriented technologies that support explicit routing, such as MPLS.

* Constraint-based routing can also be used as a way to distribute traffic onto the infrastructure, including for best-effort traffic. For example, congestion problems caused by uneven traffic distribution may be avoided or reduced by knowing the reservable bandwidth attributes of the network links and by specifying the bandwidth requirements for path selection.

* A number of enhancements to the link-state IGPs allow them to distribute additional state information required for constraint-based routing. The extensions to OSPF are described in [RFC3630], and the extensions to IS-IS are described in [RFC5305]. Some of the additional topology state information includes link attributes, such as reservable bandwidth and link resource class attribute (an administratively specified property of the link). The resource class attribute concept is defined in [RFC2702]. The additional topology state information is carried in new TLVs and sub-TLVs in IS-IS [RFC5305] or in the Opaque LSA in OSPF [RFC3630].

* An enhanced link-state IGP may flood information more frequently than a normal IGP. This is because even without changes in topology, changes in reservable bandwidth or link affinity can trigger the enhanced IGP to initiate flooding. A trade-off between the timeliness of the information flooded and the flooding frequency is typically implemented using a threshold based on the percentage change of the advertised resources to avoid excessive consumption of link bandwidth and computational resources and to avoid instability in the TED.

* In a TE system, it is also desirable for the routing subsystem to make the load-splitting ratio among multiple paths (with equal cost or different cost) configurable. This capability gives network administrators more flexibility in the control of traffic distribution across the network. It can be very useful for avoiding/relieving congestion in certain situations. Examples can be found in [XIAO] and [EVPN-UNEQUAL-LB].

* The routing system should also have the capability to control the routes of subsets of traffic without affecting the routes of other traffic if sufficient resources exist for this purpose. This capability allows a more refined control over the distribution of traffic across the network. For example, the ability to move traffic away from its original path to another path (without affecting other traffic paths) allows the traffic to be moved from resource-poor network segments to resource-rich segments. Path-oriented technologies, such as MPLS-TE, inherently support this capability as discussed in [AWD2].

* Additionally, the routing subsystem should be able to select different paths for different classes of traffic (or for different traffic behavior aggregates) if the network supports multiple classes of service (different behavior aggregates).

### 6.3. Traffic Mapping Recommendations

Traffic mapping is the assignment of traffic workload onto (pre-established) paths to meet certain requirements. Thus, while constraint-based routing deals with path selection, traffic mapping deals with the assignment of traffic to established paths that may

have been generated by constraint-based routing or by some other means. Traffic mapping can be performed by time-dependent or state-dependent mechanisms, as described in Section 4.1.

Two important aspects of the traffic mapping function are the ability to establish multiple paths between an originating node and a destination node and the capability to distribute the traffic across those paths according to configured policies. A precondition for this scheme is the existence of flexible mechanisms to partition traffic and then assign the traffic partitions onto the parallel paths (described as "parallel traffic trunks" in [RFC2702]). When traffic is assigned to multiple parallel paths, it is recommended that special care should be taken to ensure proper ordering of packets belonging to the same application (or traffic flow) at the destination node of the parallel paths.

Mechanisms that perform the traffic mapping functions should aim to map the traffic onto the network infrastructure to minimize congestion. If the total traffic load cannot be accommodated, or if the routing and mapping functions cannot react fast enough to changing traffic conditions, then a traffic mapping system may use short timescale congestion control mechanisms (such as queue management, scheduling, etc.) to mitigate congestion. Thus, mechanisms that perform the traffic mapping functions complement existing congestion control mechanisms. In an operational network, traffic should be mapped onto the infrastructure such that intra-class and inter-class resource contention are minimized (see Section 2).

When traffic mapping techniques that depend on dynamic state feedback (e.g., MPLS Adaptive Traffic Engineering (MATE) [MATE] and suchlike) are used, special care must be taken to guarantee network stability.

### 6.4. Measurement Recommendations

The importance of measurement in TE has been discussed throughout this document. A TE system should include mechanisms to measure and collect statistics from the network to support the TE function. Additional capabilities may be needed to help in the analysis of the statistics. The actions of these mechanisms should not adversely affect the accuracy and integrity of the statistics collected. The mechanisms for statistical data acquisition should also be able to scale as the network evolves.

Traffic statistics may be classified according to long-term or short-term timescales. Long-term traffic statistics are very useful for traffic engineering. Long-term traffic statistics may periodically record network workload (such as hourly, daily, and weekly variations in traffic profiles) as well as traffic trends. Aspects of the traffic statistics may also describe class of service characteristics for a network supporting multiple classes of service. Analysis of the long-term traffic statistics may yield other information such as busy-hour characteristics, traffic growth patterns, persistent congestion problems, hotspots, and imbalances in link utilization caused by routing anomalies.

A mechanism for constructing traffic matrices for both long-term and short-term traffic statistics should be in place. In multi-service IP networks, the traffic matrices may be constructed for different service classes. Each element of a traffic matrix represents a statistic about the traffic flow between a pair of abstract nodes. An abstract node may represent a router, a collection of routers, or a site in a VPN.

Traffic statistics should provide reasonable and reliable indicators of the current state of the network on the short-term scale. Some

short-term traffic statistics may reflect link utilization and link congestion status. Examples of congestion indicators include excessive packet delay, packet loss, and high resource utilization. Examples of mechanisms for distributing this kind of information include SNMP, probing tools, FTP, IGP link-state advertisements, NETCONF/RESTCONF, etc.

### 6.5. Policing, Planning, and Access Control

The recommendations in Sections 6.2 and 6.3 may be sub-optimal or even ineffective if the amount of traffic flowing on a route or path exceeds the capacity of the resource on that route or path. Several approaches can be used to increase the performance of TE systems:

* The fundamental approach is some form of planning where traffic is steered onto paths so that it is distributed across the available resources. This planning may be centralized or distributed and must be aware of the planned traffic volumes and available resources. However, this approach is only of value if the traffic that is presented conforms to the planned traffic volumes.

* Traffic flows may be policed at the edges of a network. This is a simple way to ensure that the actual traffic volumes are consistent with the planned volumes. Some form of measurement (see Section 6.4) is used to determine the rate of arrival of traffic, and excess traffic could be discarded. Alternatively, excess traffic could be forwarded as best-effort within the network. However, this approach is only completely effective if the planning is stringent and network-wide and if a harsh approach is taken to disposing of excess traffic.

* Resource-based admission control is the process whereby network nodes decide whether to grant access to resources. The basis for the decision on a packet-by-packet basis is the determination of the flow to which the packet belongs. This information is combined with policy instructions that have been locally configured or installed through the management or control planes. The end result is that a packet may be allowed to access (or use) specific resources on the node if, and only if, the flow to which the packet belongs conforms to the policy.

Combining some elements of all three of these measures is advisable to achieve a better TE system.

### 6.6. Network Survivability

Network survivability refers to the capability of a network to maintain service continuity in the presence of faults. This can be accomplished by promptly recovering from network impairments and maintaining the required QoS for existing services after recovery. Survivability is an issue of great concern within the Internet community due to the demand to carry mission-critical traffic, real-time traffic, and other high-priority traffic over the Internet. Survivability can be addressed at the device level by developing network elements that are more reliable and at the network level by incorporating redundancy into the architecture, design, and operation of networks. It is recommended that a philosophy of robustness and survivability should be adopted in the architecture, design, and operation of TE used to control IP networks (especially public IP networks). Because different contexts may demand different levels of survivability, the mechanisms developed to support network survivability should be flexible so that they can be tailored to different needs. A number of tools and techniques have been developed to enable network survivability, including MPLS Fast Reroute [RFC4090], Topology Independent Loop-free Alternate Fast Reroute for Segment Routing [SR-TI-LFA], RSVP-TE Extensions in

Support of End-to-End GMPLS Recovery [RFC4872], and GMPLS Segment Recovery [RFC4873].

The impact of service outages varies significantly for different service classes depending on the duration of the outage, which can vary from milliseconds (with minor service impact) to seconds (with possible call drops for IP telephony and session timeouts for connection-oriented transactions) to minutes and hours (with potentially considerable social and business impact). Outages of different durations have different impacts depending on the nature of the traffic flows that are interrupted.

Failure protection and restoration capabilities are available in multiple layers as network technologies have continued to evolve. Optical networks are capable of providing dynamic ring and mesh restoration functionality at the wavelength level. At the SONET/SDH layer, survivability capability is provided with Automatic Protection Switching (APS) as well as self-healing ring and mesh architectures. Similar functionality is provided by Layer 2 technologies such as Ethernet.

Rerouting is used at the IP layer to restore service following link and node outages. Rerouting at the IP layer occurs after a period of routing convergence, which may require seconds to minutes to complete. Path-oriented technologies such as MPLS [RFC3469] can be used to enhance the survivability of IP networks in a potentially cost-effective manner.

An important aspect of multi-layer survivability is that technologies at different layers may provide protection and restoration capabilities at different granularities in terms of timescales and at different bandwidth granularities (from the level of packets to that of wavelengths). Protection and restoration capabilities can also be sensitive to different service classes and different network utility models. Coordinating different protection and restoration capabilities across multiple layers in a cohesive manner to ensure network survivability is maintained at reasonable cost is a challenging task. Protection and restoration coordination across layers may not always be feasible, because networks at different layers may belong to different administrative domains.

Some of the general recommendations for protection and restoration coordination are as follows:

* Protection and restoration capabilities from different layers should be coordinated to provide network survivability in a flexible and cost-effective manner. Avoiding duplication of functions in different layers is one way to achieve the coordination. Escalation of alarms and other fault indicators from lower to higher layers may also be performed in a coordinated manner. The order of timing of restoration triggers from different layers is another way to coordinate multi-layer protection/restoration.

* Network capacity reserved in one layer to provide protection and restoration is not available to carry traffic in a higher layer: it is not visible as spare capacity in the higher layer. Placing protection/restoration functions in many layers may increase redundancy and robustness, but it can result in significant inefficiencies in network resource utilization. Careful planning is needed to balance the trade-off between the desire for survivability and the optimal use of resources.

* It is generally desirable to have protection and restoration schemes that are intrinsically bandwidth efficient.

* Failure notifications throughout the network should be timely and reliable if they are to be acted on as triggers for effective protection and restoration actions.

* Alarms and other fault monitoring and reporting capabilities should be provided at the right network layers so that the protection and restoration actions can be taken in those layers.

#### 6.6.1. Survivability in MPLS-Based Networks

Because MPLS is path-oriented, it has the potential to provide faster and more predictable protection and restoration capabilities than conventional hop-by-hop routed IP systems. Protection types for MPLS networks can be divided into four categories:

Link Protection: The objective of link protection is to protect an LSP from the failure of a given link. Under link protection, a protection or backup LSP (the secondary LSP) follows a path that is disjoint from the path of the working or operational LSP (the primary LSP) at the particular link where link protection is required. When the protected link fails, traffic on the working LSP is switched to the protection LSP at the headend of the failed link. As a local repair method, link protection can be fast. This form of protection may be most appropriate in situations where some network elements along a given path are known to be less reliable than others.

Node Protection: The objective of node protection is to protect an LSP from the failure of a given node. Under node protection, the secondary LSP follows a path that is disjoint from the path of the primary LSP at the particular node where node protection is required. The secondary LSP is also disjoint from the primary LSP at all links attached to the node to be protected. When the protected node fails, traffic on the working LSP is switched over to the protection LSP at the upstream LSR directly connected to the failed node. Node protection covers a slightly larger part of the network compared to link protection but is otherwise fundamentally the same.

Path Protection: The goal of LSP path protection (or end-to-end protection) is to protect an LSP from any failure along its routed path. Under path protection, the path of the protection LSP is completely disjoint from the path of the working LSP. The advantage of path protection is that the backup LSP protects the working LSP from all possible link and node failures along the path, except for failures of ingress or egress LSR. Additionally, path protection may be more efficient in terms of resource usage than link or node protection applied at every hop along the path. However, path protection may be slower than link and node protection because the fault notifications have to be propagated further.

Segment Protection: An MPLS domain may be partitioned into multiple subdomains (protection domains). Path protection is applied to the path of each LSP as it crosses the domain from its ingress to the domain to where it egresses the domain. In cases where an LSP traverses multiple protection domains, a protection mechanism within a domain only needs to protect the segment of the LSP that lies within the domain. Segment protection will generally be faster than end-to-end path protection because recovery generally occurs closer to the fault, and the notification doesn’t have to propagate as far.

See [RFC3469] and [RFC6372] for a more comprehensive discussion of MPLS-based recovery.

### 6.6.2. Protection Options

Another issue to consider is the concept of protection options. We use notation such as "m:n protection", where m is the number of protection LSPs used to protect n working LSPs. In all cases except 1+1 protection, the resources associated with the protection LSPs can be used to carry preemptable best-effort traffic when the working LSP is functioning correctly.

1:1 protection: One working LSP is protected/restored by one protection LSP. Traffic is sent only on the protected LSP until the protection/restoration event switches the traffic to the protection LSP.

1:n protection: One protection LSP is used to protect/restore n working LSPs. Traffic is sent only on the n protected working LSPs until the protection/restoration event switches the traffic from one failed LSP to the protection LSP. Only one failed LSP can be restored at any time.

n:1 protection: One working LSP is protected/restored by n protection LSPs, possibly with load splitting across the protection LSPs. This may be especially useful when it is not feasible to find one path for the backup that can satisfy the bandwidth requirement of the primary LSP.

1+1 protection: Traffic is sent concurrently on both the working LSP and a protection LSP. The egress LSR selects one of the two LSPs based on local policy (usually based on traffic integrity). When a fault disrupts the traffic on one LSP, the egress switches to receive traffic from the other LSP. This approach is expensive in how it consumes network but recovers from failures most rapidly.

## 6.7. Multi-Layer Traffic Engineering

Networks are often implemented as layers. A layer relationship may represent the interaction between technologies (for example, an IP network operated over an optical network) or the relationship between different network operators (for example, a customer network operated over a service provider’s network). Note that a multi-layer network does not imply the use of multiple technologies, although some form of encapsulation is often applied.

Multi-layer traffic engineering presents a number of challenges associated with scalability and confidentiality. These issues are addressed in [RFC7926], which discusses the sharing of information between domains through policy filters, aggregation, abstraction, and virtualization. That document also discusses how existing protocols can support this scenario with special reference to BGP-LS (see Section 5.1.3.10).

PCE (see Section 5.1.3.11) is also a useful tool for multi-layer networks as described in [RFC6805], [RFC8685], and [RFC5623]. Signaling techniques for multi-layer TE are described in [RFC6107].

See also Section 6.6 for examination of multi-layer network survivability.

## 6.8. Traffic Engineering in Diffserv Environments

Increasing requirements to support multiple classes of traffic in the Internet, such as best-effort and mission-critical data, call for IP networks to differentiate traffic according to some criteria and to give preferential treatment to certain types of traffic. Large numbers of flows can be aggregated into a few behavior aggregates based on some criteria based on common performance requirements in

terms of packet loss ratio, delay, and jitter or in terms of common fields within the IP packet headers.

Differentiated Services (Diffserv) [RFC2475] can be used to ensure that SLAs defined to differentiate between traffic flows are met. Classes of service can be supported in a Diffserv environment by concatenating Per-Hop Behaviors (PHBs) along the routing path. A PHB is the forwarding behavior that a packet receives at a Diffserv-compliant node, and it can be configured at each router. PHBs are delivered using buffer-management and packet-scheduling mechanisms and require that the ingress nodes use traffic classification, marking, policing, and shaping.

TE can complement Diffserv to improve utilization of network resources. TE can be operated on an aggregated basis across all service classes [RFC3270] or on a per-service-class basis. The former is used to provide better distribution of the traffic load over the network resources (see [RFC3270] for detailed mechanisms to support aggregate TE). The latter case is discussed below since it is specific to the Diffserv environment, with so-called Diffserv-aware traffic engineering [RFC4124].

For some Diffserv networks, it may be desirable to control the performance of some service classes by enforcing relationships between the traffic workload contributed by each service class and the amount of network resources allocated or provisioned for that service class. Such relationships between demand and resource allocation can be enforced using a combination of, for example:

* TE mechanisms on a per-service-class basis that enforce the relationship between the amount of traffic contributed by a given service class and the resources allocated to that class.

* Mechanisms that dynamically adjust the resources allocated to a given service class to relate to the amount of traffic contributed by that service class.

It may also be desirable to limit the performance impact of high-priority traffic on relatively low-priority traffic. This can be achieved, for example, by controlling the percentage of high-priority traffic that is routed through a given link. Another way to accomplish this is to increase link capacities appropriately so that lower-priority traffic can still enjoy adequate service quality. When the ratio of traffic workload contributed by different service classes varies significantly from router to router, it may not be enough to rely on conventional IGP routing protocols or on TE mechanisms that are not sensitive to different service classes. Instead, it may be desirable to perform TE, especially routing control and mapping functions, on a per-service-class basis. One way to accomplish this in a domain that supports both MPLS and Diffserv is to define class-specific LSPs and to map traffic from each class onto one or more LSPs that correspond to that service class. An LSP corresponding to a given service class can then be routed and protected/restored in a class-dependent manner, according to specific policies.

Performing TE on a per-class basis may require per-class parameters to be distributed. It is common to have some classes share some aggregate constraints (e.g., maximum bandwidth requirement) without enforcing the constraint on each individual class. These classes can be grouped into class types, and per-class-type parameters can be distributed to improve scalability. This also allows better bandwidth sharing between classes in the same class type. A class type is a set of classes that satisfy the following two conditions:

* Classes in the same class type have common aggregate requirements

to satisfy required performance levels.

* There is no requirement to be enforced at the level of an individual class in the class type. Note that it is, nevertheless, still possible to implement some priority policies for classes in the same class type to permit preferential access to the class type bandwidth through the use of preemption priorities.

See [RFC4124] for detailed requirements on Diffserv-aware TE.

### 6.9. Network Controllability

Offline and online (see Section 4.2) TE considerations are of limited utility if the network cannot be controlled effectively to implement the results of TE decisions and to achieve the desired network performance objectives.

Capacity augmentation is a coarse-grained solution to TE issues. However, it is simple, may be applied through creating parallel links that form part of an ECMP scheme, and may be advantageous if bandwidth is abundant and cheap. However, bandwidth is not always abundant and cheap, and additional capacity might not always be the best solution. Adjustments of administrative weights and other parameters associated with routing protocols provide finer-grained control, but this approach is difficult to use and imprecise because of the way the routing protocols interact across the network.

Control mechanisms can be manual (e.g., static configuration), partially automated (e.g., scripts), or fully automated (e.g., policy-based management systems). Automated mechanisms are particularly useful in large-scale networks. Multi-vendor interoperability can be facilitated by standardized management tools (e.g., YANG models) to support the control functions required to address TE objectives.

Network control functions should be secure, reliable, and stable as these are often needed to operate correctly in times of network impairments (e.g., during network congestion or attacks).

## 7. Inter-Domain Considerations

Inter-domain TE is concerned with performance optimization for traffic that originates in one administrative domain and terminates in a different one.

BGP [RFC4271] is the standard exterior gateway protocol used to exchange routing information between ASes in the Internet. BGP includes a decision process that calculates the preference for routes to a given destination network. There are two fundamental aspects to inter-domain TE using BGP:

Route Propagation: Controlling the import and export of routes between ASes and controlling the redistribution of routes between BGP and other protocols within an AS.

Best-path selection: Selecting the best path when there are multiple candidate paths to a given destination network. This is performed by the BGP decision process, which selects the preferred exit points out of an AS toward specific destination networks by taking a number of different considerations into account. The BGP path selection process can be influenced by manipulating the attributes associated with the process, including NEXT_HOP, LOCAL_PREF, AS_PATH, ORIGIN, MULTI_EXIT_DISC (MED), IGP metric, etc.

Most BGP implementations provide constructs that facilitate the

implementation of complex BGP policies based on pre-configured logical conditions. These can be used to control import and export of incoming and outgoing routes, control redistribution of routes between BGP and other protocols, and influence the selection of best paths by manipulating the attributes (either standardized or local to the implementation) associated with the BGP decision process.

When considering inter-domain TE with BGP, note that the outbound traffic exit point is controllable, whereas the interconnection point where inbound traffic is received typically is not. Therefore, it is up to each individual network to implement TE strategies that deal with the efficient delivery of outbound traffic from its customers to its peering points. The vast majority of TE policy is based on a "closest exit" strategy, which offloads inter-domain traffic at the nearest outbound peering point towards the destination AS. Most methods of manipulating the point at which inbound traffic enters are either ineffective or not accepted in the peering community.

Inter-domain TE with BGP is generally effective, but it is usually applied in a trial-and-error fashion because a TE system usually only has a view of the available network resources within one domain (an AS in this case). A systematic approach for inter-domain TE requires cooperation between the domains. Further, what may be considered a good solution in one domain may not necessarily be a good solution in another. Moreover, it is generally considered inadvisable for one domain to permit a control process from another domain to influence the routing and management of traffic in its network.

MPLS-TE tunnels (LSPs) can add a degree of flexibility in the selection of exit points for inter-domain routing by applying the concept of relative and absolute metrics. If BGP attributes are defined such that the BGP decision process depends on IGP metrics to select exit points for inter-domain traffic, then some inter-domain traffic destined to a given peer network can be made to prefer a specific exit point by establishing a TE tunnel between the router making the selection and the peering point via a TE tunnel and assigning the TE tunnel a metric that is smaller than the IGP cost to all other peering points. RSVP-TE protocol extensions for inter-domain MPLS and GMPLS are described in [RFC5151].

Similarly to intra-domain TE, inter-domain TE is best accomplished when a traffic matrix can be derived to depict the volume of traffic from one AS to another.

Layer 4 multipath transport protocols are designed to move traffic between domains and to allow some influence over the selection of the paths. To be truly effective, these protocols would require visibility of paths and network conditions in other domains, but that information may not be available, might not be complete, and is not necessarily trustworthy.

## 8. Overview of Contemporary TE Practices in Operational IP Networks

This section provides an overview of some TE practices in IP networks. The focus is on aspects of control of the routing function in operational contexts. The intent here is to provide an overview of the commonly used practices; the discussion is not intended to be exhaustive.

Service providers apply many of the TE mechanisms described in this document to optimize the performance of their IP networks, although others choose to not use any of them. These techniques include capacity planning, including adding ECMP options, for long timescales; routing control using IGP metrics and MPLS, as well as path planning and path control using MPLS and SR for medium timescales; and traffic management mechanisms for short timescales.

* Capacity planning is an important component of how a service provider plans an effective IP network. These plans may take the following aspects into account: location of any new links or nodes, WECMP algorithms, existing and predicted traffic patterns, costs, link capacity, topology, routing design, and survivability.

* Performance optimization of operational networks is usually an ongoing process in which traffic statistics, performance parameters, and fault indicators are continually collected from the network. This empirical data is analyzed and used to trigger TE mechanisms. Tools that perform what-if analysis can also be used to assist the TE process by reviewing scenarios before a new set of configurations are implemented in the operational network.

* Real-time intra-domain TE using the IGP is done by increasing the OSPF or IS-IS metric of a congested link until enough traffic has been diverted away from that link. This approach has some limitations as discussed in Section 6.2. Intra-domain TE approaches [RR94] [FT00] [FT01] [WANG] take traffic matrix, network topology, and network performance objectives as input and produce link metrics and load-sharing ratios. These processes open the possibility for intra-domain TE with IGP to be done in a more systematic way.

Administrators of MPLS-TE networks specify and configure link attributes and resource constraints such as maximum reservable bandwidth and resource class attributes for the links in the domain. A link-state IGP that supports TE extensions (IS-IS-TE or OSPF-TE) is used to propagate information about network topology and link attributes to all routers in the domain. Network administrators specify the LSPs that are to originate at each router. For each LSP, the network administrator specifies the destination node and the attributes of the LSP that indicate the requirements that are to be satisfied during the path selection process. The attributes may include an explicit path for the LSP to follow, or the originating router may use a local constraint-based routing process to compute the path of the LSP. RSVP-TE is used as a signaling protocol to instantiate the LSPs. By assigning proper bandwidth values to links and LSPs, congestion caused by uneven traffic distribution can be avoided or mitigated.

The bandwidth attributes of an LSP relate to the bandwidth requirements of traffic that flows through the LSP. The traffic attribute of an LSP can be modified to accommodate persistent shifts in demand (traffic growth or reduction). If network congestion occurs due to unexpected events, existing LSPs can be rerouted to alleviate the situation, or the network administrator can configure new LSPs to divert some traffic to alternative paths. The reservable bandwidth of the congested links can also be reduced to force some LSPs to be rerouted to other paths. A traffic matrix in an MPLS domain can also be estimated by monitoring the traffic on LSPs. Such traffic statistics can be used for a variety of purposes including network planning and network optimization.

Network management and planning systems have evolved and assumed a lot of the responsibility for determining traffic paths in TE networks. This allows a network-wide view of resources and facilitates coordination of the use of resources for all traffic flows in the network. Initial solutions using a PCE to perform path computation on behalf of network routers have given way to an approach that follows the SDN architecture. A stateful PCE is able to track all of the LSPs in the network and can redistribute them to make better use of the available resources. Such a PCE can form part of a network orchestrator that uses PCEP or some other configuration and management interface to instruct the signaling protocol or

directly program the routers.

SR leverages a centralized TE controller and either an MPLS or IPv6 forwarding plane but does not need to use a signaling protocol or management plane protocol to reserve resources in the routers. All resource reservation is logical within the controller and is not distributed to the routers. Packets are steered through the network using SR, and this may have configuration and operational scaling benefits.

As mentioned in Section 7, there is usually no direct control over the distribution of inbound traffic to a domain. Therefore, the main goal of inter-domain TE is to optimize the distribution of outbound traffic between multiple inter-domain links. When operating a geographically widespread network (such as for a multi-national or global network provider), maintaining the ability to operate the network in a regional fashion where desired, while continuing to take advantage of the benefits of a globally interconnected network, also becomes an important objective.

Inter-domain TE with BGP begins with the placement of multiple peering interconnection points that are in close proximity to traffic sources/destinations and offer lowest-cost paths across the network between the peering points and the sources/destinations. Some location-decision problems that arise in association with inter-domain routing are discussed in [AWD5].

Once the locations of the peering interconnects have been determined and implemented, the network operator decides how best to handle the routes advertised by the peer, as well as how to propagate the peer’s routes within their network. One way to engineer outbound traffic flows in a network with many peering interconnects is to create a hierarchy of peers. Generally, the shortest AS paths will be chosen to forward traffic, but BGP metrics can be used to prefer some peers and so favor particular paths. Preferred peers are those peers attached through peering interconnects with the most available capacity. Changes may be needed, for example, to deal with a "problem peer" who is difficult to work with on upgrades or is charging high prices for connectivity to their network. In that case, the peer may be given a reduced preference. This type of change can affect a large amount of traffic and is only used after other methods have failed to provide the desired results.

When there are multiple exit points toward a given peer, and only one of them is congested, it is not necessary to shift traffic away from the peer entirely, but only from the one congested connection. This can be achieved by using passive IGP metrics, AS_PATH filtering, or prefix filtering.

## 9. Security Considerations

In general, TE mechanisms are security neutral, and this document does not introduce new security issues.

Network security is, of course, an important issue, and TE mechanisms can have benefits and drawbacks:

* TE may use tunnels that can slightly help protect traffic from inspection and that, in some cases, can be secured using encryption.

* TE puts traffic onto predictable paths within the network that may make it easier to find and attack.

* TE often increases the complexity of operation and management of the network, which may lead to errors that compromise security.

* TE enables traffic to be steered onto more secure links or to more secure parts of the network.

* TE can be used to steer traffic through network nodes that are able to provide additional security functions.

The consequences of attacks on the control and management protocols used to operate TE networks can be significant:

* Traffic can be hijacked to pass through specific nodes that perform inspection or even to be delivered to the wrong place.

* Traffic can be steered onto paths that deliver quality that is below the desired quality.

* Networks can be congested or have resources on key links consumed.

Thus, it is important to use adequate protection mechanisms, such as authentication, on all protocols used to deliver TE.

Certain aspects of a network may be deduced from the details of the TE paths that are used. For example, the link connectivity of the network and the quality and load on individual links may be inferred from knowing the paths of traffic and the requirements they place on the network (for example, by seeing the control messages or through path-trace techniques). Such knowledge can be used to launch targeted attacks (for example, taking down critical links) or can reveal commercially sensitive information (for example, whether a network is close to capacity). Therefore, network operators may choose techniques that mask or hide information from within the network.

External control interfaces that are introduced to provide additional control and management of TE systems (see Section 5.1.2) provide flexibility to management and to customers, but they do so at the risk of exposing the internals of a network to potentially malicious actors. The protocols used at these interfaces must be secured to protect against snooping and modification, and use of the interfaces must be authenticated.

## 10. IANA Considerations

This document has no IANA actions.

## 11. Informative References


[AFD03] Pan, R., Breslau, L., Prabhakar, B., and S. Shenker, "Approximate fairness through differential dropping", ACM SIGCOMM Computer Communication Review, Volume 33, Issue 2, Pages 23-39, DOI 10.1145/956981.956985, April 2003, <https://dl.acm.org/doi/10.1145/956981.956985>.

[AJ19] Adekitan, A., Abolade, J., and O. Shobayo, "Data mining approach for predicting the daily Internet data traffic of a smart university", Journal of Big Data, Volume 6, Number 1, Page 1, DOI 10.1186/s40537-019-0176-5, February 2019, <https://journalofbigdata.springeropen.com/track/pdf/10.1186/s40537-019-0176-5.pdf>.

[ATSSS] 3GPP, "Study on access traffic steering, switch and splitting support in the 5G System (5GS) architecture", Release 16, 3GPP TR 23.793, December 2018, <https://www.3gpp.org/ftp//Specs/archive/23_series/23.793/23793-g00.zip>.

[AWD2] Awduche, D., "MPLS and traffic engineering in IP networks", IEEE Communications Magazine, Volume 37, Issue 12, Pages 42-47, DOI 10.1109/35.809383, December 1999, <https://ieeexplore.ieee.org/document/809383>.

[AWD5] Awduche, D., "An approach to optimal peering between autonomous systems in the Internet", Proceedings 7th International Conference on Computer Communications and Networks (Cat. No. 98EX226), DOI 10.1109/ICCCN.1998.998795, October 1998, <https://ieeexplore.ieee.org/document/998795>.

[E.360.1] ITU-T, "Framework for QoS routing and related traffic engineering methods for IP-, ATM-, and TDM-based multiservice networks", ITU-T Recommendation E.360.1, May 2002, <https://www.itu.int/rec/T-REC-E.360.1-200205-I/en>.

[ENHANCED-VPN] Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for NRP-based Enhanced Virtual Private Network", Work in Progress, Internet-Draft, draft-ietf-teas-enhanced-vpn-17, 25 December 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-enhanced-vpn-17>.

[Err309] RFC Errata, Erratum ID 309, RFC 3272, <https://www.rfc-editor.org/errata/eid309>.

[EVPN-UNEQUAL-LB] Malhotra, N., Ed., Sajassi, A., Rabadan, J., Drake, J., Lingala, A., and S. Thoria, "Weighted Multi-Path Procedures for EVPN Multi-Homing", Work in Progress, Internet-Draft, draft-ietf-bess-evpn-unequal-lb-21, 7 December 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-evpn-unequal-lb-21>.

[FLJA93] Floyd, S. and V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance", IEEE/ACM Transactions on Networking, Volume 1, Issue 4, Pages 397-413, DOI 10.1109/90.251892, August 1993, <https://www.icir.org/floyd/papers/early.twocolumn.pdf>.

[FT00] Fortz, B. and M. Thorup, "Internet Traffic Engineering by Optimizing OSPF Weights", Proceedings IEEE INFOCOM 2000, DOI 10.1109/INFCOM.2000.832225, March 2000, <https://www.cs.cornell.edu/courses/cs619/2004fa/documents/ospf_opt.pdf>.

[FT01] Fortz, B. and M. Thorup, "Optimizing OSPF/IS-IS Weights in a Changing World", IEEE Journal on Selected Areas in Communications, DOI 10.1109/JSAC.2002.1003042, May 2002, <https://ieeexplore.ieee.org/document/1003042>.

[GRPC] gRPC Authors, "gRPC: A high performance, open source universal RPC framework", <https://grpc.io>.

[KELLY] Kelly, F., "Notes on effective bandwidths", Oxford University Press, 1996.

[MA] Ma, Q., "Quality-of-Service Routing in Integrated Services Networks", Ph.D. Dissertation, Carnegie Mellon University, CMU-CS-98-138, January 1998, <https://apps.dtic.mil/sti/pdfs/ADA352299.pdf>.

[MATE] Elwalid, A., Jin, C., Low, S., and I. Widjaja, "MATE: MPLS Adaptive Traffic Engineering", Proceedings IEEE INFOCOM

2001, Conference on Computer Communications, Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No. 01CH37213), DOI 10.1109/INFCOM.2001.916625, August 2002, <https://www.yumpu.com/en/document/view/35140398/mate-mpls-adaptive-traffic-engineering-infocom-ieee-xplore/8>.

[MR99] Mitra, D. and K.G. Ramakrishnan, "A case study of multiservice, multipriority traffic engineering design for data networks", Seamless Interconnection for Universal Services, Global Telecommunications Conference, GLOBECOM'99, (Cat. No. 99CH37042), DOI 10.1109/GLOCOM.1999.830281, December 1999, <https://ieeexplore.ieee.org/document/830281>.

[MULTIPATH-DCCP] Amend, M., Ed., Brunstrom, A., Kassler, A., Rakocevic, V., and S. Johnson, "DCCP Extensions for Multipath Operation with Multiple Addresses", Work in Progress, Internet-Draft, draft-ietf-tsvwg-multipath-dccp-11, 12 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-tsvwg-multipath-dccp-11>.

[NETWORK-SLICES] Farrel, A., Ed., Drake, J., Ed., Rokui, R., Homma, S., Makhijani, K., Contreras, L. M., and J. Tantsura, "A Framework for Network Slices in Networks Built from IETF Technologies", Work in Progress, Internet-Draft, draft-ietf-teas-ietf-network-slices-25, 14 September 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-ietf-network-slices-25>.

[PERFORMANCE-ROUTING] Xu, X., Hegde, S., Talaulikar, K., Boucadair, M., and C. Jacquenet, "Performance-based BGP Routing Mechanism", Work in Progress, Internet-Draft, draft-ietf-idr-performance-routing-03, 22 December 2020, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-performance-routing-03>.

[QUIC-MULTIPATH] Liu, Y., Ed., Ma, Y., Ed., De Coninck, Q., Ed., Bonaventure, O., Huitema, C., and M. Khlewind, Ed., "Multipath Extension for QUIC", Work in Progress, Internet-Draft, draft-ietf-quic-multipath-06, 23 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-quic-multipath-06>.

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, DOI 10.17487/RFC0791, September 1981, <https://www.rfc-editor.org/info/rfc791>.

[RFC1102] Clark, D., "Policy routing in Internet protocols", RFC 1102, DOI 10.17487/RFC1102, May 1989, <https://www.rfc-editor.org/info/rfc1102>.

[RFC1104] Braun, H., "Models of policy based routing", RFC 1104, DOI 10.17487/RFC1104, June 1989, <https://www.rfc-editor.org/info/rfc1104>.

[RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, DOI 10.17487/RFC2205, September 1997, <https://www.rfc-editor.org/info/rfc2205>.

[RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis,

"Framework for IP Performance Metrics", RFC 2330, DOI 10.17487/RFC2330, May 1998, <https://www.rfc-editor.org/info/rfc2330>.

[RFC2386] Crawley, E., Nair, R., Rajagopalan, B., and H. Sandick, "A Framework for QoS-based Routing in the Internet", RFC 2386, DOI 10.17487/RFC2386, August 1998, <https://www.rfc-editor.org/info/rfc2386>.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, DOI 10.17487/RFC2474, December 1998, <https://www.rfc-editor.org/info/rfc2474>.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, DOI 10.17487/RFC2475, December 1998, <https://www.rfc-editor.org/info/rfc2475>.

[RFC2597] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, DOI 10.17487/RFC2597, June 1999, <https://www.rfc-editor.org/info/rfc2597>.

[RFC2678] Mahdavi, J. and V. Paxson, "IPPM Metrics for Measuring Connectivity", RFC 2678, DOI 10.17487/RFC2678, September 1999, <https://www.rfc-editor.org/info/rfc2678>.

[RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, DOI 10.17487/RFC2702, September 1999, <https://www.rfc-editor.org/info/rfc2702>.

[RFC2722] Brownlee, N., Mills, C., and G. Ruth, "Traffic Flow Measurement: Architecture", RFC 2722, DOI 10.17487/RFC2722, October 1999, <https://www.rfc-editor.org/info/rfc2722>.

[RFC2753] Yavatkar, R., Pendarakis, D., and R. Guerin, "A Framework for Policy-based Admission Control", RFC 2753, DOI 10.17487/RFC2753, January 2000, <https://www.rfc-editor.org/info/rfc2753>.

[RFC2961] Berger, L., Gan, D., Swallow, G., Pan, P., Tommasi, F., and S. Molendini, "RSVP Refresh Overhead Reduction Extensions", RFC 2961, DOI 10.17487/RFC2961, April 2001, <https://www.rfc-editor.org/info/rfc2961>.

[RFC2998] Bernet, Y., Ford, P., Yavatkar, R., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J., and E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks", RFC 2998, DOI 10.17487/RFC2998, November 2000, <https://www.rfc-editor.org/info/rfc2998>.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <https://www.rfc-editor.org/info/rfc3031>.

[RFC3086] Nichols, K. and B. Carpenter, "Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification", RFC 3086, DOI 10.17487/RFC3086, April 2001, <https://www.rfc-editor.org/info/rfc3086>.

[RFC3124] Balakrishnan, H. and S. Seshan, "The Congestion Manager",

RFC 3124, DOI 10.17487/RFC3124, June 2001,
<https://www.rfc-editor.org/info/rfc3124>.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, DOI 10.17487/RFC3168, September 2001, <https://www.rfc-editor.org/info/rfc3168>.

[RFC3175] Baker, F., Iturralde, C., Le Faucheur, F., and B. Davie, "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175, DOI 10.17487/RFC3175, September 2001, <https://www.rfc-editor.org/info/rfc3175>.

[RFC3198] Westerinen, A., Schnizlein, J., Strassner, J., Scherling, M., Quinn, B., Herzog, S., Huynh, A., Carlson, M., Perry, J., and S. Waldbusser, "Terminology for Policy-Based Management", RFC 3198, DOI 10.17487/RFC3198, November 2001, <https://www.rfc-editor.org/info/rfc3198>.

[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>.

[RFC3270] Le Faucheur, F., Ed., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, DOI 10.17487/RFC3270, May 2002, <https://www.rfc-editor.org/info/rfc3270>.

[RFC3272] Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., and X. Xiao, "Overview and Principles of Internet Traffic Engineering", RFC 3272, DOI 10.17487/RFC3272, May 2002, <https://www.rfc-editor.org/info/rfc3272>.

[RFC3469] Sharma, V., Ed. and F. Hellstrand, Ed., "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery", RFC 3469, DOI 10.17487/RFC3469, February 2003, <https://www.rfc-editor.org/info/rfc3469>.

[RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, DOI 10.17487/RFC3473, January 2003, <https://www.rfc-editor.org/info/rfc3473>.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <https://www.rfc-editor.org/info/rfc3630>.

[RFC3945] Mannie, E., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, DOI 10.17487/RFC3945, October 2004, <https://www.rfc-editor.org/info/rfc3945>.

[RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, DOI 10.17487/RFC4090, May 2005, <https://www.rfc-editor.org/info/rfc4090>.

[RFC4124] Le Faucheur, F., Ed., "Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering", RFC 4124, DOI 10.17487/RFC4124, June 2005, <https://www.rfc-editor.org/info/rfc4124>.

[RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, DOI 10.17487/RFC4203, October 2005, <https://www.rfc-editor.org/info/rfc4203>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC4340] Kohler, E., Handley, M., and S. Floyd, "Datagram Congestion Control Protocol (DCCP)", RFC 4340, DOI 10.17487/RFC4340, March 2006, <https://www.rfc-editor.org/info/rfc4340>.

[RFC4461] Yasukawa, S., Ed., "Signaling Requirements for Point-to-Multipoint Traffic-Engineered MPLS Label Switched Paths (LSPs)", RFC 4461, DOI 10.17487/RFC4461, April 2006, <https://www.rfc-editor.org/info/rfc4461>.

[RFC4594] Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ Service Classes", RFC 4594, DOI 10.17487/RFC4594, August 2006, <https://www.rfc-editor.org/info/rfc4594>.

[RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <https://www.rfc-editor.org/info/rfc4655>.

[RFC4872] Lang, J.P., Ed., Rekhter, Y., Ed., and D. Papadimitriou, Ed., "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, DOI 10.17487/RFC4872, May 2007, <https://www.rfc-editor.org/info/rfc4872>.

[RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel, "GMPLS Segment Recovery", RFC 4873, DOI 10.17487/RFC4873, May 2007, <https://www.rfc-editor.org/info/rfc4873>.

[RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, DOI 10.17487/RFC4875, May 2007, <https://www.rfc-editor.org/info/rfc4875>.

[RFC5151] Farrel, A., Ed., Ayyangar, A., and JP. Vasseur, "Inter-Domain MPLS and GMPLS Traffic Engineering -- Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 5151, DOI 10.17487/RFC5151, February 2008, <https://www.rfc-editor.org/info/rfc5151>.

[RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, DOI 10.17487/RFC5250, July 2008, <https://www.rfc-editor.org/info/rfc5250>.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <https://www.rfc-editor.org/info/rfc5305>.

[RFC5329] Ishiguro, K., Manral, V., Davey, A., and A. Lindem, Ed., "Traffic Engineering Extensions to OSPF Version 3", RFC 5329, DOI 10.17487/RFC5329, September 2008, <https://www.rfc-editor.org/info/rfc5329>.

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, DOI 10.17487/RFC5331, August 2008, <https://www.rfc-editor.org/info/rfc5331>.

[RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5357, DOI 10.17487/RFC5357, October 2008, <https://www.rfc-editor.org/info/rfc5357>.

[RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394, DOI 10.17487/RFC5394, December 2008, <https://www.rfc-editor.org/info/rfc5394>.

[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>.

[RFC5470] Sadasivan, G., Brownlee, N., Claise, B., and J. Quittek, "Architecture for IP Flow Information Export", RFC 5470, DOI 10.17487/RFC5470, March 2009, <https://www.rfc-editor.org/info/rfc5470>.

[RFC5472] Zseby, T., Boschi, E., Brownlee, N., and B. Claise, "IP Flow Information Export (IPFIX) Applicability", RFC 5472, DOI 10.17487/RFC5472, March 2009, <https://www.rfc-editor.org/info/rfc5472>.

[RFC5541] Le Roux, JL., Vasseur, JP., and Y. Lee, "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC 5541, DOI 10.17487/RFC5541, June 2009, <https://www.rfc-editor.org/info/rfc5541>.

[RFC5557] Lee, Y., Le Roux, JL., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, DOI 10.17487/RFC5557, July 2009, <https://www.rfc-editor.org/info/rfc5557>.

[RFC5559] Eardley, P., Ed., "Pre-Congestion Notification (PCN) Architecture", RFC 5559, DOI 10.17487/RFC5559, June 2009, <https://www.rfc-editor.org/info/rfc5559>.

[RFC5623] Oki, E., Takeda, T., Le Roux, JL., and A. Farrel, "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering", RFC 5623, DOI 10.17487/RFC5623, September 2009, <https://www.rfc-editor.org/info/rfc5623>.

[RFC5664] Halevy, B., Welch, B., and J. Zelenka, "Object-Based Parallel NFS (pNFS) Operations", RFC 5664, DOI 10.17487/RFC5664, January 2010, <https://www.rfc-editor.org/info/rfc5664>.

[RFC5671] Yasukawa, S. and A. Farrel, Ed., "Applicability of the Path Computation Element (PCE) to Point-to-Multipoint (P2MP) MPLS and GMPLS Traffic Engineering (TE)", RFC 5671, DOI 10.17487/RFC5671, October 2009, <https://www.rfc-editor.org/info/rfc5671>.

[RFC5693] Seedorf, J. and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement", RFC 5693, DOI 10.17487/RFC5693, October 2009, <https://www.rfc-editor.org/info/rfc5693>.

[RFC6107] Shiomoto, K., Ed. and A. Farrel, Ed., "Procedures for Dynamically Signaled Hierarchical Label Switched Paths", RFC 6107, DOI 10.17487/RFC6107, February 2011, <https://www.rfc-editor.org/info/rfc6107>.

[RFC6119] Harrison, J., Berger, J., and M. Bartlett, "IPv6 Traffic Engineering in IS-IS", RFC 6119, DOI 10.17487/RFC6119, February 2011, <https://www.rfc-editor.org/info/rfc6119>.

[RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, <https://www.rfc-editor.org/info/rfc6241>.

[RFC6372] Sprecher, N., Ed. and A. Farrel, Ed., "MPLS Transport Profile (MPLS-TP) Survivability Framework", RFC 6372, DOI 10.17487/RFC6372, September 2011, <https://www.rfc-editor.org/info/rfc6372>.

[RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, DOI 10.17487/RFC6374, September 2011, <https://www.rfc-editor.org/info/rfc6374>.

[RFC6601] Ash, G., Ed. and D. McDysan, "Generic Connection Admission Control (GCAC) Algorithm Specification for IP/MPLS Networks", RFC 6601, DOI 10.17487/RFC6601, April 2012, <https://www.rfc-editor.org/info/rfc6601>.

[RFC6805] King, D., Ed. and A. Farrel, Ed., "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, DOI 10.17487/RFC6805, November 2012, <https://www.rfc-editor.org/info/rfc6805>.

[RFC7011] Claise, B., Ed., Trammell, B., Ed., and P. Aitken, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information", STD 77, RFC 7011, DOI 10.17487/RFC7011, September 2013, <https://www.rfc-editor.org/info/rfc7011>.

[RFC7149] Boucadair, M. and C. Jacquenet, "Software-Defined Networking: A Perspective from within a Service Provider Environment", RFC 7149, DOI 10.17487/RFC7149, March 2014, <https://www.rfc-editor.org/info/rfc7149>.

[RFC7285] Alimi, R., Ed., Penno, R., Ed., Yang, Y., Ed., Kiesel, S., Previdi, S., Roome, W., Shalunov, S., and R. Woundy, "Application-Layer Traffic Optimization (ALTO) Protocol", RFC 7285, DOI 10.17487/RFC7285, September 2014, <https://www.rfc-editor.org/info/rfc7285>.

[RFC7390] Rahman, A., Ed. and E. Dijk, Ed., "Group Communication for the Constrained Application Protocol (CoAP)", RFC 7390, DOI 10.17487/RFC7390, October 2014, <https://www.rfc-editor.org/info/rfc7390>.

[RFC7426] Haleplidis, E., Ed., Pentikousis, K., Ed., Denazis, S., Hadi Salim, J., Meyer, D., and O. Koufopavlou, "Software-Defined Networking (SDN): Layers and Architecture Terminology", RFC 7426, DOI 10.17487/RFC7426, January 2015, <https://www.rfc-editor.org/info/rfc7426>.

[RFC7471] Giacalone, S., Ward, D., Drake, J., Atlas, A., and S. Previdi, "OSPF Traffic Engineering (TE) Metric

Extensions", RFC 7471, DOI 10.17487/RFC7471, March 2015, <https://www.rfc-editor.org/info/rfc7471>.

[RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <https://www.rfc-editor.org/info/rfc7491>.

[RFC7551] Zhang, F., Ed., Jing, R., and R. Gandhi, Ed., "RSVP-TE Extensions for Associated Bidirectional Label Switched Paths (LSPs)", RFC 7551, DOI 10.17487/RFC7551, May 2015, <https://www.rfc-editor.org/info/rfc7551>.

[RFC7567] Baker, F., Ed. and G. Fairhurst, Ed., "IETF Recommendations Regarding Active Queue Management", BCP 197, RFC 7567, DOI 10.17487/RFC7567, July 2015, <https://www.rfc-editor.org/info/rfc7567>.

[RFC7665] Halpern, J., Ed. and C. Pignataro, Ed., "Service Function Chaining (SFC) Architecture", RFC 7665, DOI 10.17487/RFC7665, October 2015, <https://www.rfc-editor.org/info/rfc7665>.

[RFC7679] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Delay Metric for IP Performance Metrics (IPPM)", STD 81, RFC 7679, DOI 10.17487/RFC7679, January 2016, <https://www.rfc-editor.org/info/rfc7679>.

[RFC7680] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Loss Metric for IP Performance Metrics (IPPM)", STD 82, RFC 7680, DOI 10.17487/RFC7680, January 2016, <https://www.rfc-editor.org/info/rfc7680>.

[RFC7923] Voit, E., Clemm, A., and A. Gonzalez Prieto, "Requirements for Subscription to YANG Datastores", RFC 7923, DOI 10.17487/RFC7923, June 2016, <https://www.rfc-editor.org/info/rfc7923>.

[RFC7926] Farrel, A., Ed., Drake, J., Bitar, N., Swallow, G., Ceccarelli, D., and X. Zhang, "Problem Statement and Architecture for Information Exchange between Interconnected Traffic-Engineered Networks", BCP 206, RFC 7926, DOI 10.17487/RFC7926, July 2016, <https://www.rfc-editor.org/info/rfc7926>.

[RFC7950] Bjorklund, M., Ed., "The YANG 1.1 Data Modeling Language", RFC 7950, DOI 10.17487/RFC7950, August 2016, <https://www.rfc-editor.org/info/rfc7950>.

[RFC8033] Pan, R., Natarajan, P., Baker, F., and G. White, "Proportional Integral Controller Enhanced (PIE): A Lightweight Control Scheme to Address the Bufferbloat Problem", RFC 8033, DOI 10.17487/RFC8033, February 2017, <https://www.rfc-editor.org/info/rfc8033>.

[RFC8034] White, G. and R. Pan, "Active Queue Management (AQM) Based on Proportional Integral Controller Enhanced (PIE) for Data-Over-Cable Service Interface Specifications (DOCSIS) Cable Modems", RFC 8034, DOI 10.17487/RFC8034, February 2017, <https://www.rfc-editor.org/info/rfc8034>.

[RFC8040] Bierman, A., Bjorklund, M., and K. Watsen, "RESTCONF Protocol", RFC 8040, DOI 10.17487/RFC8040, January 2017, <https://www.rfc-editor.org/info/rfc8040>.

[RFC8051] Zhang, X., Ed. and I. Minei, Ed., "Applicability of a

Stateful Path Computation Element (PCE)", RFC 8051, DOI 10.17487/RFC8051, January 2017, <https://www.rfc-editor.org/info/rfc8051>.

[RFC8189] Randriamasy, S., Roome, W., and N. Schwan, "Multi-Cost Application-Layer Traffic Optimization (ALTO)", RFC 8189, DOI 10.17487/RFC8189, October 2017, <https://www.rfc-editor.org/info/rfc8189>.

[RFC8231] Crabbe, E., Minei, I., Medved, J., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", RFC 8231, DOI 10.17487/RFC8231, September 2017, <https://www.rfc-editor.org/info/rfc8231>.

[RFC8259] Bray, T., Ed., "The JavaScript Object Notation (JSON) Data Interchange Format", STD 90, RFC 8259, DOI 10.17487/RFC8259, December 2017, <https://www.rfc-editor.org/info/rfc8259>.

[RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <https://www.rfc-editor.org/info/rfc8279>.

[RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <https://www.rfc-editor.org/info/rfc8281>.

[RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <https://www.rfc-editor.org/info/rfc8283>.

[RFC8290] Hoeiland-Joergensen, T., McKenney, P., Taht, D., Gettys, J., and E. Dumazet, "The Flow Queue CoDel Packet Scheduler and Active Queue Management Algorithm", RFC 8290, DOI 10.17487/RFC8290, January 2018, <https://www.rfc-editor.org/info/rfc8290>.

[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>.

[RFC8453] Ceccarelli, D., Ed. and Y. Lee, Ed., "Framework for Abstraction and Control of TE Networks (ACTN)", RFC 8453, DOI 10.17487/RFC8453, August 2018, <https://www.rfc-editor.org/info/rfc8453>.

[RFC8570] Ginsberg, L., Ed., Previdi, S., Ed., Giacalone, S., Ward, D., Drake, J., and Q. Wu, "IS-IS Traffic Engineering (TE) Metric Extensions", RFC 8570, DOI 10.17487/RFC8570, March 2019, <https://www.rfc-editor.org/info/rfc8570>.

[RFC8571] Ginsberg, L., Ed., Previdi, S., Wu, Q., Tantsura, J., and C. Filsfils, "BGP - Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions", RFC 8571, DOI 10.17487/RFC8571, March 2019, <https://www.rfc-editor.org/info/rfc8571>.

[RFC8655] Finn, N., Thubert, P., Varga, B., and J. Farkas,

"Deterministic Networking Architecture", RFC 8655, DOI 10.17487/RFC8655, October 2019, <https://www.rfc-editor.org/info/rfc8655>.

[RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <https://www.rfc-editor.org/info/rfc8664>.

[RFC8684] Ford, A., Raiciu, C., Handley, M., Bonaventure, O., and C. Paasch, "TCP Extensions for Multipath Operation with Multiple Addresses", RFC 8684, DOI 10.17487/RFC8684, March 2020, <https://www.rfc-editor.org/info/rfc8684>.

[RFC8685] Zhang, F., Zhao, Q., Gonzalez de Dios, O., Casellas, R., and D. King, "Path Computation Element Communication Protocol (PCEP) Extensions for the Hierarchical Path Computation Element (H-PCE) Architecture", RFC 8685, DOI 10.17487/RFC8685, December 2019, <https://www.rfc-editor.org/info/rfc8685>.

[RFC8795] Liu, X., Bryskin, I., Beeram, V., Saad, T., Shah, H., and O. Gonzalez de Dios, "YANG Data Model for Traffic Engineering (TE) Topologies", RFC 8795, DOI 10.17487/RFC8795, August 2020, <https://www.rfc-editor.org/info/rfc8795>.

[RFC8803] Bonaventure, O., Ed., Boucadair, M., Ed., Gundavelli, S., Seo, S., and B. Hesmans, "0-RTT TCP Convert Protocol", RFC 8803, DOI 10.17487/RFC8803, July 2020, <https://www.rfc-editor.org/info/rfc8803>.

[RFC8896] Randriamasy, S., Yang, R., Wu, Q., Deng, L., and N. Schwan, "Application-Layer Traffic Optimization (ALTO) Cost Calendar", RFC 8896, DOI 10.17487/RFC8896, November 2020, <https://www.rfc-editor.org/info/rfc8896>.

[RFC8938] Varga, B., Ed., Farkas, J., Berger, L., Malis, A., and S. Bryant, "Deterministic Networking (DetNet) Data Plane Framework", RFC 8938, DOI 10.17487/RFC8938, November 2020, <https://www.rfc-editor.org/info/rfc8938>.

[RFC8955] Loibl, C., Hares, S., Raszuk, R., McPherson, D., and M. Bacher, "Dissemination of Flow Specification Rules", RFC 8955, DOI 10.17487/RFC8955, December 2020, <https://www.rfc-editor.org/info/rfc8955>.

[RFC8972] Mirsky, G., Min, X., Nydell, H., Foote, R., Masputra, A., and E. Ruffini, "Simple Two-Way Active Measurement Protocol Optional Extensions", RFC 8972, DOI 10.17487/RFC8972, January 2021, <https://www.rfc-editor.org/info/rfc8972>.

[RFC9000] Iyengar, J., Ed. and M. Thomson, Ed., "QUIC: A UDP-Based Multiplexed and Secure Transport", RFC 9000, DOI 10.17487/RFC9000, May 2021, <https://www.rfc-editor.org/info/rfc9000>.

[RFC9023] Varga, B., Ed., Farkas, J., Malis, A., and S. Bryant, "Deterministic Networking (DetNet) Data Plane: IP over IEEE 802.1 Time-Sensitive Networking (TSN)", RFC 9023, DOI 10.17487/RFC9023, June 2021, <https://www.rfc-editor.org/info/rfc9023>.

[RFC9040] Touch, J., Welzl, M., and S. Islam, "TCP Control Block

Interdependence", RFC 9040, DOI 10.17487/RFC9040, July 2021, <https://www.rfc-editor.org/info/rfc9040>.

[RFC9113] Thomson, M., Ed. and C. Benfield, Ed., "HTTP/2", RFC 9113, DOI 10.17487/RFC9113, June 2022, <https://www.rfc-editor.org/info/rfc9113>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

[RFC9262] Eckert, T., Ed., Menth, M., and G. Cauchie, "Tree Engineering for Bit Index Explicit Replication (BIER-TE)", RFC 9262, DOI 10.17487/RFC9262, October 2022, <https://www.rfc-editor.org/info/rfc9262>.

[RFC9298] Schinazi, D., "Proxying UDP in HTTP", RFC 9298, DOI 10.17487/RFC9298, August 2022, <https://www.rfc-editor.org/info/rfc9298>.

[RFC9315] Clemm, A., Ciavaglia, L., Granville, L. Z., and J. Tantsura, "Intent-Based Networking - Concepts and Definitions", RFC 9315, DOI 10.17487/RFC9315, October 2022, <https://www.rfc-editor.org/info/rfc9315>.

[RFC9332] De Schepper, K., Briscoe, B., Ed., and G. White, "Dual-Queue Coupled Active Queue Management (AQM) for Low Latency, Low Loss, and Scalable Throughput (L4S)", RFC 9332, DOI 10.17487/RFC9332, January 2023, <https://www.rfc-editor.org/info/rfc9332>.

[RFC9350] Psenak, P., Ed., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", RFC 9350, DOI 10.17487/RFC9350, February 2023, <https://www.rfc-editor.org/info/rfc9350>.

[RFC9439] Wu, Q., Yang, Y., Lee, Y., Dhody, D., Randriamasy, S., and L. Contreras, "Application-Layer Traffic Optimization (ALTO) Performance Cost Metrics", RFC 9439, DOI 10.17487/RFC9439, August 2023, <https://www.rfc-editor.org/info/rfc9439>.

[RFC9502] Britto, W., Hegde, S., Kaneriya, P., Shetty, R., Bonica, R., and P. Psenak, "IGP Flexible Algorithm in IP Networks", RFC 9502, DOI 10.17487/RFC9502, November 2023, <https://www.rfc-editor.org/info/rfc9502>.

[RFC9552] Talaulikar, K., Ed., "Distribution of Link-State and Traffic Engineering Information Using BGP", RFC 9552, DOI 10.17487/RFC9552, December 2023, <https://www.rfc-editor.org/info/rfc9552>.

[RR94] Rodrigues, M. and K.G. Ramakrishnan, "Optimal routing in shortest-path data networks", Bell Labs Technical Journal, Volume 6, Issue 1, Pages 117-138, DOI 10.1002/bltj.2267, August 2002, <https://onlinelibrary.wiley.com/doi/abs/10.1002/bltj.2267>.

[SLDC98] Suter, B., Lakshman, T.V., Stiliadis, D., and A.K. Choudhury, "Design considerations for supporting TCP with per-flow queueing", Proceedings IEEE INFOCOM ’98, DOI 10.1109/INFCOM.1998.659666, April 1998, <https://ieeexplore.ieee.org/document/659666>.

[SR-TE-POLICY]
Previdi, S., Filsfils, C., Talaulikar, K., Ed., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-segment-routing-te-policy-26, 23 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-segment-routing-te-policy-26>.

[SR-TI-LFA]
Bashandy, A., Litkowski, S., Filsfils, C., Francois, P., Decraene, B., and D. Voyer, "Topology Independent Fast Reroute using Segment Routing", Work in Progress, Internet-Draft, draft-ietf-rtgwg-segment-routing-ti-lfa-13, 16 January 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-rtgwg-segment-routing-ti-lfa-13>.

[TE-QoS-ROUTING]
Ash, G., "Traffic Engineering & QoS Methods for IP-, ATM-, & Based Multiservice Networks", Work in Progress, Internet-Draft, draft-ietf-tewg-qos-routing-04, October 2001, <https://datatracker.ietf.org/doc/html/draft-ietf-tewg-qos-routing-04>.

[WANG] Wang, Y., Wang, Z., and L. Zhang, "Internet traffic engineering without full mesh overlaying", Proceedings IEEE INFOCOM 2001, DOI 10.1109/INFCOM.2001.916782, April 2001, <https://ieeexplore.ieee.org/document/916782>.

[XIAO] Xiao, X., Hannan, A., Bailey, B., and L. Ni, "Traffic Engineering with MPLS in the Internet", IEEE Network, Volume 14, Issue 2, Pages 28-33, DOI 10.1109/65.826369, March 2000, <https://courses.cs.washington.edu/courses/cse561/02au/papers/xiao-mpls-net00.pdf>.

[YARE95] Yang, C. and A. Reddy, "A Taxonomy for Congestion Control Algorithms in Packet Switching Networks", IEEE Network, Pages 34-45, DOI 10.1109/65.397042, August 1995, <https://ieeexplore.ieee.org/document/397042>.

## Appendix A. Summary of Changes since RFC 3272

The changes to this document since [RFC3272] are substantial and not easily summarized as section-by-section changes. The material in the document has been moved around considerably, some of it removed, and new text added.

The approach taken here is to list the contents of both [RFC3272] and this document saying, respectively, where the text has been placed and where the text came from.

### A.1. RFC 3272

* Section 1.0 ("Introduction"): Edited in place in Section 1.
  - Section 1.1 ("What is Internet Traffic Engineering?"): Edited in place in Section 1.1.
  - Section 1.2 ("Scope"): Moved to Section 1.3.
  - Section 1.3 ("Terminology"): Moved to Section 1.4 with some obsolete terms removed and a little editing.
* Section 2.0 ("Background"): Retained as Section 2 with some text removed.

- Section 2.1 ("Context of Internet Traffic Engineering"): Retained as Section 2.1.

- Section 2.2 ("Network Context"): Rewritten as Section 2.2.

- Section 2.3 ("Problem Context"): Rewritten as Section 2.3.

  - Section 2.3.1 ("Congestion and its Ramifications"): Retained as Section 2.3.1.

- Section 2.4 ("Solution Context"): Edited as Section 2.4.

  - Section 2.4.1 ("Combating the Congestion Problem"): Reformatted as Section 2.4.1.

- Section 2.5 ("Implementation and Operational Context"): Retained as Section 2.5.

* Section 3.0 ("Traffic Engineering Process Model"): Retained as Section 3.

- Section 3.1 ("Components of the Traffic Engineering Process Model"): Retained as Section 3.1.

- Section 3.2 ("Measurement"): Merged into Section 3.1.

- Section 3.3 ("Modeling, Analysis, and Simulation"): Merged into Section 3.1.

- Section 3.4 ("Optimization"): Merged into Section 3.1.

* Section 4.0 ("Historical Review and Recent Developments"): Retained as Section 5, but the very historic aspects have been deleted.

- Section 4.1 ("Traffic Engineering in Classical Telephone Networks"): Deleted.

- Section 4.2 ("Evolution of Traffic Engineering in the Internet"): Deleted.

- Section 4.3 ("Overlay Model"): Deleted.

- Section 4.4 ("Constraint-Based Routing"): Retained as Section 5.1.3.1, but moved into Section 5.1.

- Section 4.5 ("Overview of Other IETF Projects Related to Traffic Engineering"): Retained as Section 5.1 with many new subsections.

  - Section 4.5.1 ("Integrated Services"): Retained as Section 5.1.1.1.

  - Section 4.5.2 ("RSVP"): Retained as Section 5.1.3.2 with some edits.

  - Section 4.5.3 ("Differentiated Services"): Retained as Section 5.1.1.2.

  - Section 4.5.4 ("MPLS"): Retained as Section 5.1.3.3.

  - Section 4.5.5 ("IP Performance Metrics"): Retained as Section 5.1.3.6.

  - Section 4.5.6 ("Flow Measurement"): Retained as

Section 5.1.3.7 with some reformatting.

- o Section 4.5.7 ("Endpoint Congestion Management"): Retained as Section 5.1.3.8.

- Section 4.6 ("Overview of ITU Activities Related to Traffic Engineering"): Deleted.

- Section 4.7 ("Content Distribution"): Retained as Section 5.2.

* Section 5.0 ("Taxonomy of Traffic Engineering Systems"): Retained as Section 4.

  - Section 5.1 ("Time-Dependent Versus State-Dependent"): Retained as Section 4.1.

  - Section 5.2 ("Offline Versus Online"): Retained as Section 4.2.

  - Section 5.3 ("Centralized Versus Distributed"): Retained as Section 4.3 with additions.

  - Section 5.4 ("Local Versus Global"): Retained as Section 4.4.

  - Section 5.5 ("Prescriptive Versus Descriptive"): Retained as Section 4.5 with additions.

  - Section 5.6 ("Open-Loop Versus Closed-Loop"): Retained as Section 4.6.

  - Section 5.7 ("Tactical vs Strategic"): Retained as Section 4.7.

* Section 6.0 ("Recommendations for Internet Traffic Engineering"): Retained as Section 6.

  - Section 6.1 ("Generic Non-functional Recommendations"): Retained as Section 6.1.

  - Section 6.2 ("Routing Recommendations"): Retained as Section 6.2 with edits.

  - Section 6.3 ("Traffic Mapping Recommendations"): Retained as Section 6.3.

  - Section 6.4 ("Measurement Recommendations"): Retained as Section 6.4.

  - Section 6.5 ("Network Survivability"): Retained as Section 6.6.

    - o Section 6.5.1 ("Survivability in MPLS Based Networks"): Retained as Section 6.6.1.

    - o Section 6.5.2 ("Protection Option"): Retained as Section 6.6.2.

  - Section 6.6 ("Traffic Engineering in Diffserv Environments"): Retained as Section 6.8 with edits.

  - Section 6.7 ("Network Controllability"): Retained as Section 6.9.

* Section 7.0 ("Inter-Domain Considerations"): Retained as Section 7.

* Section 8.0 ("Overview of Contemporary TE Practices in Operational IP Networks"): Retained as Section 8.

* Section 9.0 ("Conclusion"): Removed.

* Section 10.0 ("Security Considerations"): Retained as Section 9 with considerable new text.

## A.2. This Document

* Section 1: Based on Section 1 of [RFC3272].
  - Section 1.1: Based on Section 1.1 of [RFC3272].
  - Section 1.2: New for this document.
  - Section 1.3: Based on Section 1.2 of [RFC3272].
  - Section 1.4: Based on Section 1.3 of [RFC3272].
* Section 2: Based on Section 2 of [RFC3272].
  - Section 2.1: Based on Section 2.1 of [RFC3272].
  - Section 2.2: Based on Section 2.2 of [RFC3272].
  - Section 2.3: Based on Section 2.3 of [RFC3272].
    - Section 2.3.1: Based on Section 2.3.1 of [RFC3272].
  - Section 2.4: Based on Section 2.4 of [RFC3272].
    - Section 2.4.1: Based on Section 2.4.1 of [RFC3272].
  - Section 2.5: Based on Section 2.5 of [RFC3272].
* Section 3: Based on Section 3 of [RFC3272].
  - Section 3.1: Based on Sections 3.1, 3.2, 3.3, and 3.4 of [RFC3272].
* Section 4: Based on Section 5 of [RFC3272].
  - Section 4.1: Based on Section 5.1 of [RFC3272].
  - Section 4.2: Based on Section 5.2 of [RFC3272].
  - Section 4.3: Based on Section 5.3 of [RFC3272].
    - Section 4.3.1: New for this document.
    - Section 4.3.2: New for this document.
  - Section 4.4: Based on Section 5.4 of [RFC3272].
  - Section 4.5: Based on Section 5.5 of [RFC3272].
    - Section 4.5.1: New for this document.
  - Section 4.6: Based on Section 5.6 of [RFC3272].
  - Section 4.7: Based on Section 5.7 of [RFC3272].
* Section 5: Based on Section 4 of [RFC3272].
  - Section 5.1: Based on Section 4.5 of [RFC3272].
    - Section 5.1.1.1: Based on Section 4.5.1 of [RFC3272].

      - Section 5.1.1.2: Based on Section 4.5.3 of [RFC3272].
      - Section 5.1.1.3: New for this document.
      - Section 5.1.1.4: New for this document.
      - Section 5.1.1.5: New for this document.
      - Section 5.1.2.1: New for this document.
      - Section 5.1.2.2: New for this document.
      - Section 5.1.2.3: New for this document.
      - Section 5.1.3.1: Based on Section 4.4 of [RFC3272].
        + Section 5.1.3.1.1: New for this document.
      - Section 5.1.3.2: Based on Section 4.5.2 of [RFC3272].
      - Section 5.1.3.3: Based on Section 4.5.4 of [RFC3272].
      - Section 5.1.3.4: New for this document.
      - Section 5.1.3.5: New for this document.
      - Section 5.1.3.6: Based on Section 4.5.5 of [RFC3272].
      - Section 5.1.3.7: Based on Section 4.5.6 of [RFC3272].
      - Section 5.1.3.8: Based on Section 4.5.7 of [RFC3272].
      - Section 5.1.3.9: New for this document.
      - Section 5.1.3.10: New for this document.
      - Section 5.1.3.11: New for this document.
      - Section 5.1.3.12: New for this document.
      - Section 5.1.3.13: New for this document.
      - Section 5.1.3.14: New for this document.
      - Section 5.1.3.15: New for this document.
    - Section 5.2: Based on Section 4.7 of [RFC3272].
* Section 6: Based on Section 6 of [RFC3272].
    - Section 6.1: Based on Section 6.1 of [RFC3272].
    - Section 6.2: Based on Section 6.2 of [RFC3272].
    - Section 6.3: Based on Section 6.3 of [RFC3272].
    - Section 6.4: Based on Section 6.4 of [RFC3272].
    - Section 6.5: New for this document.
    - Section 6.6: Based on Section 6.5 of [RFC3272].
      - Section 6.6.1: Based on Section 6.5.1 of [RFC3272].
      - Section 6.6.2: Based on Section 6.5.2 of [RFC3272].

- Section 6.7: New for this document.

- Section 6.8: Based on Section 6.6 of [RFC3272].

- Section 6.9: Based on Section 6.7 of [RFC3272].

* Section 7: Based on Section 7 of [RFC3272].

* Section 8: Based on Section 8 of [RFC3272].

* Section 9: Based on Section 10 of [RFC3272].

## Acknowledgments

Much of the text in this document is derived from [RFC3272]. The editor and contributors to this document would like to express their gratitude to all involved in that work. Although the source text has been edited in the production of this document, the original authors should be considered as contributors to this work. They were:

Daniel O. Awduche
Movaz Networks

Angela Chiu
Celion Networks

Anwar Elwalid
Lucent Technologies

Indra Widjaja
Bell Labs, Lucent Technologies

XiPeng Xiao
Redback Networks

The acknowledgements in [RFC3272] were as below. All people who helped in the production of that document also need to be thanked for the carry-over into this new document.

| The authors would like to thank Jim Boyle for inputs on the
| recommendations section, Francois Le Faucheur for inputs on
| Diffserv aspects, Blaine Christian for inputs on measurement,
| Gerald Ash for inputs on routing in telephone networks and for
| text on event-dependent TE methods, Steven Wright for inputs on
| network controllability, and Jonathan Aufderheide for inputs on
| inter-domain TE with BGP. Special thanks to Randy Bush for
| proposing the TE taxonomy based on "tactical vs strategic"
| methods. The subsection describing an "Overview of ITU Activities
| Related to Traffic Engineering" was adapted from a contribution by
| Waisum Lai. Useful feedback and pointers to relevant materials
| were provided by J. Noel Chiappa. Additional comments were
| provided by Glenn Grotefeld during the working last call process.
| Finally, the authors would like to thank Ed Kern, the TEWG co-
| chair, for his comments and support.

The early draft versions of this document were produced by the TEAS Working Group's RFC3272bis Design Team. The full list of members of this team is:

Acee Lindem
Adrian Farrel

Aijun Wang
Daniele Ceccarelli
Dieter Beller
Jeff Tantsura
Julien Meuric
Liu Hua
Loa Andersson
Luis Miguel Contreras
Martin Horneffer
Tarek Saad
Xufeng Liu

The production of this document includes a fix to the original text resulting from an errata report #309 [Err309] by Jean-Michel Grimaldi.

The editor of this document would also like to thank Dhruv Dhody, Gyan Mishra, Joel Halpern, Dave Taht, John Scudder, Rich Salz, Behcet Sarikaya, Bob Briscoe, Erik Kline, Jim Guichard, Martin Duke, and Roman Danyliw for review comments.

This work is partially supported by the European Commission under Horizon 2020 grant agreement number 101015857 Secured autonomic traffic management for a Tera of SDN flows (Teraflow).

## Contributors

The following people contributed substantive text to this document:

Gert Grammel
Email: ggrammel@juniper.net

Loa Andersson
Email: loa@pi.nu

Xufeng Liu
Email: xufeng.liu.ietf@gmail.com

Lou Berger
Email: lberger@labn.net

Jeff Tantsura
Email: jefftant.ietf@gmail.com

Daniel King
Email: daniel@olddog.co.uk

Boris Hassanov
Email: bhassanov@yandex-team.ru

Kiran Makhijani
Email: kiranm@futurewei.com

Dhruv Dhody
Email: dhruv.ietf@gmail.com

Mohamed Boucadair

Email: mohamed.boucadair@orange.com

## Author’s Address

Adrian Farrel (editor)
Old Dog Consulting
Email: adrian@olddog.co.uk